This is an electronic version of an article published in Fergusson DM. A brief introduction to structural equation models. In: Verhulst P, Koot H. (Eds.). *Handbook of Childhood Psychiatric Epidemiology*. Oxford: Oxford University Press, 1995; 122-145.

A Brief Introduction to Structural Equation Models

D M Fergusson

Christchurch Health and Development Study

Christchurch School of Medicine

Christchurch Hospital

Christchurch

New Zealand

Introduction

Over the past two decades there has been a rapid increase in the use of what have come to be known as structural equation models, covariance structure models or LISREL models. These models describe the measurement and causal structure of a set of observed variables in terms of a system of simultaneous linear equations which summarise and encapsulate the assumptions of some prespecified conceptual model of the measurement and causal structure of the observed variables. Applications of these methods may be found in many areas including Economics, Psychology, Sociology and Behavioral Genetics. At the same time, the use of these methods for analysing epidemiological data has been limited and, in particular, the full potential for methods of structural equation modelling in the analysis of child psychiatric epidemiological data has yet to be realised.

In part, this situation may reflect the fact that many epidemiologists, whilst being aware of structural equation modelling techniques, are unsure about the logic of these methods and how they are applied in practice. The major aim of this chapter is to present an introductory and user friendly account of the logic and application of structural equation modelling methods with this account being aimed at a perceived audience of epidemiologists and analysts who are aware of structural equation modelling techniques but who may be unsure about the application of these methods. For this reason the chapter is presented at an introductory level and presents only a limited account of the formal mathematical and statistical foundations of the method. This does not, however, imply that an understanding of the mathematical and statistical foundations of structural equation models can be ignored and, indeed, the effective application of these methods does require that the user understand these issues, and particularly issues relating to model estimation, model identification and model testing.

Information on the mathematical foundations of structural equation models can be found in a number of sources. The early work of Duncan (1975), whilst being somewhat dated in its approach, provides a very useful introduction to model specification and model identification and introduces the basic covariance algebra which underlies these models. The books by Heise (1975) and Kenny (1979) develop these topics further. Long (1983a; 1983b) provides an excellent general introduction to issues of model specification and identification using LISREL. Everitt (1984) provides a useful general introduction to latent variable models (1984). More advanced treatments of the mathematical and statistical foundations of structural equation models may be found in Aigner & Goldberger (1977), McDonald (1978), Joreskog & Sorbom (1979), Browne (1982) and Joreskog & Wold (1982).

The chapter is presented in three general sections. In the first section I present an introductory account of the general LISREL model. The second section illustrates the use of this model through the use of a worked example which examines a five variable model which is sufficiently simple to be specified without recourse to complex matrix algebra but at the same time illustrates many of the principles of model specification, identification, latent variables and model testing. In the final section I attempt to provide an overview of both the advantages and disadvantages of structural equation modelling methods.

The LISREL Model

While there have been many attempts to specify the general form of structural equation models, (see, for example, McDonald, 1978; and Joreskog & Wold, 1982), probably the most useful formulation has been given in the LISREL model described by Joreskog and his associates (Joreskog, 1973; Joreskog & Sorbom, 1989).

Model Specification. In its original formulation, the LISREL model was described by an extremely general series of matrix algebra statements which used a complex system of Greek notation. The complexity with which the model was presented has probably inhibited the use of structural equation modelling methods amongst users who found difficulty with the matrix algebra of the original model. However, in the most recent version of the LISREL program (LISREL 8) many of the complexities of the original model statement have been removed and the program now provides a user friendly method of specifying structural equation models on the basis of path diagrams. To introduce the general LISREL model I will describe construction of these models by reference to path diagrams rather than in terms of the more general matrix algebra formulation of the model.

The point of departure for LISREL models is the matrix of correlations of a set of observed variables. The LISREL model distinguishes between two types of observed variables: a) the x variables which are indicators of latent exogenous or independent variables and b) the y variables which are indicators of latent endogenous or dependent variables. The model represents the relationships between the x and y variables by a system of simultaneous linear equations. These equations may be classified into: a) measurement model equations which describe the relationship between the observed x, y variables and corresponding latent variables; b) the structural equation model equations which describe the relationship between the latent constructs.

The general approach to formulating a LISREL model is most easily illustrated by way of an example. Figure 1 shows a simple four variable model which illustrates many of the principles used to construct such models. This figure assumes that the observed data comprises two x variables (x_1, x_2) and two variables (y_1, y_2). To provide a concrete illustration, the variables x_1, x_2 might be independent measures of childhood body lead burden and the variables y_1, y_2 measures of childhood intelligence. The model makes the following assumptions:

1) The observed measures x1, x2 are fallible measures of a latent variable T_1 (e.g. the child's true but non-observed body lead burden). It is assumed that the

child's status on the latent variable T_1 influences the observed test scores x_1, x_2 but that these scores are also influenced by errors of measurement U_1, U_2. These errors are assumed to be uncorrelated with each other and with the latent variable T_1.

2) A similar set of assumptions is made about the relationships between the observed measures y_1, y_2 and a second latent variable T_2 (e.g. the child's true cognitive ability score).

3) Finally it is assumed that the subject's score on the non-observed variable T_1 (true body lead burden) may causally influence his/her score on the latent variable T_2 (true cognitive ability) but T_2 is also influenced by other non-observed factors represented by the disturbance term E_1.

INSERT FIGURE 1 HERE

The path model in Figure 1 can be presented by the following system of simultaneous linear equations and assumptions.

1) Measurement Model Equations: **2) Structural Equation Model:**

$X_1 = G_1 T_1 + U_1$

$X_2 = G_2 T_1 + U_2$

$Y_1 = G_3 T_2 + U_3$

$Y_1 = G_4 T_2 + U_4$

$T_2 = B T_1 + E_1$

3) Assumptions:

$E(x_i) = E(y_i) = E(T_i) = 0$ $(i = 1,2)$; $\mathrm{Cov}(U_i U_j) = 0$ $(i = j)$

$\mathrm{Cov}(U_i T_k) = 0$ $(i = 1 \ldots 4;\ k = 1,2)$; $\mathrm{Cov}(U_i E_1) = 0$;

$\mathrm{Var}(T_1) = \mathrm{Var}(T_2) = 1$; $\mathrm{Cov}(E_1 T_1) = 0$

The general assumptions underlying this model are: a) that all observed and latent variables are scaled to a mean of zero (thus avoiding the introduction of intercept terms into the model); b) that all relationships represented in the model are adequately described by an additive and linear model; c) that the disturbance terms of Model U_i are uncorrelated with each other and with the latent variables T_1, T_2 and the disturbance term E_1; d) E_1 is uncorrelated with T_1; e) the latent variables are scaled to have unit variance (to fix the scale of measurement of these variables). Subject to these assumptions providing a realistic account of the data, each of the model parameters has a clear substantive interpretation. These interpretations are as follows:

1) The coefficients G_i linking the observed test variables to the corresponding latent variables describe the regression relationships that exist between the observed variables and latent variables. If the test variables are measured in standardised form with a mean of zero and a unit variance the coefficients G represent the correlations between the non-observed latent variables and the observed test scores.

2) The coefficient B describes the regression relationship between the latent variable T_1 (true body lead level) and the latent variable T_2 (true cognitive ability), after errors of measurement in the observed indicators x, y have been taken into account. Since the latent trait variables are measured in standardised units the coefficient B also represents the correlation that exists between the latent trait variables.

Whilst the model in Figure 1 is clearly very simple, the general principles of formulating more complex LISREL models follow a similar logic in which a general theoretical model is expressed as a path diagram of relationships between variables and this path diagram is represented by a system of simultaneous linear equations.

Identification. While it is possible to specify a wide range of models using LISREL, it is not the case that all such models can be solved. For some model specifications the information in the observed variance/covariance matrix will not be sufficient to secure estimates of some or all model parameters. Such models are said to be under-identified and cannot be solved without either reformulating the model to overcome the sources of the identification problems or introducing additional information to solve the identification problem.

Models that are not under-identified are identified, which means that it is possible to estimate all model parameters on the basis of the observed variance/covariance matrix. Exactly identified models are those in which all model parameters are identified and the number of the model parameters is equal to the number of observed variances and covariances. Such models are not amenable to falsification through goodness of fit methods since the fitted model must exactly reproduce the observed variance/covariance matrix as a matter of mathematical necessity.

Models which are identified and have fewer parameters than observed variance and covariances are said to be over-identified. Over-identification is a desirable feature of a model since over-identified models do not fit the observed data as a matter of mathematical necessity and thus can be subject to test on the basis of goodness of fit measures.

Parameter Estimation. To solve a structural equation model requires that a set of model parameters representing the coefficients of a system of linear equations are estimated on the basis of the information contained in the variance/covariance or correlation matrix of a set of observed variables x, y. Over the years a range of methods have been evolved for solving such equations. Perhaps the most widely used method has been based on maximum likelihood estimation methods using iterative computer based methods (Joreskog & Sorbom, 1979). In these methods various numerical algorithms have been developed to estimate parameter values which maximise the likelihood of the parameters conditional on the properties of the observed matrix of variances and covariances. The maximum likelihood approach, however, assumes that the observed variables have a multivariate normal distribution and in many applications this assumption may be unrealistic. Whilst it has been found that parameter estimates based on maximum likelihood methods are generally robust to violation of the normality assumption these violations may influence both estimates of standard errors and model fit (Joreskog & Sorbom, 1989). For this reason a range of alternative methods of estimation to deal with different situations have been suggested. These methods include generalised least squares methods (Joreskog & Sorbom, 1989), methods for the analysis of dichotomous and polytomous variables (Muthen, 1984) and estimation methods which are distribution free (Browne, 1984). In cases where concern exists about the choice of estimation method it is generally prudent to fit the same model using a range of methods to determine the extent to which parameter values are sensitive to the choice of estimation method. However empirical experience suggests that in most cases the differences in parameter estimates yielded by different methods tend to be small.

Goodness of Fit. For over-identified models the problem of assessing the fit of the model to the observed data arises. There is no single test of goodness of fit that will prove to be adequate in all circumstances, but there are a number of approaches that may be used in assessing model fit.

Perhaps the most useful general test is through visual inspection of the residual variances/covariances obtained from taking the differences between the observed variances/covariances and the variances/covariances estimated from the model. If these residuals are small, relative to the original variances/covariances and are distributed in a random or haphazard way with respect to sign, this can be taken as good evidence of model fit. Joreskog & Sorbom (1989) provide a number of ways of examining the properties of such residuals to detect systematic deviations between the fitted model and the observed data.

The maximum likelihood method of estimation yields a formal goodness-of-fit test in the form of a log likelihood chi-square statistic. This statistic, which is analogous to the conventional chi-square goodness-of-fit statistic, provides a measure of the fit of the model to the observed data. However, this test is known to be sensitive to large sample size and departures from multivariate normality and for this reason is seldom used as a formal test of fit. More often the log likelihood chi-square is used for comparing the fit of competing models of the data (Joreskog & Sorbom, 1989).

Finally a range of indices have been suggested to measure the differences between the observed data and the fitted model. These indices are discussed by Joreskog & Sorbom (1989) and Bentler & Bonnett (1980) and may be used to assess various aspects of model fit.

A Worked Example

The previous sections of this chapter provide a general introduction to the logic, background and methodology of structural equation models. However many readers may find these general principles more accessible and theoretically meaningful by working through an example which shows the application of structural equation modelling methods to a specific problem. Below I present a worked example which has the advantage of illustrating many of the principles of model formulation, model specification, model fitting and testing but at the same time is restricted to a small number of variables so that the structure of the proposed models and assumptions can be examined in detail.

The example concerns a problem which has been of continuing interest in the area of childhood psychiatric epidemiology: the extent to which exposure to maternal depressive symptoms is associated with increased rates of problem behaviour in childhood. There have been a number of studies that have examined the association between maternal depressive symptoms and disruptive problem behaviours in childhood (for reviews of these studies see Downey &

Coyne, 1970; Gelfand & Teti, 1990; Rutter, 1990). In general, these studies have found that children exposed to maternal depression have increased rates of problem behaviour. These findings are, however, complicated by the presence of two problems relating to the measurement, interpretation and validity of measures of child behaviour. These problems are:

1) The validity and interpretation of behavioural report data: In all studies reported to date, the measurement of child behaviour has been based on report data provided by parents or teachers. These measures are likely to be subject to fallibility as measures of the child's general predisposition to engage in disruptive problem behaviours and this fallibility has been highlighted by the fact that it is well known that reports of the same children provided by different sources tend to be only modestly correlated. In a review of this issue, Achenbach, McConaughy & Howell (1987) report correlations between parent and teacher reports of child behaviour which suggest that these correlations are in the region of +.30. Such correlations imply substantial disagreements between reports of the same child described by different sources. In turn, the lack of strong cross-informant correlations raises complex issues about the interpretation, meaning and validity of child behaviour reports derived from a single source. (See also Achenbach in this volume).

2) Contamination of maternal report by maternal mental state: A closely related problem is that, in some studies, associations between maternal depression and child behaviour have been assessed on the basis of maternal report data (e.g. Billings & Moos, 1983; Cohler et al., 1983; Ghodsian et al., 1984). This raises the possibility that correlations between measures of depression and child behaviour may have been contaminated by the use of non-independent methods of measuring these variables. In particular, it could be suggested that maternal depression may colour the reporting of child behaviour leading to spurious correlations between maternally reported depressive symptoms and maternally reported child behaviour.

The analytic challenges posed by these possibilities are twofold: a) to devise methods for analysing and representing child behaviour reports in a way which reconciles differences in reports derived from different sources; b) to estimate the extent to which it is possible that maternal depression contaminates the reporting of child behaviour. Figure 2 presents two alternative models of the relationship between observed measures of maternal depression and child behaviour. Both models assume that maternal depression has been measured by two indicator variables x_1, x_2 which represent reported maternal depressive symptoms and that child behaviour has been described by reports from three sources x_3, x_4, x_5 (maternal report, teacher report and child report).

Model 1 (Figure 2a) is a baseline model which assumes that:

1) Maternal depression is a non-observed latent variable (T_1) which influences the fallible observed measures x_1, x_2. However the observed variables are also influenced by non-observed errors of measurement and other sources of disturbance represented by the variables U_1, U_2.

2) Child problem behaviour is a non-observed latent variable (T_2) which influences the observed report measures x_3, x_4, x_5. However the observed variables are also influenced by the disturbance terms U_3, U_4, U_5 which present sources of measurement error and test unreliability in the observed measures x_3, x_4, x_5.

3) The latent variables of maternal depression (T_1) and child behaviour (T_2) are correlated variables.

Model 2 (Figure 2b) extends the assumptions of Model 1 by assuming that in addition to the relationships implied by Model 1, maternal depression may influence maternal reports of child behaviour. Thus Model 1 represents a situation in which maternal depression and child behaviours are (or may be) correlated whereas Model 2 assumes that maternal depression may act to influence maternal reporting of child behaviour in addition to the assumptions of Model 1.

INSERT FIGURE 2 HERE

The conceptual models shown in Figure 2 can be presented as a series of simultaneous linear equations relating the observed variables x_1 to the non-observed latent variables (T_1, T_2) and the non-observed disturbance variables (U_i). These equations are shown below.

Model 1	**Model 2**
$x_1 = G_1 T_1 + U_1$	$x_1 = G_1 T_1 + U_1$
$x_2 = G_2 T_1 + U_2$	$x_2 = G_2 T_1 + U_2$
$x_3 = G_3 T_2 + U_3$	$x_3 = G_3 T_1 + G_4 T_2 + U_3$
$x_4 = G_4 T_2 + U_4$	$x_4 = G_5 T_2 + U_4$
$x_5 = G_5 T_2 + U_5$	$x_5 = G_6 T_2 + U_5$
$Cov (T_1, T_2) = .$	$Cov (T_1, T_2) = .$
$Var (T_1) = Var (T_2) = 1$	$Var (T_1) = Var (T_2) = 1$

Where the coefficients G are the factor loadings linking the observed variables x_1 ... x_5 to the latent variables of maternal depression (T_1) and child conduct problems (T_2) and . is the correlation between T_1, T_2. It can be seen that four of the five equations of both models are identical in form but that the models differ in the specification of the x_3 equation. In Model 1, the measure x_3 (maternal report of child behaviour) is assumed to reflect the child's behavioural tendencies but is not directly influenced by maternal depression whereas Model 2 assumes that maternal reports of child behaviour may be influenced by maternal depression independently of the general association between the latent variables of maternal depression and child behaviour.

To test the fit of these alternative models, data on maternal depression and child behavioural tendencies was gathered during the course of a longitudinal study of a birth cohort of New Zealand children (Fergusson, Lynskey & Horwood, in press). The data to be analysed is presented in Table 1. This Table reports the matrix of correlations of five measures observed when children were aged 12 years. These measures are:

1) Measures of maternal depression based on split half scores derived from the Levine Pilowsky Depression Inventory (Pilowsky & Boulton, 1970; Pilowsky, Levine & Boulton, 1968). The Levine Pilowsky Depression Inventory is a 37 item measure of depression. To construct two measures of maternal depression, the inventory was divided at random to produce two measures of maternal depression with each measure representing the number of depressive symptoms reported by the child's mother at age 13 years.

2) Measures of child conduct problems were obtained from reports from three sources: the child's mother, the child's class teacher and from self report. Maternal and teacher reports were obtained from measures which combined the items of the Rutter (Rutter, Tizard & Whitmore, 1970) and Conners (Conners, 1969; 1970) maternal and teacher questionnaires. Child report measures were based on an 11 item inventory based on DSM-III-R criteria (American Psychiatric Association, 1987) for oppositional defiant disorder. In all cases measures of child behaviour were scored using the number of problem behaviours reported by each source.

Inspection of Table 1 suggests the following general conclusions.

1) Measures of maternal depression were strongly correlated (r = +.901) suggesting that the measure of depression had high internal consistency.
2) Measures of childhood behaviour were only moderately correlated with correlations between different behaviour reports ranging from .359 to .440

suggesting quite substantial disagreements between reports of the same child described by different sources.

3) For all measures of child behaviour there was evidence of positive correlations between maternal depression and child behaviour implying a tendency for behaviour problems to increase with increases in maternal depression. The size of the correlations, however, varied with the source of reporting with the correlation between maternal depression and child behaviour ($r = .270, .278$) being substantially larger than the correlation between maternal depression and reports of child behaviour provided by children ($r = .054, .059$) or teachers ($r = .092, .097$).

INSERT TABLE 1 HERE

The models described previously were fitted to the data in Table 1 using the program LISREL 7 and methods of maximum likelihood estimation. The results of model fitting are summarised in Table 2 which shows two indices of fit for the models. The first measure is the log likelihood chi square statistic. This measures the general goodness of fit between the observed correlations and those implied by the fitted model parameters. The second measure is the overall goodness of fit index proposed by Joreskog & Sorbom (1989). This index ranges between 0 and 1 with 1 implying a perfectly fitting model.

Both sets of indices lead to a common conclusion about the adequacy of Models 1 and 2 as descriptions of the observed data. On the basis of both indices, Model 1 clearly fits the data poorly: the log likelihood chi square value is highly significant (42.16, d.f. = 4, $p<.001$) implying the presence of detectable deviations between the observed data and the data implied by model estimates. In addition, the overall goodness of fit measure is not high (GOF = .921). On the other hand, Model 2 shows a good fit to the data on the basis of both the log likelihood chi square value (0.58, d.f. = 3, $p>.90$) and the overall goodness of fit measure (GOF

= .999). Further it is possible to conduct a test of the improvement of fit provided by Model 2 when compared with Model 1. This test may be conducted by taking differences in the log likelihood chi square measures and the corresponding degrees of freedom and interpreting the differences in chi square values as a chi square variate. This test shows that Model 2 produced a clearly significant improvement in fit over Model 1 (LRX^2 = 41.58, d.f. = 1, p<.001).

Collectively the results in Table 2 lead to the conclusion that whilst the observed data were not consistent with the assumptions of Model 1 they were consistent with the assumptions of Model 2.

INSERT TABLE 2 HERE

The fitted Model 2 estimates are shown in Figure 3. This Figure may be interpreted as follows:

1) The coefficients relating the observed measures of depression to the latent variable of maternal depression are estimates of the correlations between the non-observed factor and the observed test scores. In this instance these coefficients are large (.94, .96) suggesting strong associations between maternal reported depression and the non-observed latent variable.

2) Similarly, the coefficients linking the observed behaviour report data to the latent variable of child conduct problems are estimates of the correlation between the latent variable and the observed indicators. In this case the coefficients (.56, .72, .61) suggest that there are generally quite moderate associations between the latent variable and the observed tests suggesting that the observed tests were subject to considerable fallibility as measures of the child's general tendencies to problem behaviours.

3) The coefficient relating maternal reported behaviour to maternal depression is an estimate of the extent to which maternal depression influences maternal reporting behaviour, independently of the general association between maternal depression and child behaviour. This coefficient was .22 ($p<.001$) suggesting a small but detectable tendency for increasing maternal depression to be associated with a tendency for mothers to over-report child behaviour problems relative to the latent criterion measure of child conduct problems.

4) Finally, the model shows that despite contamination in maternal reports of behaviour, there is evidence of a small positive correlation (. = +0.12, $p<.001$) between maternal depression and child behaviour. This correlation is somewhat smaller than the observed correlations between maternal depression and maternal reported behaviour (r = +0.270, +0.278). This may be explained by the fact that the correlations between maternally reported behaviour and maternal depression are inflated by a tendency for maternal depression to influence maternal reporting behaviours. At the same time this correlation is somewhat larger than the correlation between maternal depression and the observed teacher and child report measures. This may be explained by the fact that the model takes into account reporting errors in both maternal depression and child behaviour measures and these reporting errors tend to attenuate the observed correlations.

INSERT FIGURE 3 HERE

In general, the model in Figure 3 appears to provide a consistent and coherent account of the relationships between maternal depression and child behaviour reports which takes into account both the possibility of both errors of measurement in the observed report data and the possibility of contamination of maternally reported behaviour by maternal mental state. The data are consistent with a theory that assumes that whilst maternal depression and child behaviour are correlated variables, the correlation between maternal depression and maternal reported behaviour may over-estimate this association. At the same time the above analysis has considered only two of an array of possible models of the relationship between maternal depression and child behaviour reports. Whilst it has been shown that these data are consistent with a particular theory of the associations between variables, this does not imply that alternative models of this relationship cannot be proposed. In particular, in a review of the evidence on the association between maternal depression and child behaviour, Richters (1992) has argued that any bias apparent in maternal reports could equally well be due to depressed women reporting more accurately and to other sources being less accurate than depressed women. Given this it is clear that there is a need for further tests of the assumption that maternal depression colours maternal reports of child behaviour and leads to depressed women over-reporting childhood problem behaviour.

The Advantages and Liabilities of Structural Equation Models

The increased use of structural equation modelling methods in the social sciences has tended to produce a polarisation of opinion about the utility of these methods. In some areas, and notably behavioural genetics, the use of such methods has been greeted with considerable enthusiasm (see, for example, Heath, Neale, Hewitt, Eaves & Fulker, 1989) whereas in other areas there has been considerable skepticism expressed about these methods with concerns being expressed about the extent to which structural equation models adequately test causal assumptions (Baumrind, 1983) about the utility of latent variable methods as opposed to the analysis of measured variables (Martin, 1982), and about the extent to which formal methods of statistical modelling may both distort data analysis and inhibit the full exploration of data (Brown, Harris & Lemyre, 1991). All of these criticisms have some foundation but largely reflect the fact that structural equation models do not provide a complete solution to all problems of causal and measurement analysis. Rather, these methods are likely to be well applied in some conditions where the basic assumptions of both the modelling process and the statistical foundations of the model are well met and be less suited to the analysis of other types of problem. In the final analysis the utility of structural equation models does not depend on the formal statistical aspects of the model but rather on the skill and insight with which these models are applied to solve problems in a realistic and theoretically meaningful way.

The advantages of structural equation modelling are almost self-evident. Providing that the investigator has a well specified conceptual theory of his/ her data; providing that this theory can be realistically represented by a set of identified simultaneous linear equations and providing that data of sufficient quality and quantity exist to test the theory then structural equation modelling methods provide a powerful means of both hypothesis testing and theory generation. The liabilities of the approach arise from failure of one or more of these conditions to be satisfied.

Perhaps the least well documented aspect of model building is that of devising an adequate theoretical framework on which to base a model. Typically, theories in the social sciences are not expressed with sufficient precision for one to argue that a particular theoretical perspective implies that a particular structural model should describe the observed data. This situation almost invariably places the investigator in the situation of having to make various assumptions to bridge the gap between the conceptual theory and the statistical model which is alleged to represent the conceptual theory. Typically, such assumptions will include assumptions about the scale characteristics of variables, the nature of relationships between variables and, perhaps most importantly, the causal linkages which are assumed to be non-existent. The extent to which it is possible to justify such assumptions will depend on the state of theory and knowledge in a particular area and, in general, it seems likely that structural equation modelling methods are likely to be most powerful and effective in areas in which there is a large amount of prior theoretical and empirical knowledge about the processes under study and least likely to be effective in areas in which theory and data are sparse.

An important decision which must be made at an early stage of model construction concerns the metric on which variables are assumed to be measured. This issue is likely to be of particular importance in the area of childhood psychiatric epidemiology. In particular, there have been two very different ways in which behavioural variation has been measured in this area. In many studies, behavioural variation is measured by case/non-case measures in which the population of children is assumed to belong to distinct groups of disordered and non-disordered children (see, for example, McGee et al in this volume). In other studies founded on a psychometric tradition, behavioural variation is assumed to conform to a dimensional model in which the severity of disturbance ranges from none to severe. Whilst for purposes of data display and data description it is perfectly possible to represent the same set of symptom measures as scales or categories, at the point that causal or structural accounts of the data are presented some clear decision must be reached about the appropriate metric on which to measure behavioural variation.

In general, if it is believed that the population can be classified into disordered and non-disordered groups of children displaying qualitatively different patterns of behaviour then the application of structural equation modelling methods is likely to produce a distorted and probably misleading account of the data. Under such circumstances any attempts at modelling the causal structure of the data should be based on methods suitable for the analysis of qualitative or categorical measures. (The general latent class model developed by Goodman (1974a; 1974b) provides a body of theory for the analysis of qualitative variables which has many similarities and analogies with structural equation modelling of continuous variables). If, on the other hand, it is believed that behavioural variation has dimensional properties in which the severity of behavioural disturbance ranges from none to severe than structural equation modelling methods may be applicable to such data.

Even assuming that issues relating to the "scales versus categories" debate can be satisfactorily resolved, the application of structural equation models to dimensional variables may still pose problems. In particular, the central assumption of such models is that structural relationships between variables can be represented by an additive and linear model. However, it has been argued by a number of authors that additive and linear models may not be applicable in many areas and that models which assume non-linear or interactive structures are likely to be more realistic. To take one example, Brown & Harris (1978) have argued that the factors leading to the onset of depression combine interactively so that depression develops only in those who are both vulnerable to this condition and who are exposed to an appropriate provoking factor. It may be argued that the use of linear and additive models in situations where data structures are likely to be non-linear and interactive may be potentially misleading (Rutter, 1983).

At the same time there are a number of ways in which structural equation models can be extended to include some types of interactive assumptions. For example, a common interactive assumption is that different subpopulations respond to the same risk factor in different ways. The Brown and Harris theory above is an example of such an assumption since this model implies that the responses of vulnerable and non-vulnerable subjects to the same provoking factor may be different. Such models can be tested using so-called multiple group models in which the sample is partitioned into a number of different subgroups and different models are fitted to each group. A very useful account of the problems of modelling population heterogeneity in this way has been provided by Muthen (1989). However, in other applications it may be less easy to include interactive assumptions in structural equation models. All of these considerations clearly suggest that at the point when structural equation models of any data set are considered it is important to weigh the assumptions of linearity and additivity which underlie these models to determine the extent to which these assumptions are consistent with prior theory and knowledge about a given area.

The process of translating a causal theory to a causal model also requires that the investigator confronts the issue of the identification status of the model. It is not the case that all causal theories can be expressed as identified models. For example, it is often tempting to assume theories in which all variables are related to each other simultaneously. Such models will invariably be underidentified, and Duncan (1975) has described these models as "hopelessly underidentified" to underline the point that theory rather than data analysis is needed to identify the causal structure of variables. The effects of confronting the identification problem are usually to force the investigator to make certain simplifying assumptions which, in effect, impose the condition that certain causal relationships do not exist. The result of this process is to make structural equation models approximations, perhaps caricatures, of the complex reality they represent.

Almost invariably, attempts to build structural equation models will be compromised by the fact that the data available to test such theories are limited in both quantity and quality. The major quantitative limitation on data is usually that the model or process of data collection may omit some variables which are of theoretical relevance. The omitted-variable problem has been a fertile ground for research criticism since it is possible to take almost any causal model and postulate the presence of omitted variables that impugn the validity of the causal interpretation. Such criticism has both constructive and destructive aspects. The constructive aspect of the argument is that the suggestion that certain variables are omitted leads to the development of tests of the difference between the original model and the revised model containing the previously omitted variables. The destructive aspect of this approach is the belief (which is apparently held by many research critics) that the act of listing omitted variables automatically impugns the validity of a causal model.

Even when problems of model specification, identification, variable scaling and estimation have been overcome there may still be problems concerning the interpretation of parameters of structural equation models. Usually these models are fitted to data from reasonably large samples with the result that the parameter estimates describe the relationships that exist "on average" between variables. However, such models do not take account of inter-individual variations in responsiveness and implicitly assume that the parameters of the model apply with equal force to all individuals. However, it is possible that substantial inter-individual variation in responsivity may exist. Thus, for example, a dosage of lead which may be sufficient to cause serious neurological damage in one child may be less harmful in another child. This view presents an extreme case of the problem of interaction described earlier and assumes that the causal parameters of the model are specific to each subject and that, as a consequence, averaging these parameters across subjects does not adequately describe the causal process. Recently there have been attempts to address this problem through the use of so-called components of variance or hierarchical linear models (Byrk, 1989; Goldstein, 1987) which, subject to the availability of repeated measures on each subject, permit the estimation of both between and within individual parameters. When used in conjunction with structural equation modelling methods these methods may make it possible to assess the extent to which it is necessary to estimate individual specific parameters and the extent to which parameters averaged across individuals provide an adequate account of the causal and other processes under study.

Consideration of the problems of specifying, fitting, testing and interpreting structural equation models suggests that there are many points at which the process may fail. Existing theory may not be sufficiently precise to suggest compelling causal models; in the process of model specification and identification compromises may be made that vitiate the assumptions of the original theory; observed data may be of insufficient quality or quantity to sustain the model-fitting process; and the use of parameter estimates based on sample data rather than individual data may misrepresent the causal processes which are occurring in populations. These problems are, of course, not peculiar to structural equation models and any attempt to construct compelling causal models of correlational evidence needs to confront such problems of model specification, identification, estimation, measurement and interpretation. In the final analysis the major contribution of current structural equation modelling methods may not be in the area of developing substantive models but rather in the area of sensitising research workers to the theoretical, formal and empirical problems which must be confronted when causal accounts of correlational evidence are offered.

REFERENCES

Achenbach, T.M., McConaughy, S.M., & Howell, C.T. (1987). Child/adolescent behavioural and emotional problems: Implications of cross-informant correlations for situational specificity. **Psychological Bulletin, 101(2)**, 213-232.

Aigner, D.J., & Goldberger, A.S. (1977). **Latent variables in socio-economic models**. Amsterdam: North Holland.

American Psychiatric Association (1987). **Diagnostic and statistical manual of mental disorders (3rd ed., rev.).** Washington, DC: Author.

Baumrind, D. (1983). Specious causal attributions in the social sciences: The reformulated stepping-stone theory of heroin use as exemplar. **Journal of Personality and Social Psychology, 45**, 1289-1298.

Bentler, P.M., & Bonnett, D.G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. **Psychological Bulletin, 88,** 588-606.

Billings, A.G., Moos, R.H. (1983). Comparisons of children of depressed and non-depressed parents: a social-environmental perspective. **Journal of Abnormal Child Psychology, 11**, 483-486.

Brown, G.W., & Harris, T. (1978). **The social origins of depression**. London: Tavistock.

Brown, G.W., Harris, T.O., & Lemyre, L. (1991). Now you see it, now you don't - some considerations on multiple regression. In D. Magnusson, L.R. Bergman, G. Rudinger & B. Törestad (Eds.). **Problems and methods in longitudinal research: Stability and Change**. Cambridge: Cambridge University Press, pp. 67-94.

Browne, M.W. (1982). Covariance structures. In Hawkins, D.M. (Ed.). **Topics in applied multivariate analysis**. Cambridge: Cambridge University Press (pp. 72-141).

Browne, M.W. (1984). Asymptotically distribution-free methods for the analysis of covariance structures. **British Journal of Mathematical and Statistical Psychology, 37**, 62-83.

Byrk, A.S. (1989). **An introduction to HLM: Computer program and users' guide**. Chichago: University of Chicago.

Cohler, B.J., Gallant, D.H., Grunebaum, H., & Kaufman, C. (1983). Social adjustment among schizophrenic, depressed, and well mothers and their school-aged children. In Morrison, H.L. (Ed.). **Children of depressed parents**. New York: Grune & Stratton, pp 65-97.

Conners, C.K. (1969). A teacher ratings scale for use in drug studies with children. **American Journal of Psychiatry, 126**, 884-888.

Conners, C.K. (1970). Symptom patterns in hyperkinetic, neurotic and normal children. **Child Development, 41**, 667-682.

Downey, G., & Coyne, J.C. (1990). Children of depressed parents: An integrative review. **Psychological Bulletin, 108**, 50-76.

Duncan, O.D. (1975). **Introduction to structural equation models**. New York: Academic Press.

Everitt, B.S. (1984). **An introduction to latent variable models**. London: Chapman and Hall.

Fergusson, D.M., Lynskey, M.T., & Horwood, L.J. (In press). The effect of maternal depression on maternal ratings of child behavior. **Journal of Abnormal Child Psychology**.

Gelfand, D.M., & Teti, D.M. (1990). The effects of maternal depression on children. **Clinical Psychology Review, 10,** 329-353.

Ghodsian, M., Zajicek, E., Wolkind, S. (1984). A longitudinal study of maternal depression and child behaviour problems. **Journal of Child Psychology and Psychiatry, 25**, 91-109.

Goldstein, H. (1987). **Multilevel models in educational and social research**. London: Oxford University Press.

Goodman, L.A. (1974a). The analysis of systems of qualitative variables when some variables are unobservable. Part 1: A modified latent structure approach. **American Journal of Sociology, 79,** 1179-1259.

Goodman, L.A. (1974b). Exploratory latent structure analysis using both identifiable and unidentifiable models. **Biometrika, 61,** 215-231.

Heath, A.C., Neale, M.C., Hewitt, J.K., Eaves, L.J., & Fulker, D.W. (1989). Testing structural equation models for twin data using LISREL. **Behavior Genetics, 19**, 9-35.

Heise, D.R. (1975). **Causal analysis**. New York: Wiley.

Joreskog, K.G. (1973). A general method for estimating a linear structural equation system. In Goldberger, A.S., and Duncan, O.D. (Eds.). **Structural equation models in the social sciences**. New York: Seminar, pp. 85-112.

Joreskog, K.G., & Sorbom, D. (1979). **Advances in factor analysis and structural equation models**. Cambridge, MA: Abt Books.

Joreskog, K.G., & Sorbom, D. (1989). **LISREL 7: User's reference guide**. Mooresville, IN: Scientific Software.

Joreskog, K.G., & Wold, H. (Eds.). (1982). **Systems under indirect observation: causality, structure, prediction**. Amsterdam: North Holland.

Kenny, D.A. (1979). **Correlation and causality**. New York: Wiley.

Long, J.S. (1983a). **Confirmatory factor analysis**. Beverly Hills: Sage.

Long, J.S. (1983b). **Covariance structure models: an introduction to LISREL**. Beverly Hills: Sage

Martin, J.A. (1982). Application of structural modeling with latent variables to adolescent drug use: A reply to Huba, Wingard, and Bentler. **Journal of Personality and Social Psychology, 43**, 598-603.

McDonald, R.P. (1978). A simple comprehensive model for the analysis of covariance structures. **British Journal of Mathematical and Statistical Psychology, 31**, 59-72.

Muthen, B. (1984). A general structural equation model with dichotomous, ordered categorical and continuous latent variable indicators. **Psychometrika, 49**, 115-132.

Muthen, B.O. (1989). Latent variable modelling in heterogenous populations. **Psychometrika, 54**, 557-585.

Pilowsky, I., & Boulton, D.M. (1970). Development of a questionnaire-based decision rule for classifying depressed patients. **British Journal of Psychiatry, 116**, 647.

Pilowsky, I., Levine, S., & Boulton, D.M. (1969). The classification of depression by numerical taxonomy. **British Journal of Psychiatry, 115**, 937-945.

Richters, J.E. (1992). Depressed mothers as informants about their children: A critical review of the evidence for distortion. **Psychological Bulletin, 112(3)**, 485-499.

Rutter, M. (1983). Statistical and personal interactions: facets and perspectives. In Magnusson, D., and Allen V. (Eds.). **Human development: an interactional perspective**. New York: Academic Press, pp. 295-319.

Rutter, M. (1990). Commentary: Some focus and process considerations regarding effects of parental depression on children. **Developmental Psychology, 26**, 60-67.

Rutter, M., Tizard, J., & Whitmore, K. (1970). **Education, health and behaviour**. London: Longman.

Table 1: Matrix of correlations of maternal depression and measures of child conduct disorder

at 12 years

Conduct Disorder	**Maternal Depression**				
	First Measure	Second Measure	Maternal Rating	Teacher Rating	Child Rating
MATERNAL DEPRESSION					
First Measure	1				
Second Measure	.901	1			
CONDUCT DISORDER					
Maternal Rating	.270	.278	1		
Teacher Rating	.092	.097	.423	1	
Child Rating	.054+	.059+	.359	.440	1

N = 768

+ Correlation not significantly different from zero (p>.05)

Table 2: Log likelihood chi-square goodness of fit and adjusted goodness of fit

index for models 1 and 2

Model	Log-likelihood Chi-square	Degrees of Freedom	Probability	Adjusted Goodness of Fit
One	42.16	4	<.001	.921
Two	0.58	3	>.90	.999