Soc Psychiatry (1986) 21: 53-62

Social Psychiatry

The effects of test reliability on relationships between measures of life events and depression

D. M. Fergusson and L. J. Horwood

Department of Paediatrics, Christchurch Clinical School of Medicine, Christchurch, New Zealand

Summary. The paper develops a measurement model for assessing the reliabilities of life event inventories. This method is applied to life event data collected in a longitudinal study of a sample of over 1100 New Zealand children and their families. The findings suggest that the life event inventory was of low-to-modest reliability, with reliability estimates ranging from 0.67 to 0.72. The reliability estimates were applied to the correlations between life event and depression measures collected at two time periods spaced 1 year apart to obtain estimates of the correlations between measures corrected for the effects of test reliability. On the basis of observed and corrected correlations, a non-recursive path model was fitted to the data to examine the effects of measurement error on the coefficients of the model. This analysis showed the presence of quite marked differences in the coefficients of the fitted model. In particular, the model solved assuming that the data were of perfect reliability suggested the presence of reciprocal causation between life event and depression measures, whereas when test reliability was taken into account, the model suggested a uni-directional pattern of causation, in which life event measures influenced depression measures. Nonetheless, both models showed that the predominant direction of association was from life events to depression rather than vice versa. Various theoretical implications of these findings are discussed.

Introduction

In recent years there have been a large number of studies of the relationship between life events and depressive illness or symptoms in women (Paykel et al. 1969; Thomson and Hendrie 1972; Cooper and Sylph 1973; Gersten et al. 1977; Horowitz et al. 1977; Hornstra and Klassen 1977; Brown and Harris 1978; Warheit 1979; Andrews 1981; Brown and Prudo 1981; Benjaminsen 1981; Finlay-Jones and Brown 1981; Henderson et al. 1981; Stewart and Salt 1981; Fergusson and Horwood 1984). As a general rule, it has been found that increasing reports of life events are associated with increased rates of depressive illness and/or symptoms. This persistent correlation between life event measures and depression has led, naturally, to the assumption that adverse life events may precipitate or provoke depression, and this assumption has been incorporated into a number of theoretical models (Brown and Harris 1978; Henderson et al. 1981). At the same time, there have been a number of critiques of the view that there is a substantial causal association between life event measures and depression. These criticisms have been most clearly summarised by Tennant et al. (1981) who outlined a number of problems in interpreting the correlation.

Amongst the criticisms that were raised about life event research was that life event inventories tend to be of low-to modest reliability and that life event measures explain a relatively small amount of the variation in depression. Tennant et al. cite reliabilities which range from 0.50 to 0.75; the amount of variance in rates of illness explained by life events has typically been found to be less than 10% (Rabkin and Struening 1976; Tennant et al. 1981). However, it is possible that these findings are closely related since the effects of low test reliability are to impart a downward bias to the observed correlations between life events and illness (Guilford and Fruchter 1973). In view of this, it is possible that the small amounts of variance explained may merely reflect the fact that life event schedules have low reliability and, consequently, that the true correlations between life event and depression measures corrected for the effects of measurement error may be substantially larger than the observed correlations.

A second major criticism has concerned the direction of causality between life event and depression measures (Brown and Harris 1978; Tennant et al. 1981). While it has commonly been assumed that the direction of causality is such that life events tend to provoke depression, it is possible that the relationship operates in the opposite direction as a result of a reporting bias amongst depressed women. In particular, if depressed women tend to report more life events simply as a consequence of their mental state, then the apparent correlations between life events and depression may reflect nothing more profound than that a high rate of life event reporting is symptomatic of depressive illness. This issue was explored in a previous paper (Fergusson and Horwood 1984) in which we developed a structural equation model designed to examine a possible reciprocal relationship between life event measures and depression measures. This analysis suggested the possibility of a reciprocal association but, nonetheless, the predominant direction of causality was from life event measures to depression measures rather than vice versa. However, the model was solved on the basis of the observed correlations between life event and depression measures and thus implicitly assumed that the observed data were error free. In view of the suspect reliability of life event measures, this assumption may be unrealistic, as the observed correlations were likely to have provided underestimates of the true correlations between the measures. Furthermore, while the elements of the correlation matrix between the observed variables may have had a downward bias, the same conditions may not have applied to the coefficients of a structural equation model since the effects of measurement error may influence these coefficients in complex ways.

This aims of this paper are two-fold: firstly, to develop a method for measuring the reliability of a particular life event inventory and, secondly, to apply these reliability estimates to the observed correlations reported in the previous analysis and thus re-estimate the proposed model, taking into account the effects of measurement error. The theoretical basis of the reliability model is developed below.

The measurement of reliability for life event inventories

In the situations where life event inventories comprise a series of n items $z_1, z_2 \ldots z_n$ to which the subject makes responses, the life event score is typically estimated by a weighted linear composite of the items:

$$x = \sum w_i z_i$$

Where x is the estimated life event score for the subject and w_i is the weight attached to the ith life event item. However since x, the observed life event score, is made up of a sum of fallible indicator measures which are subject to error, it is not free from error. If it is assumed that the sources of error influencing x are random, then the *measurement model* linking the observed score x to the true but non-observed score X is:

$$x = X + u$$

Where u is a random error variable having the property that $E(u) = 0$; $\text{Cov}(Xu) = 0$.

From the above, it is easy to show that estimates of the reliability of the life event measure may be obtained from either administering the inventory to the same subject on two separate occasions or administering the inventory to both the respondent and a separate informant. However, both estimates rely on strong assumptions involving the independence of error terms and the stability of the true score X:

1. For the test/retest procedure, the spacing of the tests must be such that there is very little or no change in the true life event scores as a result of intertemporal variation; in addition, the sources of error on the measurement must be uncorrelated over time. In practice these criteria tend to conflict since short spacing of testing times minimises the effects of intertemporal variation but increases the possibility of correlated errors of measurement as a result of carryover effects between testing sessions.

2. For the informant/respondent method, it must be assumed that both respondent and informant have the same true score values (or at least these scores are exactly linearly related) and that the sources of error influencing their reporting behaviour are uncorrelated.

At first sight, it would seem that measures of reliability based on the internal structure of the inventory (i.e. split half or alpha-coefficients) may overcome these problems. However, it may be shown that these methods are not applicable to life event scores. The model underlying both the split half and coefficient alpha-measures assumes that each observed item z_i is a fallible estimate of a common underlying trait measured by the items. This is clearly incorrect for life event items which do not measure a common underlying trait but rather are a set of heterogeneous items grouped together because it is assumed that they have a common effect rather than arising from a common source. The inapplicability of split half-and alpha-coefficients can be seen from the following hypothetical situation. Consider a population in which life events occur to individuals at random but they are reported with perfect accuracy. Under these circumstances, the reported life event score is perfectly reliable and the true score is equal to the observed score. However, because the life event items are uncorrelated, both the split half and coefficient alpha-values will have expected values of zero, implying a test with no reliability whatsoever. This situation is not a paradox but arises because tests of internal consistency are not applicable to inventories of heterogeneous items (Guilford and Fruchter 1973).

An alternative method of estimation has been suggested by the work of Heise (1969) and Wiley and Wiley (1970). This method is applicable when life event scores have been measured on three occasions, even though the spacing of the measurements may be such as to preclude the application of the test/retest method. The model is based on two sets of equations:

1. The measurement model which describes the linkages between the observed life event score x_i at time i ($i=1, 2, 3$) and the corresponding non-observed true score X_i at time i.

2. The structural equation model which describes the linkages between the true scores X_i across time.

The measurement model is:

$$x_i = X_i + u_i$$
$$\mathrm{Cov}(X_i u_i) = 0;\ \mathrm{Cov}(u_i u_j) = 0\ (i \neq j)$$

The structural equation model for a three wave model is:

$$X_1 = e_1$$
$$X_2 = B_{21} X_1 + e_2$$
$$X_3 = B_{32} X_2 + e_3$$
$$\mathrm{Cov}(e_i, e_j) = 0\ (i \neq j)$$

The path diagram corresponding to the model is shown in Fig. 1. In its general form, the model is underidentified requiring the estimation of eight unknowns on the basis of six observed variances and covariances. However, this problem can be overcome by introducing the assumption that the variances of the disturbance terms u_i in the measurement model are constant (Wiley and Wiley 1970). Alternatively, it may be assumed that the test reliability is constant over time (Heise 1969). From the estimated coefficients of the model, it is possible to obtain estimates of the variances of the u_i's. Using these, the reliability of each observed score x_i may be estimated from:

$$r_{ii} = \frac{\sigma^2_{x_i} - \sigma^2_{u_i}}{\sigma^2_{x_i}}$$

Where r_{ii} is the test reliability, $\sigma^2_{x_i}$ is the variance of the observed score and $\sigma^2_{u_i}$ is the estimated error variance.

The remainder of this paper explores the application of this method of reliability estimation to data collected during the course of a longitudinal study of a birth cohort of New Zealand children and examines the effects of test reliability of the structure of associations between life event and depression measures.

Method

The data were collected during the sixth, seventh and eighth stages of the Christchurch Child Development Study. In this study, a birth cohort of 1265 children and their families was studied at birth, at 4 months and at annual intervals to the age of 6 years, using a combination of an interview with the child's mother supplemented by other documentary sources. The methods of data collection and quality control have been described in previous papers (Fergusson et al. 1981; Fergusson and Horwood 1984).

As part of the data collection, information on life event measures and depression measures were collected as follows:

1. Life event measures were obtained at 4, 5 and 6 years, using a structured 20-item schedule based on

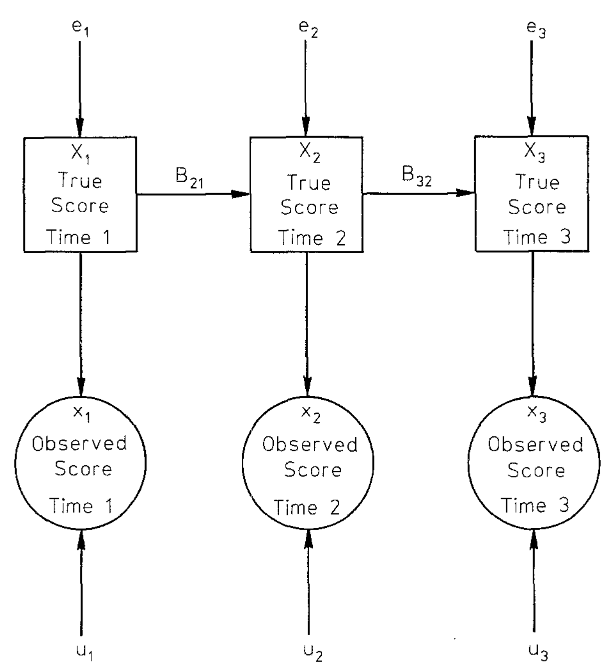

Fig. 1. Three-wave model of observed and true life event scores

Table 1. Matrix of intercorrelations between measures of life events taken at three times

	Age (years)		
	4	5	6
4 years	1	0.46	0.29
5 years		1	0.43
6 years			1

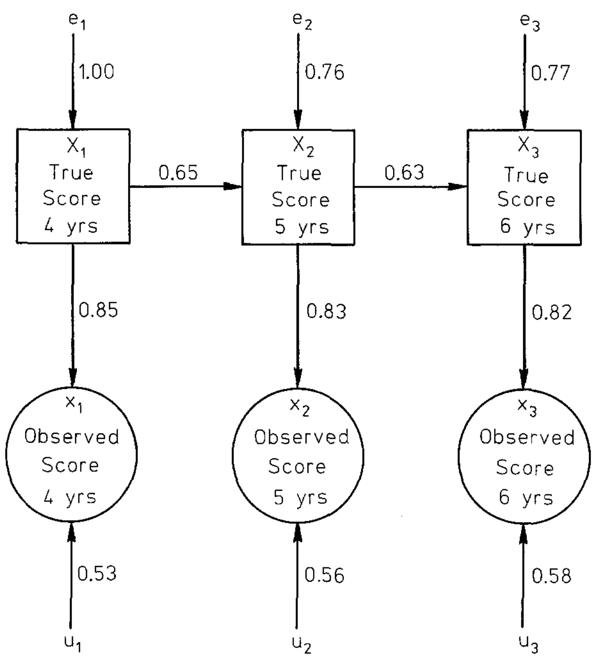

Fig. 2. Fitted reliability model for life events

the Holmes and Rahe (1967) Social Readjustment Rating Scale. The 20 items in the scale have been described in a previous paper (Fergusson and Horwood 1984) and related to common areas which involved stress or readjustment: financial problems, marital problems, family discord, ill health, bereavement, involvement with the law, pregnancy, changes in employment and residential change. For each year, a life event score was constructed using an unweighted sum of the 20 items. A simple unweighted sum was used as previous analysis (Fergusson and Horwood 1984), showed that various methods of weighting items, using both the weights described by Masuda and Holmes (1967) and empirical weights obtained from multiple regression analysis, did not significantly improve prediction or explanation of the depression measure. In addition, an unweighted sum provided a more robust and easy to interpret measure.

2. The extent of depressive symptoms displayed by the mother was based on a modified version of the Levine/Pilowsky depression questionnaire (Pilowsky et al. 1969, Pilowsky and Boulton 1970), which was administered to the mother when her child was aged 5 years and 6 years old. The questionnaire was scored using a simple unweighted sum of the number of depressive symptoms reported by the woman. This method of scoring was adopted since both factor analysis and clustering methods suggest that the questionnaire measured a single underlying dimension relating to the severity of the depressive symptoms reported by the woman. An unweighted sum was used as it was found that this sum was highly correlated with a weighted sum, using the least-squares estimates of the score on a single factor model. Reliability of the measure was obtained from coefficient alpha-estimates and this suggested that the test was highly internally consistent. Coefficient alpha (Cronbach 1951) was 0.91 for the 5-year measure and 0.93 for the six-year measure.

Sample sizes

The analysis is based upon a sample of 1103 women for whom data on both depressive symptoms and life events were available when the child was aged 5 and 6 years. This sample represented 87% of the 1265 women whose children entered the research and 95% of those women whose child was still alive and resident in New Zealand.

Results

Estimates of the reliability of life event measures

Table 1 shows the matrix of intercorrelations of the three life event measures taken at 4, 5 and 6 years. Application of the estimation methods described by Wiley and Wiley (1970) to the results in Table 1 yielded the path diagram shown in Fig. 2. In this diagram all coefficients are expressed in standardised form. The following inferences may be drawn from the analysis.

1. The standardised coefficients linking the observed scores x_i to the corresponding true scores X_i are the square roots of the test reliabilities (under the assumptions of the model). The solved model shows that the reliability of the life event measure at 4 years was 0.72, at 5 years 0.69 and at 6 years 0.67. The declining test reliability with time reflects the fact that the variance of the life event score showed a slight tendency to reduce over time. The results thus suggest that the reliability of the measure was in the region of 0.65 to 0.70, and this result appears to be consistent with previous reliability estimates which have ranged from 0.50 to 0.75 (see Tennant et al. 1981).

Table 2. Matrix of observed and disattenuated correlations. (Observed correlations given in the lower segment of the matrix, reliabilities in the leading diagonal)

		Time 1		Time 2	
		Life events	Depression	Life events	Depression
Time 1	Life events	0.69	0.49	0.63	0.40
	Depression	0.39	0.91	0.29	0.52
Time 2	Life events	0.43	0.23	0.67	0.44
	Depression	0.32	0.48	0.35	0.93

The alternative method of estimation suggested by Heise (1969) gave a reliability estimate of 0.69 for all three measures.

2. The true life event scores show substantial autocorrelations over time and for each year, the coefficient linking the life event score to the score in the preceding year was in excess of 0.60. This suggests that true life event scores had considerable stability over time, implying that reporting a large number of life events in 1 year was associated with high reports in the next. Given the life events studied (i.e. financial problems, ill health, marital problems, etc.), these high-stability coefficients are not unexpected since many of the life events described relate to situations which are likely to be ongoing and chronic rather than short-term acute events.

The relationship between life events and depression over time

The results of the reliability analysis suggest the presence of substantial measurement error in the life event scores, with the result that the observed correlations involving these scores may have a marked downward bias. The extent of this bias may be seen from Table 2, which gives the matrix of disattenuated correlations between the life event measures and depression measures at two time intervals. The disattenuated correlation between two variables X_i and X_j is defined by:

$$r_{X_iX_j} = r_{x_ix_j} / (r_{ii})^{\frac{1}{2}} (r_{jj})^{\frac{1}{2}}$$

Where $r_{X_iX_j}$ is the estimated correlation between the true scores X_i, X_j and $r_{x_ix_j}$ is the observed correlation, and r_{ii}, R_{jj} are the respective test reliabilities. The disattenuated correlation is an estimate of the true correlation between the observed scores purged for the effects of measurement error. In computing Table 2, the reliability estimates for the life event measure were derived from the analysis given above and the test reliabilities for the depression measure were the coefficient alpha values. These reliabilities are shown in the leading diagonal of the matrix. The entries above the leading diagonal give the disattenuated correlations and those below the diagonal, the observed correlations. It is apparent that the effects of disattenuation are to increase the correlations quite substantially, and this is most marked for correlations involving life event measures. Thus while the observed correlation between life events at time 1 and time 2 was 0.43, the corresponding disattenuated correlation is 0.63. Similarly, the observed correlations between life events and depression increased from 0.35:0.39 to 0.44:0.49, following adjustment for measurement error. More generally, the matrix clearly shows that the effects of the modest reliability of the life event measures is to produce a marked downward bias in the observed correlations involving life event measures.

In a previous paper (Fergusson and Horwood 1984), we proposed that the structure of the data in Table 2 could be described by a two-equation nonrecursive model specified by:

$$x_3 = p_{31}x_1 + p_{34}x_4 + p_{e3}E_3$$
$$x_4 = p_{42}x_2 + p_{43}x_3 + p_{e4}E_4$$
$$\mathrm{Cov}(x_iE_j) = 0 \; (i = 1, 2; j = 3, 4)$$

Where x_1, x_3 are the life event measures taken at 5 and 6 years, respectively; x_2, x_4 are the depression measures taken at 5 and 6 years. All variables x_i and disturbances E_i are assumed to be in standardised form. It is assumed that the exogenous variables x_1, x_2 are uncorrelated with the disturbances E_3, E_4 but the disturbance terms E_3, E_4 may be correlated. The model is just identified and the coefficients may be estimated by indirect least-squares methods (Fergusson and Horwood 1984). Figure 3 shows the model solved on the data assuming that the observed variables were errorless (i.e. were of perfect reliability). However, as the reliability analysis shows, the assumption of perfect reliability is quite unrealistic and in fact, as may be seen from Table 2, the estimated correlations between the true scores differ quite markedly from those which exist between the observed scores. Thus, to take account of measurement error, it is necessary to elaborate the model by making explicit acknowledgement of the fact that the observed variables are fallible estimates of the true but non-observed variables. This may be done by expanding the original model above to include a set of measurement equations that specify the relationship between the observed scores and the true scores. The model is:

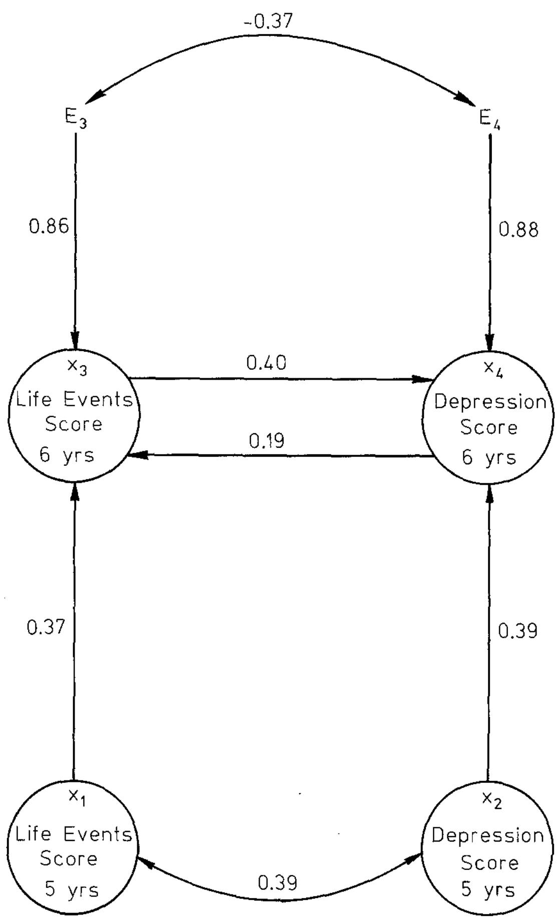

Fig. 3. Fitted non-recursive model of life events and depression assuming observed variables are errorless

The measurement model

$x_1 = a_{11}X_1 + u_1$
$x_2 = a_{22}X_2 + u_2$
$x_3 = a_{33}X_3 + u_3$
$x_4 = a_{44}X_4 + u_4$
$\mathrm{Cov}(u_i u_j) = 0$; $\mathrm{Cov}(X_i u_j) = 0$ for $i \neq j$

In this model both the observed score x_i and the corresponding true score X_i are assumed to be standardised variables with unit variance, the coefficient a_{ii} denotes the square root of the test reliability (as in Fig. 2) and u_i is an error term which is uncorrelated with X_i and uncorrelated across measuring periods.

The structural equation model

$X_3 = p_{31}X_1 + p_{34}X_4 + p_{e3}E_3$
$X_4 = p_{42}X_2 + p_{43}X_3 + p_{e4}E_4$
$\mathrm{Cov}(X_i E_j) = 0$ ($i = 1, 2; j = 3, 4$)

The model is of the same form as the original model but differs in that the variables X_i are the true scores rather than the observed scores.

As it stands, the model is underidentified since it involves the estimation of ten unknowns on the basis of six observed correlations. However, the model may be identified if estimates of the reliabilities of the tests are available and these are given in Table 2. Assuming that test reliabilities are fixed for the model, there are two equivalent methods of solving the equations. The first is to apply the method of indirect least squares estimation to the matrix of disattenuated correlations to estimate the coefficients of the structural equation model. The alternative method is to apply Joreskog and Sorbom's LISREL model (Joreskog 1973; Joreskog & Sorbom 1976) to the data, using the reliability estimates to fix the size of the variance of the disturbance parameters u_1, u_2, u_3, u_4. In practice, both methods will yield the same solution since the model is exactly identified and the diagram below has been solved using LISREL.

The fitted model is shown in Fig. 4. To aid in the interpretation of the results, all coefficients for both the measurement model and the structural equation model are expressed in standardised form.

Comparison of the model in Fig. 4 with the previous model in Fig. 3, which assumed that the data were errorless shows marked differences between the findings:

1. The most obvious difference is that coefficients involving the variables X_1 (life events at time 1) and X_2 (depression at time 1) are larger in the model solved assuming the data were contaminated by measurement error than for the model in Fig. 3. This is a direct consequence of the fact that measurement error imparts a downward bias to the observed correlations.

2. The properties of the reciprocal path between life events at time 2 and depression at time 2 differ between the two models. Both models show that the predominant direction of causation is from life events to depression rather than vice versa, but they differ on both the size and sign of the feed-back loop from depression to life events. In the model solved assuming errorless data, there is a small positive feed-back coefficient ($p_{34} = +0.19$), implying that increased depression was associated with an increased tendency to report life events. In the model solved assuming the presence of measurement error, the coefficient is very small and negative ($p_{34} = -0.03$). Furthermore, the small negative coefficient is not significantly different from zero, suggesting that for the model in Fig. 4, there were no feed-back effects from depression to life events.

3. The properties of the disturbance terms for both models have altered. In the model in Fig. 3, there is a substantial negative correlation between the disturbance terms E_3, E_4 ($r_{E_3E_4} = -0.37$). This

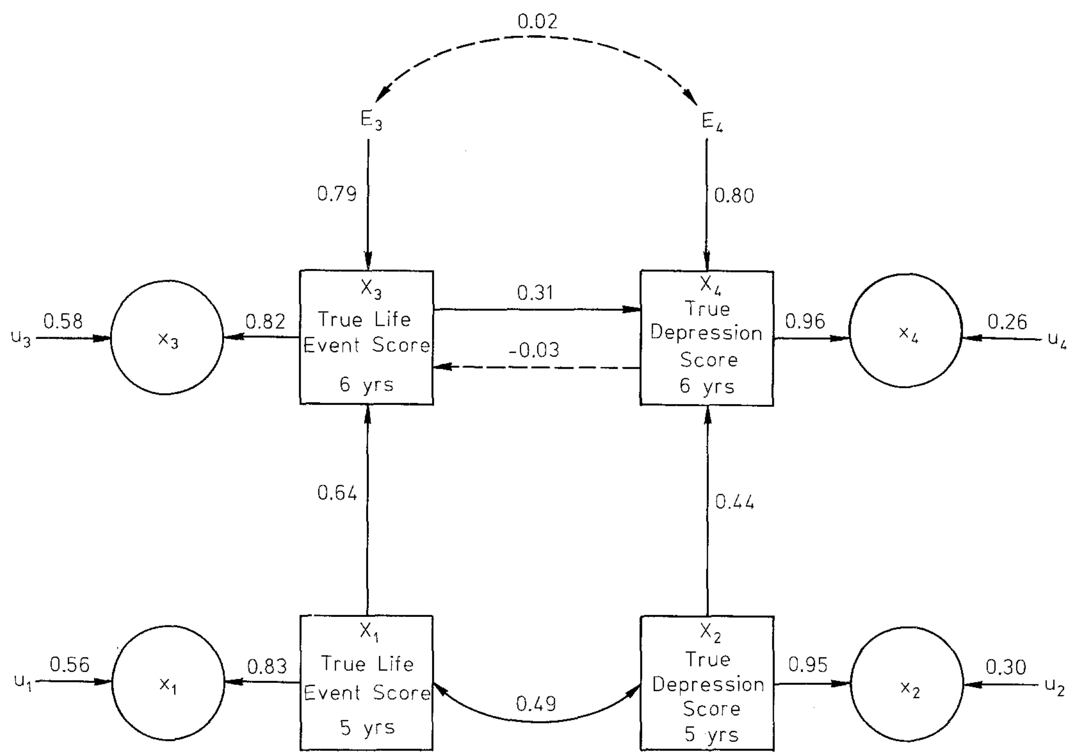

Fig. 4. Fitted non-recursive model assuming the presence of measurement error. (All variables and disturbance terms are measured in standard score units)

negative correlation implies that the factors influencing susceptibility to life events are negatively correlated with the factors influencing susceptibility to depressive symptoms (once the effects of previous life events and depressive symptoms have been taken into account). In the model solved assuming measurement error, the correlation is small and not significant ($r_{E_3E_4} = +0.02$), suggesting that there is no correlation between the disturbances. This absence of correlation is, of course, consistent with the finding in that there is no feed-back from life events to depression.

The accumulated findings of the fitted model in Figure 4 suggest that two of the coefficients of the model are, in fact, very close to zero and as a consequence the structural equation model may be rewritten as an overidentified model:

$$X_3 = p_{31}X_1 + p_{e3}E_3$$
$$X_4 = p_{42}X_2 + p_{43}X_3 + p_{e3}E_4$$
$$\text{Cov}(E_3E_4) = 0;\ \text{Cov}(X_iE_j) = 0\ (i = 1, 2; j = 3, 4)$$

Where the scores X_i are the true score measures. The presence of the overidentifying restrictions ($p_{34} = 0$; Cov (E_3, E_4) = 0) makes it possible to extend the model to estimate further parameters. In particular, consider the 'full' model of the data.

$$X_3 = p_{31}X_1 + p_{32}X_2 + p_{34}X_4 + p_{e3}E_3$$
$$X_4 = p_{41}X_1 + p_{42}X_2 + p_{43}X_3 + p_{e4}E_4$$
$$\text{Cov}(X_iE_j) = 0\ (i = 1, 2; j = 3, 4)$$

This model extends the previous non-recursive models by permitting the possibility that there are coefficients (p_{32}, p_{41}) across the 'diagonals' of the model: that is, life events at time 1 may influence depression at time 2 and depression at time 1 may influence life events at time 2. In its general form, the model is underidentified, requiring the estimation of eight model parameters on the basis of six observed covariances (assuming the measurement model is fixed). However, from previous models, it is known that $p_{34} = 0$; Cov(E_3E_4) = 0. By setting these parameters to zero, the model can be shown to be identified and takes the form shown in Fig. 5, which essentially gives a recursive model with two correlated variables X_1, X_2. The solved model shows that in fact the diagonal path coefficients (p_{41}, p_{32}) of the model are also very small and statistically non-significant and this confirms the findings of a previous paper (Fergusson and Horwood 1984). This model suggests the following conclusions about the structure of the variables over time:

1. Both the life event and depression measures show substantial autocorrelations over time, implying that

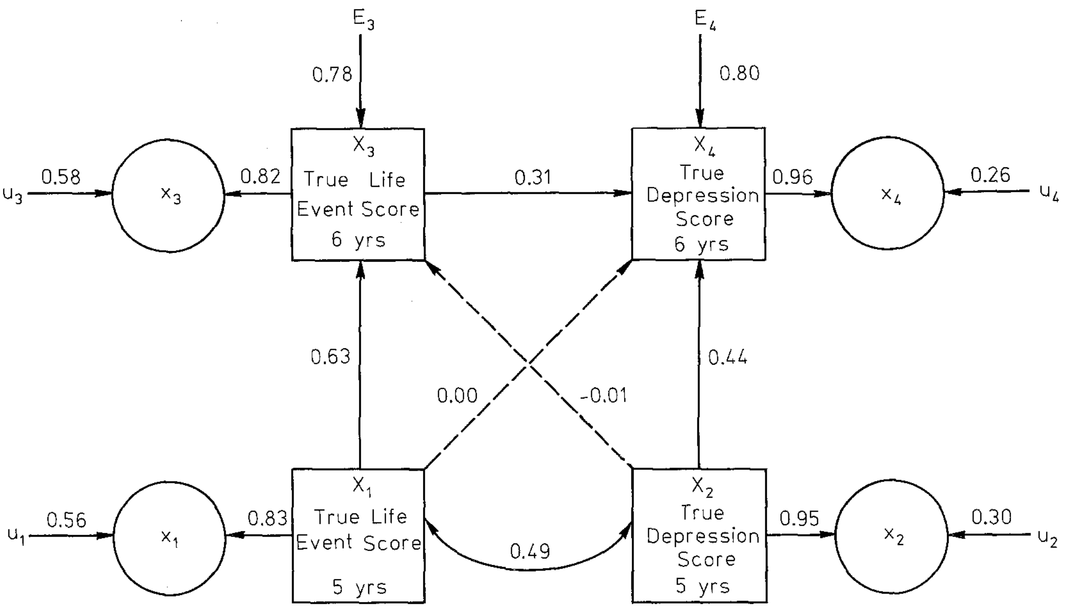

Fig. 5. Fitted model assuming: (a) known test reliabilities; (b) absence of reciprocal causation; (c) uncorrelated disturbances (E_3, E_4). (All variables and disturbance terms are measured in standard score units)

high life event or depression scores at one time were associated with high scores at a later time.

2. The correlation between life events and depression reflects a unidirectional causal relationship from life events to depression in which there is little or no reciprocal causation from depression to life events.

3. When the previous history of life events and depression are taken into account, the disturbance terms of the model are uncorrelated.

Discussion

The reliability analysis presented in this paper conforms previous findings which suggest that life event inventories based on the Holmes and Rahe Social Readjustment Rating Scale are of low-to-modest reliability. The estimates obtained in this study ranged from 0.66 to 0.72, and these estimates appear to be consistent with the range of 0.50 to 0.75 quoted by Tennant et al. (1981). The relatively low reliability of life event measures serves to clarify some of the criticisms that have been made about the small amounts of variance explained by life event scores. Specifically, the effect of low test reliability is to impart a downward bias to correlations between life event and depression measures. The extent of this bias is evident from this analysis: while the observed correlations between life events and depression ranged from 0.35 to 0.39, the corresponding correlations corrected for unreliability ranged from 0.44 to 0.49. The latter values clearly suggest the presence of quite substantial associations between life event and depression measures, indicating that one reason for the low observed correlations between life event and depression scores is that life event measures are contaminated by substantial measurement error.

However, while the effects of low reliability were to impart a downward bias to the observed correlations, the consequences of this bias were more complex when non-recursive path models were fitted to the data. In our original model fitted under the assumption that the data were of perfect reliability, the results suggest that while the predominant direction of causation was from life event to depression, there was a small positive feed-back loop from depression to life events. This latter path implies that increased depressive symptoms may be associated with a tendency either to experience or report a greater number of life events. In the model solved taking measurement unreliability into account, there was no significant feed-back loop.

Inspection of the disturbance terms of the fitted model suggests that the model based on the corrected correlations was the more satisfactory. In particular, in the model solved assuming data were of perfect reliability, there was a substantial negative correlation between the disturbance terms of the structural equations. This negative correlation is both unexpected and inexplicable since it implies that the sources of disturbance influencing life event measures were negatively related to the sources of disturbance influencing the depression measure. It is hard to produce a theoretical justification for this result. Further, Luskin (1978) has argued that anomalous correlations between disturbance terms are indicative of the presence of specification error in a model. Thus the unexplained negative correlation between disturbances suggests that something was amiss with the original model. However, when this model was solved taking into account measurement error, the disturbance terms were found to be uncorrelated. Given that there was no significant feed-back in this model, the presence of uncorrelated disturbance terms seems to be quite reasonable.

At a theoretical level, the results suggest that the subject's probability of displaying a singificant number of depressive symptoms is influenced by at least two sets of factors: (a) the subject's current exposure to adverse life events; (b) the subject's history of previous depression. These findings may be interpreted in terms of the distinction between provoking and predisposing factors suggested by Brown and Harris (1978). In particular, the measure of previous depression is clearly a 'proxy' variable, summarising the subject's tendencies or susceptibility to depression, and thus may be considered to be a predisposing factor. On the other hand, the measure of current exposure to life events is clearly a provoking factor. While the findings may be interpreted in terms of the distinction made by Brown and Harris, it is important to note that the theoretical model used here differs from the Brown and Harris model. In particular, Brown and Harris argue that provoking and predisposing factors combine in a non-additive and interactive fashion, whereas the model proposed here assumes that the effects of these factors are linear and additive.

Whilst the development of measurement and structural equation models provides a number of insights into the possible structure of life event and depression measures over time, as with all modelling methods, these insights are purchased at the cost of making a number of simplifying assumptions. The major assumption that permeates the model-fitting process is that both errors of measurement and disturbance terms in models are uncorrelated across measurement periods.

In the reliability model, it is assumed that measurement errors are uncorrelated with each other and with true score values. However, given that the variables are lagged measures of the same construct, it is open to question whether the former condition will be satisfied. Wheaton et al. (1977) have shown, using a multiple indicators approach, that when errors of measurement are correlated between measuring periods, the estimation of reliability is biased upward and, in fact, the reliability coefficient may exceed unity. In addition, the life event scores analysed had a positively skewed distribution and this raises the possibility that errors of measurement may be correlated with true score values. Thus, it is possible that departures from the assumptions of the model may have led to misleading estimates of test reliability. At the same time, however, the values obtained by this analysis appear to be similar to those reported by other studies and, to this extent, the results are consistent with prior knowledge of the reliability of the life event measures.

In the structural equation model it is assumed that the disturbances on the measures taken at the second time period are uncorrelated with the initial measurements. Again, this is a strong assumption which is unlikely to be exactly satisfied. The variables in the model are, in fact, lagged measures of the same construct, and it would be expected that the non-observed factors which influenced depression (say) at time 1 would be similar to the non-observed disturbance factors influencing depression at time 2. Thus one might expect the disturbance terms to be correlated with the initial value of the depression variable. A similar argument applies to the life event measures. However, to ensure that the model is identified, it is necessary to assume the absence of these correlations. The effect of a violation of this assumption is to produce a biased and inconsistent estimate of the coefficients linking variables between measurement periods (Hibbs 1974). The principle involved can most easily seen by considering the interpretation of the correlation between depression at the two measuring periods. In particular, while the model assumes that depression at the first measuring period 'causes' depression at the later measurement period, it is quite clear that the measure of depression at the first period is, in effect, a proxy variable which stands for a number of unmeasured variables reflecting the subject's susceptibility to depression. Thus the coefficient linking the depression measure at time 1 to the depression measure at time 2 is not an unalloyed estimate of the extent to which early depression leads to later depression but reflects the effects of common but non-measured factors which influence depression at both time periods. A technical account of the problems of estimation and inference in models using lagged endogenous variables is given by Hibbs (1974).

By now, the reader may be convinced that the strong assumptions required to build structural equation models of life events and depression are such as to render any inference from these models suspect. At the same time, the assumptions used differ little from those customarily employed in more familiar tabular or regression analyses, the major difference being that the structural equation modelling approach makes explicit the strong assumptions that are required to draw causal inferences from correlational data.

Furthermore, very rarely in epidemiological research are conclusions based on the application of a specific method to a single data set but rather rely on the convergence of evidence from a variety of sources using a variety of methods (Susser 1973). In this respect, the conclusions drawn from the present analysis show considerable agreement with the findings of studies which have examined the correlation between life event and depression measures using empirical techniques. In these studies, life event inventories have been refined by either a dating of life events to ensure that life events preceded the onset of depressive illness (Paykel et al. 1969; Thomson and Hendrie 1972; Brown and Harris 1978) or through the elimination of life events which might have been caused by the respondent's agency (Brown and Harris 1978). In general, these studies have confirmed the view that the correlation between life events and depression cannot be explained by a tendency for depressed individuals to either report or experience a greater number of life events. This conclusion, of course, is consistent with the findings of the present analysis which has looked at the problem using statistical rather than empirical methods.

Acknowledgement. This research was funded by grants from the Medical Research Council of New Zealand and the National Children's Health Research Foundation.

References

Andrews G (1981) A prospective study of life events and psychological symptoms. Psychol Med 11: 795-801

Benjaminsen D (1981) Stressful life events preceding the onset of neurotic depression. Psychol Med 11: 369-378

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

Brown G, Prudo R (1981) Psychiatric disorder in a rural and urban population: 1. Aetiology of depression. Psychol Med 11: 581-599

Cooper B, Sylph J (1973) Life events and the onset of neurotic illness: an investigation in general practice. Psychol Med 3: 421-435

Cronbach L (1951) Coefficient alpha and the internal structure of tests. Psychology 16: 297-334

Fergusson D, Horwood L, Shannon F (1981) Birth placement and childhood disadvantage. Soc Sci Med [E] 15: 315-326

Fergusson D, Horwood L (1984) Life events and depression in women: a structural equation model. Psychol Med 14: 881-889

Finlay-Jones R, Brown G (1981) Types of stressful life events and the onset of anxiety and depression disorders. Psychol Med 11: 803-815

Gersten J, Langner T, Eisenberg L, Simcha-Fagan O (1977) An evaluation of the aetiologic role of stressful life change events in psychiatric disorder. J Health Soc Behav 18: 228-244

Guilford J, Fruchter B (1973) Fundamental statistics in psychology and education. McGraw-Hill, New York

Heise D (1969) Separating reliability and stability in test-retest correlation. Am Soc Rev 34: 93-101

Henderson A, Bryne D, Duncan-Jones P (1981) Neurosis and the social environment. Academic Press, Sydney

Hibbs D (1974) Problems of statistical estimation and causal inference in time-series regression models. In: Costner H (ed) Sociological methodology 1973-1974. Jossey-Bass, San Francisco

Holmes T, Rahe R (1967) The social readjustment rating scale. J Psychosom Res 11: 213-218

Hornstra R, Klassen D (1977) The course of depression. Compr Psychiatry 18: 119-125

Horowitz H, Schaffer D, Hirito D, Wilner N, Levin B (1977) Life event questionnaires for measuring presumptive stress. Psychosom Med 39: 413-431

Joreskog K (1973) A general method for estimating a linear structural equation system. In: Goldberger A, Duncan O (eds) Structural equation models for the social sciences. Seminar, New York

Joreskog K, Sorbom D (1976) LISREL - estimation of linear structural equation systems by maximum likelihood methods. International Educational Services, Chicago

Luskin R (1978) Estimating and interpreting correlations between disturbances and residual path coefficients in non-recursive (and recursive) causal models. Am J Pol Sci 22: 444-474

Masuda M, Holmes T (1967) Magnitude estimations of social readjustments. J Psychosom Res 11: 219-225

Paykel E, Myers J, Dieuelt M, Klevmen G, Lindenthal J, Pepper M (1969) Life events and depression: a controlled study. Arch Gen Psychiatry 21: 753-760

Pilowsky I, Levine S, Boulton D (1969) The classification of depression by numerical taxonomy. Br J Psychiatry 115: 937-945

Pilowsky I, Boulton D (1970) Development of a questionnaire based decision rule for classifying depressed patients. Br J Psychiatry 116: 647-650

Rabkin J, Struening E (1976) Life events, stress and illness. Science 194: 1013-1020

Stewart A, Salt P (1981) Life stress, life styles, depression and illness in adult women. J Pers Soc Psychol 40: 1063-1069

Susser M (1973) Causal thinking in the health sciences. Oxford University Press, London

Tennant C, Bebbington P, Hurry J (1981) The role of life events in depressive illness: is there a substantial causal relation? Psychol Med 11: 379-389

Thomson K, Hendrie H (1972) Environmental stress in primary depressive illness. Arch Gen Psychiatry 26: 130-132

Warheit G (1979) Life events, coping, stress and depressive symptomatology. Am J Psychiatry 136: 502-507

Wheaton B, Muthen B, Alwin D, Summers G (1977) Assessing reliability and stability in panel models. In: Heise D (ed) Sociological methodology. Jossey-Bass, San Francisco

Wiley D, Wiley J (1970) The estimation of measurement error in panel data. Am Soc Rev 35: 112-117

Accepted 13 September 1985

Dr. D. M. Fergusson
The Christchurch Child Development Study,
Department of Paediatrics
Christchurch Clinical School of Medicine
Christchurch
New Zealand