

Structural equation modeling of repeated retrospective reports of childhood maltreatment

DAVID M. FERGUSSON, L. JOHN HORWOOD & JOSEPH M. BODEN

University of Otago, Christchurch School of Medicine and Health Sciences, Christchurch, New Zealand

Key words

childhood sexual abuse,
childhood physical abuse,
structural equation modelling,
longitudinal study

Correspondence

David M. Fergusson, Christchurch
Health and Development Study,
University of Otago, Christchurch
School of Medicine and Health
Sciences, PO Box 4345,
Christchurch, New Zealand.
Telephone (+64) 3 372 0406
Fax (+64) 3 372 0407
E-mail: dm.fergusson@otago.ac.nz

Abstract

The present study examined errors of reporting, including unreliability and recall bias, in retrospective reports of childhood maltreatment. Data were gathered on retrospective reports of childhood sexual and physical abuse and current mental health in a birth cohort of over 980 participants studied at ages 18 and 21. A structural equation model was developed to estimate the contributions of test unreliability and recall bias to reporting variation. Retrospective abuse reports were of modest test–retest reliability ($r_{tt} = 0.50$ approximately). Recall bias accounted for <1% of report variance. As a consequence the observed correlations between maltreatment and mental health closely approximated the corresponding estimated true correlations. The results of the study suggested that, for this cohort, errors of measurement in reports of childhood maltreatment did not pose a significant threat to study validity. Copyright © 2011 John Wiley & Sons, Ltd.

Received 19 March 2009;
revised 17 June 2009;
accepted 12 October 2009

Introduction

Over the last three decades there has been an increasing body of research evidence that supports the conclusion that exposure to childhood maltreatment, including sexual and physical abuse, is associated with increased risks of later mental health problems and related problems of adjustment.

In particular, children exposed to maltreatment have been shown to be at increased risks of later adjustment problems using data gathered from both prospective and retrospective designs with these associations being shown

to persist following controls for confounding (for reviews see: Chalk *et al.*, 2002; Cicchetti and Toth, 2005; Fergusson and Mullen, 1999; Finkelhor, 1990; Holmes and Slap, 1998; Johnson, 2004; Kaplan *et al.*, 1999; Mullen *et al.*, 2000; National Clearinghouse on Child Abuse and Neglect Information, 2005; Newton and Vandeven, 2008; Putnam, 2003; Stirling *et al.*, 2008; Wissow, 1995). Nonetheless, some uncertainties remain.

One of the more controversial issues in this area concerns the use of retrospective reports of child abuse. Because of the practical and ethical problems associated with obtaining prospective measures of abuse in childhood,

the majority of studies in this area have relied on the retrospective reports of adults recalling their childhood (DiLillo *et al.*, 2006; Fergusson and Mullen, 1999). In turn this use of retrospective reports has raised concerns about the reliability and validity of such reports as measures of exposure to abuse in childhood (Maughan and Rutter, 1997; White *et al.*, 2007; Widom and Morris, 1997; Widom *et al.*, 2004; Widom and Shepard, 1996; Widom *et al.*, 1999; Williams, 1994).

There have been several studies that have examined the reliability of retrospectively gathered reports of child abuse using a test–retest paradigm in which the same sample of respondents is questioned about these issues on two occasions (DiLillo *et al.*, 2006; Dill *et al.*, 1991; Fergusson *et al.*, 2000; Friedrich *et al.*, 1997; Fry *et al.*, 1996; Kooiman *et al.*, 2002; Leserman *et al.*, 1995; Lipschitz *et al.*, 1999; Martin *et al.*, 1993; Meyer *et al.*, 1996; Mills *et al.*, 2007). This research has shown wide variation in the estimates of the reliability of retrospective reports with kappa values ranging from 0.30 to 0.82. However, the studies with high kappa values have employed clinical samples and have had relatively short delays between assessments (Kooiman *et al.*, 2002; Lipschitz *et al.*, 1999; Meyer *et al.*, 1996). Studies which have employed non-clinical samples and those with longer delay between assessments have shown a more modest level of reliability with kappa values ranging from 0.30 to 0.65 (DiLillo *et al.*, 2006; Dill *et al.*, 1991; Fergusson *et al.*, 2000; Friedrich *et al.*, 1997; Fry *et al.*, 1996; Leserman *et al.*, 1995). These findings clearly suggest the presence of quite substantial unreliability in the retrospective reports of child abuse.

A further and influential criticism of the use of retrospective reports of child abuse has been that the errors of reporting are related to current mental health, thus leading to inflated and artifactual correlations between retrospective reports of child abuse and current mental health. On this issue Widom and colleagues noted "... a major problem in making inferences about the association between retrospectively assessed child adversity and later health outcomes is not random measurement error but recall bias (Raphael, 1987). The net effect of recall bias is to lead artifactually to an inflation of measures of association by creating differential accuracy of reports of childhood adversity in reports of those with relatively poor versus relatively good health" (Widom *et al.*, 2004).

Such concerns have led to a widespread advocacy that research into the mental health effects of child abuse should be based on prospective studies in which abuse is assessed in childhood and mental health in later life (Tajima *et al.*, 2004; Widom *et al.*, 2004). These designs avoid the type of recall bias identified by Widom *et al.* (2004) but have a

number of limitations. Specifically, because of the practical and ethical difficulties of assessing child abuse during childhood most prospective research has been confined to clinic or agency samples of children known to be exposed to child abuse. Because these children have come to attention it is likely that some attempt has been made to mitigate the effects of child abuse. There is no guarantee that the findings for selected populations of children coming to attention will mirror the linkages between child abuse and mental health that exist in unselected and untreated population samples. For these reasons it is important to examine the extent to which estimates of the association between child abuse and mental health obtained from retrospective reports are subject to a recall bias in which the effects of current mental state on the reporting of past child abuse inflates estimate of the associations between reported child abuse and current mental health.

In this paper we explore a method for addressing the issues of reporting unreliability and recall bias in retrospective reports of child abuse by constructing a structural equation model. This model estimates the associations between child abuse and later mental health taking into account: (a) unreliability in the retrospective reporting of abuse; (b) the effects of current mental health on variation in retrospective reports of child abuse. This model relies on the use of a repeated measures design which, subject to identifying assumptions, permits the estimation of model parameters representing: (a) test unreliability in the reporting of child abuse; (b) the effects of current mental health on the reporting of abuse; and (c) the association between child abuse and mental health net of the effects of test unreliability and recall bias arising from the effects of current mental health on retrospective reports of abuse. A description of the key features and assumptions of the model is given later. A more formal statistical formulation is given in the Methods section of this paper.

A structural equation model of reporting error in reports of childhood maltreatment

Consider a sample of young adults studied on two occasions (t_1 , t_2) at which retrospective reports of child abuse (X) and measures of current mental health (Y) are obtained. Let X_1 , X_2 denote the reports of child abuse at times t_1 , t_2 , and Y_1 , Y_2 the corresponding measures of mental health.

Figure 1 shows a structural equation model of the relationships between the two measures observed at the two times. This model assumes that:

- (1) The measures X_1 , X_2 are fallible indicators of the individual's true but non-observed history of child

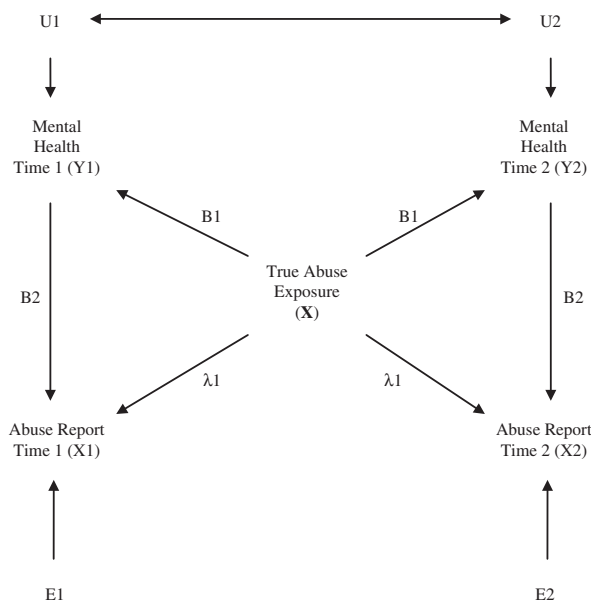


Figure 1 Theoretical model of associations between child abuse reports and mental health status at two times allowing for contamination of abuse reports by current mental health status.

abuse represented by the latent variable X . This component of the model specifies a measurement model linking the observed retrospective reports to a latent variable representing the individual's true but non-observed history of abuse.

- (2) The latent variable X is related to mental health variables Y_1, Y_2 by a regression model. If all variables in the model are standardized the parameter B_1 represents the correlation between the latent variable X and the observed measures of mental health.
- (3) The observed measures of mental health Y_1, Y_2 are permitted to influence the retrospective reports of child abuse (X_1, X_2). This pathway represents the effects of current mental health on the reporting of child abuse net of any effect of child abuse on mental health.
- (4) The model also includes the error or disturbance terms E_1, E_2 and U_1, U_2 . The terms E_1, E_2 represent sources of measurement error in the retrospective abuse reports X_1, X_2 . These errors are assumed to be uncorrelated with each other and with all other variables in the model (aside from their respective abuse reports X_1, X_2). The terms U_1, U_2 represent the variation in the measures of mental health Y_1, Y_2 that is not explained by the latent variable X . These terms are assumed to be uncorrelated with E_1, E_2 but are permitted to be correlated with each other to allow for the effects of common non-observed factors that influence mental health outcomes at both times.

- (5) Finally the model assumes that the model parameters which describe the associations between observations obtained at the same time are the same for both times of observations.

While the model in Figure 1 has a relatively simple structure, it can be used to obtain three sets of estimates which have an important bearing on the issues of test reliability and recall bias in retrospective reports of child abuse:

- (1) First, as shown in the Methods section, the parameter estimates can be used to decompose the variance in reports of child abuse into components reflecting: (a) variance attributable to test unreliability; (b) variance due to the effects of current mental health on the reporting of child abuse; (c) variance reflecting the true but non-observed histories of abuse of the cohort.
- (2) Second, when the model parameters are fully standardized the parameter B_1 represents the correlation between reported child abuse and mental health corrected for test unreliability in abuse reports and the effects of current mental health on the reporting of child abuse.
- (3) Third, the parameter B_2 represents the effects of current mental state on the reporting of child abuse. The size of this parameter provides a guide to the extent to which current mental state is likely to bias abuse reports.

As with all such models these estimates are purchased at the cost of making a number of assumptions that ensure the proposed model is identified and parameter estimates can be obtained.

The key identifying conditions for the model in Figure 1 are:

- (1) Parameter stability: The model parameters that describe within time associations are the same at both times of measurement. This assumption implies that the same processes acting in the same way influence errors in retrospective reports of abuse at both times.
- (2) Independence of errors: As noted earlier the model assumes that E1 and E2 are uncorrelated with each other and with U1, U2. These assumptions imply that after the effects of any recall bias are taken into account errors in the reporting of child abuse are unrelated to each other and to the mental health outcomes Y1, Y2.

In the remainder of this paper, we apply the model in Figure 1 to data gathered on retrospective reports of child abuse and current measures of mental health outcomes in a birth cohort of New Zealand young people studied at ages 18 and 21.

Method

The data described in this report were gathered during the course of the Christchurch Health and Development Study (CHDS). The CHDS is a longitudinal study of an unselected birth cohort of 1265 children (635 males, 630 females) born in the Christchurch (New Zealand) urban region over a four month period during mid-1977. This cohort has been studied at birth, four months, one year and annual intervals to age 16 years, and at ages 18 and 21 years using information gathered from a combination of sources, including parent interviews, teacher reports, psychometric testing, child interviews, medical, police and other records. An overview of the study design has been given previously (Fergusson and Horwood, 2001; Fergusson *et al.*, 1989).

Measures

At age 18 and 21 years, sample members were interviewed on a structured questionnaire that examined a range of mental health issues, including childhood exposure to sexual or physical abuse, symptoms of psychiatric disorders and related problems of adjustment. Interviews typically lasted between 1.5 to two hours and were administered in private by trained and experienced female interviewers recruited for the project. In all cases, the

release of interview data was subject to signed and informed consent from the respondent. The following measures were used in the present analysis.

Childhood sexual abuse

As part of the interview conducted at ages 18 and 21 years, participants were asked whether, before the age of 16, anyone had ever attempted to involve them in any of a series of 15 sexual activities when they did not want this to happen. These activities spanned: (a) non-contact episodes, including indecent exposure, public masturbation by others and unwanted sexual propositions or lewd suggestions; (b) incidents involving sexual contact in the form of sexual fondling, genital contact or attempts to undress the respondent; (c) incidents involving attempted or completed vaginal, oral or anal intercourse (Fergusson *et al.*, 1996b). Young people who reported having experienced any of these behaviors before the age of 16 were then asked, for each perpetrator involved, a further series of questions concerning the nature and extent of abuse, the characteristics of the perpetrator, abuse disclosure and treatment seeking or counseling subsequent to abuse. Information on these issues was gathered using a combination of pre-coded survey items and open-ended questions (Fergusson *et al.*, 1996b). For the purposes of the present analysis a scale measure of abuse exposure was derived by classifying participants according to the most severe form of sexual abuse reported at each age. This scale was zero to three: (0) participant reported no sexual abuse; (1) participant reported incident(s) involving non-contact abuse only; (2) participant reported incident(s) involving physical contact with the perpetrator, but which did not involve attempted or completed sexual penetration; (3) participant reported incident(s) of abuse that involved attempted or completed vaginal, oral or anal intercourse.

Childhood physical abuse

The assessment of physical abuse was based on young people's reports of parental use of physical punishment. At ages 18 and 21 years, respondents were asked to report on the extent to which their parents used methods of physical punishment during their childhood years (prior to age 16). Reports were made on a five-point scale: (1) parent never used physical punishment; (2) parent seldom used physical punishment; (3) parent regularly used physical punishment; (4) parent used physical punishment too often or too severely; (5) parent used physical punishment in a harsh and abusive way. Separate ratings were obtained for the child's mother figure and father figure wherever possible. Ratings for both parents were

then combined into a single rating at each age by classifying the participant's exposure to physical abuse based on the most severe form of physical punishment reported for either parent (Fergusson and Lynskey, 1997). Owing to the small numbers in the last two categories these were combined in the present analysis into a single category representing those who reported overly frequent or severe physical punishment or harsh and abusive treatment by either parent.

Psychiatric adjustment (16–21 years)

At ages 18 and 21, participants were questioned concerning their psychiatric symptomatology between the ages of 16 to 18 years and 18 to 21 years respectively, using a questionnaire that combined elements of the Composite International Diagnostic Interview (CIDI) (World Health Organization, 1993) and the Self-Report Delinquency Inventory (SRDI) (Elliott and Huizinga, 1989). On the basis of these data, the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) (American Psychiatric Association, 1994) symptom criteria were used to classify young people according to a series of psychiatric disorder diagnoses over each assessment period. These disorders included: (a) major depression; (b) anxiety disorders (generalized anxiety disorder, panic/agoraphobia, specific phobia, social phobia); (c) conduct disorder; and (d) alcohol, cannabis or other illicit drug dependence. Items from the CIDI were used to assess depression, anxiety disorders and substance dependence, while items from the SRDI were used to assess the presence of conduct disorder in the sample. A detailed description of these measures has been provided by Fergusson *et al.* (1996a). In addition, at each interview participants were questioned about the occurrence of suicidal behaviors during the assessment period, including suicidal thoughts and suicide attempts. Those who reported experiencing suicidal thoughts or making a suicide attempt in a given interview period were classified as having suicidal ideation/attempt in that period. To provide an overall measure of the burden of mental disorder, the five individual measures (depression, anxiety disorder, conduct disorder, substance dependence, suicidal ideation/attempt) were summed to obtain a count of the number of mental health problems reported for each interval.

Statistical analysis

The model in Figure 1 was fitted to the variance/covariance matrix of the observed abuse report and mental health problem measures using asymptotic distribution free

weighted least squares (WLS) methods (Browne, 1984). These methods provide a useful alternative to maximum likelihood (ML) estimation in a context where the observed data are unlikely to conform to conventional assumptions of multivariate normality. Model fitting was conducted using Mplus software (Muthén and Muthén, 2007). Model goodness-of-fit was assessed on the basis of a number of indices including: (a) the model chi square goodness-of-fit statistic; (b) the root mean squared error of approximation (RMSEA), values of RMSEA less than 0.05 are assumed to be indicative of a well-fitting model; (c) the Comparative Fit Index (CFI), this index varies between zero and one with values close to one indicating a well fitting model (Joreskog and Sorbom, 1993).

Sample size and sample bias

A total of 1053 participants provided data on abuse exposure and mental health problems on at least one of the assessments at age 18 or 21 years. This sample represented 83% of the original cohort of 1265 participants. However, since not all participants were assessed on all measures at both 18 and 21 years, the sample numbers with complete data on both assessments were somewhat reduced, with sample sizes of $N=983$ for physical abuse reports (78% of the cohort) and $N=980$ for sexual abuse reports (77% of the cohort). The structural equation models reported in this paper were based on these reduced samples. The following procedures were used to address potential sample selection issues relating to missing data and processes of sample attrition.

First, all models were re-fitted in Mplus to the full sample of 1053 participants who had at least partial data using missing data imputation methods that assumed data were missing at random (Muthén and Muthén, 2007). Since data imputation was not available in Mplus for WLS analysis, the re-analysis was conducted using ML. Second, data weighting methods were used to test for possible selection bias resulting from processes of sample attrition. These methods involved a two stage process. (a) In the first instance, a sample selection model was constructed by using data gathered at birth to predict inclusion in the analysis sample. This showed that there were statistically significant ($p < 0.05$) tendencies for the obtained sample to underrepresent young people from more socially disadvantaged backgrounds (low parental education, low socio-economic status, single parent family). On the basis of the fitted selection model, the sample was then post-stratified into a series of groups and the probability of study participation estimated for each group. (b) All structural models were then re-fitted with the data for

each individual weighted by the inverse of the probability of study participation. The analyses were conducted using estimation procedures that were robust to the data weighting assumptions (Muthén and Muthén, 2007). In all cases the analyses produced essentially identical conclusions to the findings reported here, suggesting that the effects of missing data and selection bias on the results were likely to be minimal.

Results

Agreement between abuse reports

Tables 1 and 2 show the joint frequency distribution of the abuse reports obtained at ages 18 and 21 years. Table 1 shows measures of childhood sexual abuse and Table 2 childhood physical abuse. Examination of Tables 1 and 2 show:

- (1) *Childhood sexual abuse*: The marginal distributions of the severity of sexual abuse reported at ages 18 and 21 years are generally very similar: overall 10% of the sample reported some form of sexual abuse at age 18 compared to 8.5% at age 21; and between 7–8%

reported contact abuse or abuse involving attempted/completed intercourse at each age. Despite this similarity, the overall agreement between abuse reports at the two ages was modest. The test–retest correlation between the two measures of abuse severity was $r_{tt} = 0.49$, suggesting that there was substantial unreliability in the reporting of abuse exposure at the two times.

- (2) *Childhood physical abuse*: The reports of childhood physical abuse show a similar pattern of results, with very similar marginal distributions of abuse severity at ages 18 and 21, but substantial inconsistency in the reporting of physical abuse over time. The correlation between abuse reports at the two ages was $r_{tt} = 0.51$, again suggesting the presence of substantial unreliability in the reporting of physical abuse.

Associations between abuse reports and mental health status

Table 3 shows the matrix of correlations between abuse reports and the total number of mental health problems reported at ages 18 and 21 years. Again, the correlations

Table 1 Joint frequency distribution of childhood sexual abuse reports at 18 and 21 years

18 Year report	21 Year report				Percentage of sample
	None	Non-contact	Contact	Intercourse	
None	843	7	13	17	89.8
Non-contact	17	4	1	2	2.5
Contact	26	1	9	9	4.6
Intercourse	11	1	4	15	3.2
Percentage of sample	91.5	1.3	2.8	4.4	100

$r_{tt} = 0.49$; $p < 0.0001$.

Table 2 Joint frequency distribution of childhood physical abuse reports at 18 and 21 years

18 Year report	21 Year report				Percentage of sample
	None	Occasional	Regular	Severe/harsh	
None	42	72	0	0	11.6
Occasional	36	663	46	13	77.1
Regular	0	39	26	8	7.4
Severe/harsh	2	9	11	16	3.9
Percentage of sample	8.1	79.7	8.4	3.8	100

$r_{tt} = 0.51$; $p < 0.0001$.

Table 3 Correlations between measures of mental health (16–18, 18–21 years) and abuse reports (18 and 21 years)

Measure	Number of mental health problems		Childhood abuse report	
	16–18 Years	18–21 Years	18 Years	21 Years
<i>Childhood sexual abuse</i>				
Number of mental health problems				
16–18 years	1.00			
18–21 years	0.40	1.00		
Childhood sexual abuse report				
18 years	0.33	0.25	1.00	
21 years	0.23	0.27	0.49	1.00
<i>Childhood physical abuse</i>				
Number of mental health problems				
16–18 years	1.00			
18–21 years	0.40	1.00		
Childhood physical abuse report				
18 years	0.19	0.13	1.00	
21 years	0.15	0.13	0.51	1.00

are shown separately for childhood sexual abuse and childhood physical abuse. Examination of Table 3 shows evidence of moderate associations between reported exposure to childhood sexual abuse (<16 years) and subsequent mental health, with correlations ranging from $r=0.23$ to 0.33 ($p < 0.0001$). For childhood physical abuse the associations were more modest ($r=0.13$ to 0.19 ; $p < 0.0001$).

Modeling reporting errors and bias in abuse reports

To examine the effects of test unreliability and recall bias in retrospective reports, the model in Figure 1 was fitted to the data in Table 3. Separate models were fitted for childhood sexual abuse and physical abuse. The results of these analyses are depicted in Figure 2 which shows the fitted model coefficients (with standardized coefficients in parentheses) and tests of goodness-of-fit for each model. Examination of Figure 2 suggests the following conclusions:

- (1) For both models: (a) values of the model chi square were non-significant; (b) the RMSEA was between .00-.02; (c) the CFI was $>.99$. All statistics indicate excellent fit.
- (2) The standardised coefficients linking latent true abuse exposure to the observed abuse reports provide estimates of the correlations between observed and true abuse status. For childhood sexual abuse these correlations ranged from .66 to .70, and for physical abuse from .69 to .72.
- (3) The standardised coefficients linking current mental health to the observed abuse reports at the same age provide estimates of the component of the correlation between concurrently assessed measures of abuse status and mental health that was attributable to a reporting bias resulting from current mental health status. For childhood sexual abuse there was evidence of a modest but statistically significant reporting bias ($p < .05$) in abuse reports, with standardised coefficients ranging from .09 to .10. For physical abuse this bias was both small (with standardised coefficients of .03) and statistically non-significant.
- (4) The standardised coefficients linking latent true abuse exposure to mental health at ages 18, 21 years provide estimates of the true correlation between abuse exposure and mental health adjusted for both unreliability in abuse reports and contamination of abuse reports by current mental health status. For childhood sexual abuse these correlations ranged from .29 to .31 ($p < .0001$), and for physical abuse from .16 to .17 ($p < .001$).
- (5) For both models there was evidence of a correlation between the disturbances (U_1, U_2) on the number of mental health problems at ages 18, 21 years. These correlations ranged from .34 to .38 ($p < .0001$), reflecting the fact that childhood exposure to sexual or physical abuse explained only a small fraction of the total correlation between mental health outcomes at the two times.

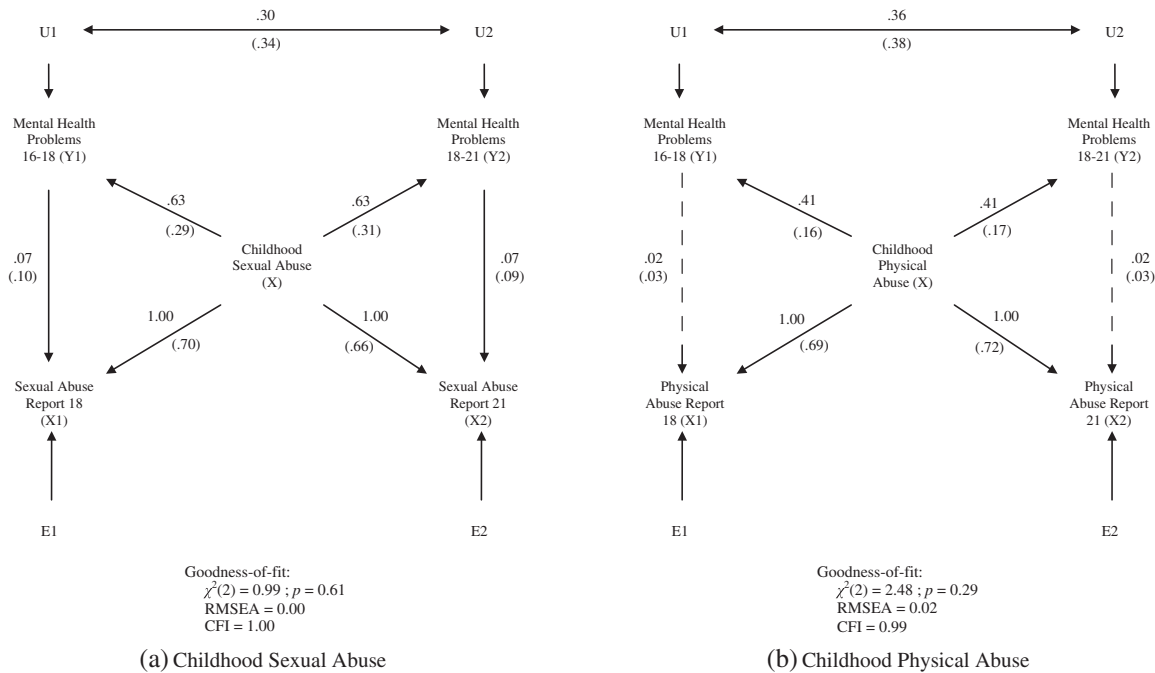


Figure 2 Fitted structural equation models estimating bias in abuse reports from current mental health status. (Standardized model coefficients are given in parentheses, dashed arrow indicates coefficient not significantly different from zero, $p > 0.05$.)

(6) From the fitted model coefficients it was possible to estimate the components of variance in the observed abuse reports that were attributable to: (a) reporting bias due to current mental health; (b) errors of measurement due to measurement unreliability; and (c) true abuse exposure. For sexual abuse, reporting bias accounted for less than 1% of reporting variance; and measurement unreliability accounted for between 46% to 52% of this variance, with the remainder due to true abuse exposure. For physical abuse, the component due to bias in abuse reports was essentially zero (<0.1%) and the component due to measurement unreliability ranged from 48% to 51%, with the remainder attributable to true abuse exposure.

To examine the consequences of reporting errors on the estimated correlations between childhood maltreatment and mental health Table 4 compares the observed within time correlations between abuse reports and number of mental health problems at ages 18 and 21 with the corresponding estimated true correlations from the structural equation models. Table 4 shows:

(1) For sexual abuse, the observed correlations (0.33, 0.27) were very similar to the estimated true correlations (0.29, 0.31).

Table 4 Comparison of observed within time correlations between abuse reports and number of mental health problems at 18 and 21 years with estimated true correlations

Measure	Observed correlation	True correlation
<i>Sexual abuse reports</i>		
18 years	0.33	0.29
21 years	0.27	0.31
<i>Physical abuse reports^a</i>		
18 years	0.19	0.19
21 years	0.13	0.20

^aThe true correlations for physical abuse report data were estimated from a model in which the recall bias parameter (B2) was fixed to zero.

(2) For physical abuse, the observed correlations (0.19, 0.13) were the same as or smaller than the estimated true correlations (0.19, 0.20).

Both comparisons show that errors of measurement in the reporting of childhood maltreatment appeared to have negligible effects on the correlations between maltreatment and mental health outcomes.

Effects of permitting E1, E2 to be correlated

As noted earlier the models in Figures 1 and 2 are identified by virtue of the assumption that the errors in abuse reports E1, E2 are uncorrelated after the effects of current mental state on the reporting of X1, X2 has been taken into account. The model thus could be mis-specified if E1, E2 were correlated between measurement periods. While it is not possible to estimate a correlation between E1, E2 it is possible to conduct a sensitivity analysis of the effects of varying the size of the correlation between E1 to E2 on the model conclusions. This issue is explored in Table 5 which shows estimates of: (a) the correlation between true child abuse and mental health; and (b) the extent of the bias in observed abuse reports due to current mental health. These estimates were obtained by extending the model in Figure 1 to include a correlation between the errors E1, E2 and fixing this correlation over a plausible range of values from zero to +0.30. Table 5 leads to the following conclusions:

- (1) With an increasing positive correlation between E1, E2 there are corresponding increases in the size of the estimated correlation between true child abuse X and current mental health Y1, Y2. These results imply that, subject to the condition that the correlation between E1, E2 is non-negative, the model in Figure 1 in fact gives a lower limit estimate of the correlation between child abuse and mental health.
- (2) The parameter B2 representing the effect of current mental health on the retrospective abuse reports X1, X2 does not vary with the correlation between E1, E2. This result reflects the fact that estimation of B2 does not depend in any way on the correlation between E1, E2.

Discussion

This paper has presented a structural equation model designed to examine the effects of errors of measurement in retrospective reports of childhood maltreatment. A key

feature of this model is that it provides estimates of the correlation between childhood maltreatment and mental health outcomes corrected for both unreliability in the reporting of maltreatment and the effects of any recall bias process in which current mental health influences the recall of past maltreatment. This model was applied to retrospective measures of childhood sexual abuse and physical abuse collected on a birth cohort of young adults studied at age 18 and 21 years. The key findings of this analysis and their implications are discussed.

Childhood sexual abuse

The test–retest reliability for childhood sexual abuse was modest ($r_{tt}=0.49$) suggesting that just over 50% of the variance in these reports was due to errors of measurement. This finding is generally consistent with those reported in other studies which have found poor to moderate test–retest agreement in retrospective reports of childhood maltreatment (DiLillo *et al.*, 2006; Dill *et al.*, 1991; Fergusson *et al.*, 2000; Friedrich *et al.*, 1997; Fry *et al.*, 1996; Leserman *et al.*, 1995; Martin *et al.*, 1993). Such findings clearly suggest that errors of measurement in retrospective reports of childhood sexual abuse are large and these errors may pose an important threat to study validity depending on the nature and source of the measurement errors.

The structural equation model developed in this paper suggested that retrospective reports of childhood sexual abuse were subject to two sources of measurement error. First, a substantial amount (around 50%) of the variation in abuse reports represented measurement errors due to reporting unreliability that were uncorrelated with mental health status. The effects of such error would be to bias observed correlations between reported abuse and mental health outcomes downwards with the result that the observed correlation would be a conservative estimate of the true correlation between childhood maltreatment and

Table 5 Sensitivity analysis of the impact of varying assumptions about the size of the correlation between measurement errors in abuse reports on estimated model parameters

Size of correlation between measurement errors in abuse reports (E1, E2)	Correlations between true abuse status and mental health		Bias in abuse reports due to mental health	
	Sexual abuse	Physical abuse	Sexual abuse	Physical abuse
0.00	0.29, 0.31	0.16, 0.17	0.09–0.10	0.03
0.10	0.31, 0.32	0.17, 0.18	0.09–0.10	0.03
0.20	0.34, 0.36	0.19, 0.20	0.09–0.10	0.03
0.30	0.39, 0.42	0.21, 0.23	0.09–0.10	0.03

mental health. However, reporting of sexual abuse was also influenced by a small recall bias which explained about 1% of the reported variance. The effects of this bias were such that correlations between reports of abuse and mental health outcomes would be biased upwards by about 0.10. The net effects of the competing tendencies for test unreliability to bias correlations downward and recall bias to bias correlations upwards was that the observed correlations between reported sexual abuse and mental health outcomes ($r = 0.33, 0.27$) proved to be very similar to the estimated true correlations ($r = 0.29, 0.31$). These results suggest that for this cohort, errors of measurement in the reporting of childhood sexual abuse did not pose a significant threat to study validity. Nonetheless the results did suggest the presence of a small recall bias in which current mental state influenced the reporting of childhood sexual abuse.

Childhood physical abuse

In common with retrospective reports of childhood sexual abuse, reports of childhood physical abuse had a modest test–retest reliability ($r_{tt} = 0.51$) suggesting the presence of substantial errors of measurement in these reports. Further modeling of response errors suggested that all of the error variation in retrospective reports of physical abuse was attributable to errors of measurement that were uncorrelated with current mental state measures. This result is consistent with the view that sources of measurement error in reports of physical abuse were due to unreliability in reporting rather than to systematic errors of measurement. As a consequence, the observed correlations between reports of physical abuse and mental health are the same as or smaller than the true correlations.

Implications

The findings in this present study have a number of important implications for debates regarding the use of retrospective reports of childhood maltreatment in studies of maltreatment and adult mental health. Specifically, they serve to address the concerns that have been raised about the use of retrospective reports because of errors of measurement in these reports (Maughan and Rutter, 1997; White *et al.*, 2007; Widom and Morris, 1997; Widom *et al.*, 2004; Widom and Shepard, 1996; Widom *et al.*, 1999; Williams, 1994). These concerns have focused, particularly, on the issue of recall bias and it has been argued that because of the threat to validity posed by recall bias, the use of retrospective report data is suspect and that prospective

study of samples of children known to be exposed to abuse provides a more trustworthy research approach (White *et al.*, 2007; Widom and Morris, 1997; Widom *et al.*, 2004; Widom and Shepard, 1996).

However, the findings of the present study do not support these claims. In particular, the study estimates suggest three conclusions regarding the effects of recall bias. First, the size of any recall bias was small, accounting for less than 1% of reporting variance, and was confined to retrospective reporting of childhood sexual abuse. Second, the effects of any recall bias in biasing correlations upward were largely offset by the effects of test unreliability that biased correlations downwards. Finally, the net effects of these errors of measurement were that, for sexual abuse, the observed correlations were found to be very similar to the estimated true correlations whilst for physical abuse the observed correlations tended to underestimate the true correlations. Neither of these results is supportive of claims that errors of measurement pose major threats to the validity of studies that use retrospective reports of childhood maltreatment. Rather, the findings suggest that although these reports are subject to substantial errors of measurement, these errors do not unduly influence estimates of the associations between exposure to childhood maltreatment and adult mental health.

Limitations

There are a number of important limitations that should be recognized. First and foremost the findings apply to a specific cohort, studied at a specific time, using a specific set of measurement methods. There is no guarantee that the findings reported here will generalize to other samples studied at other times and using other methods. Second, the estimates of measurement error obtained in this study depend critically on a number of assumptions that underlie the model. These assumptions rely on patterns of stability and change in report data to secure estimates of measurement error. As a consequence the model cannot detect errors of measurement that are not manifest in patterns of change and stability. For example, the model cannot detect errors of measurement that arise from respondents consistently misreporting their childhood exposure to maltreatment.

In addition, for the model to be identified (estimable) it is necessary to make assumptions about both the symmetry of model parameters across time and the absence of a correlation between errors of measurement in abuse reports. The symmetry assumption could be questioned, particularly if the repeated measurements are far apart in time. However, this assumption is also

testable since the model is overidentified and lack of symmetry will be reflected in lack of model fit. The issue of the absence of correlation between errors in abuse reports is more problematic. However, as was shown in the supplementary analysis, if such a correlation did exist the estimated correlations between true abuse exposure and mental health reported in Table 4 are likely to be lower limit estimates of the true correlation. Further, the estimates of bias in abuse reports due to mental health would be unchanged.

References

- American Psychiatric Association. (1994) *Diagnostic and Statistical Manual of Mental Disorders* (4th ed.), Washington, DC, American Psychiatric Association.
- Browne M.W. (1984) Asymptotically distribution-free methods for the analysis of covariance structures. *British Journal of Mathematical and Statistical Psychology*, **37**(1), 62–83.
- Chalk R., Gibbons A., Scarupa H.J. (2002) *The Multiple Dimensions of Child Abuse and Neglect: New Insights into an Old Problem*, Washington, DC, Child Trends.
- Cicchetti D., Toth S.L. (2005) Child maltreatment. *Annual Review of Clinical Psychology*, **1**(1), 409–438.
- DiLillo D., Fortier M.A., Hayes S.A., Trask E., Perry A.R., Messman-Moore T., Fauchier A., Nash C. (2006) Retrospective assessment of childhood sexual and physical abuse: A comparison of scaled and behaviorally specific approaches. *Assessment*, **13**(3), 297–312.
- Dill D.L., Chu J.A., Grob M.C., Eisen S.V. (1991) The reliability of abuse history reports: A comparison of two inquiry formats. *Comprehensive Psychiatry*, **32**(2), 166–169.
- Elliott D.S., Huizinga, D. (1989) Improving self-reported measures of delinquency. In Klein M.W. (ed.) *Cross-national Research in Self-reported Crime and Delinquency*, pp. 155–186, Boston, MA, Kluwer.
- Fergusson D.M., Horwood L.J. (2001) The Christchurch Health and Development Study: Review of findings on child and adolescent mental health. *The Australian and New Zealand Journal of Psychiatry*, **35**(3), 287–296.
- Fergusson D.M., Horwood L.J., Lynskey M.T. (1996a) Childhood sexual abuse and psychiatric disorder in young adulthood: II. Psychiatric outcomes of childhood sexual abuse. *Journal of the American Academy of Child and Adolescent Psychiatry*, **35**(10), 1365–1374.
- Fergusson D.M., Horwood L.J., Shannon F.T., Lawton J.M. (1989) The Christchurch Child Development Study: A review of epidemiological findings. *Paediatric and Perinatal Epidemiology*, **3**(3), 278–301.
- Fergusson D.M., Horwood L.J., Woodward L.J. (2000) The stability of child abuse reports: A longitudinal study of young adults. *Psychological Medicine*, **30**(3), 529–544.
- Fergusson D.M., Lynskey M.T. (1997) Physical punishment/maltreatment during childhood and adjustment in young adulthood. *Child Abuse and Neglect*, **21**(7), 617–630.
- Fergusson D.M., Lynskey M.T., Horwood L.J. (1996b) Childhood sexual abuse and psychiatric disorder in young adulthood: I. Prevalence of sexual abuse and factors associated with sexual abuse. *Journal of the American Academy of Child and Adolescent Psychiatry*, **35**(10), 1355–1364.
- Fergusson D.M., Mullen P.E. (1999) *Childhood Sexual Abuse – An Evidence Based Perspective*, Thousand Oaks, CA, Sage.
- Finkelhor D. (1990) Early and long-term effects of child sexual abuse: An update. *Professional Psychology: Research and Practice*, **21**(5), 325–330.
- Friedrich W.N., Tally N.J., Panser L., Fett S., Zinsmeister A.R. (1997) Concordance of reports of childhood abuse by adults. *Child Maltreatment*, **2**(2), 164–171.
- Fry R.P.W., Rozewicz L.M., Crisp A.H. (1996) Interviewing for sexual abuse: Reliability and effect of interviewer gender. *Child Abuse and Neglect*, **20**(8), 725–729.
- Holmes W.C., Slap G.B. (1998) Sexual abuse of boys: Definition, prevalence, correlates, sequelae, and management. *Journal of the American Medical Association*, **280**(21), 1855–1862.
- Johnson C.F. (2004) Child sexual abuse. *Lancet*, **364**(9432), 462–470.
- Joreskog K.G., Sorbom D. (1993) *LISREL 8: Structural Equation Modelling with the SIMPLIS Command Language*, Chicago, IL, Scientific Software International.
- Kaplan S.J., Pelcovitz D., Labruna V. (1999) Child and adolescent abuse and neglect research: A review of the past 10 years. Part I: Physical and emotional abuse and neglect. *Journal of the American Academy of Child and Adolescent Psychiatry*, **38**(10), 1214–1222.
- Kooiman C.G., Ouwehanda A.W., ter Kuile M.M. (2002) The Sexual and Physical Abuse Questionnaire (SPAQ): A screening instrument for adults to assess past and current experiences of abuse. *Child Abuse and Neglect*, **26**(9), 939–953.
- Leserman J., Drossman D.A., Zhiming L. (1995) The reliability and validity of a sexual and physical abuse history questionnaire in female patients with gastrointestinal disorders. *Behavioral Medicine*, **21**(3), 141–150.
- Lipschitz D.S., Bernstein D.P., Winegar R.K., Southwick S.M. (1999) Hospitalized adolescents' reports of sexual and physical abuse: A comparison of two self-report measures. *Journal of Traumatic Stress*, **12**(4), 641–654.
- Martin J., Anderson J., Romans S., Mullen P., O'Shea M. (1993) Asking about child sexual abuse: Methodological implications of a two stage survey. *Child Abuse and Neglect*, **17**(3), 383–392.
- Maughan B., Rutter M. (1997) Retrospective reporting of childhood adversity: Issues in assessing long-term recall. *Journal of Personality Disorders*, **11**(1), 19–33.
- Meyer I.H., Muenzenmaier K., Cancienne J., Struening E. (1996) Reliability and validity of a measure of sexual and physical abuse histories among women with serious mental illness. *Child Abuse and Neglect*, **20**(3), 213–219.
- Mills K., Teeson M., Darke S., Ross J. (2007) Reliability of self-reported trauma exposure among people with heroin dependence: A

Acknowledgements

This research was funded by grants from the Health Research Council of New Zealand, the National Child Health Research Foundation, the Canterbury Medical Research Foundation and the New Zealand Lottery Grants Board.

Declaration of interest statement

The authors have no competing interests.

- longitudinal investigation. *Journal of Traumatic Stress*, **20**(3), 313–323.
- Mullen P.E., King N.J., Tonge B.J. (2000) Child sexual abuse: An overview. *Behaviour Change*, **17**(1), 2–14.
- Muthén, L.K., Muthén, B.O. (2007) *Mplus Users' Guide* (5th ed.), Los Angeles, CA, Muthén and Muthén.
- National Clearinghouse on Child Abuse and Neglect Information (2005) Long-term consequences of child abuse and neglect. http://nccanch.acf.hhs.gov/pubs/factsheets/long_term_consequences.cfm [23 August 2005].
- Newton A.W., Vandeven A.M. (2008) Update on child maltreatment. *Current Opinion in Pediatrics*, **20**(2), 205–212.
- Putnam F.W. (2003) Ten-year research update review: Child sexual abuse. *Journal of the American Academy of Child and Adolescent Psychiatry*, **42**(3), 269–278.
- Raphael K.G. (1987) Recall bias: A proposal for assessment and control. *International Journal of Epidemiology*, **16**(2), 167–169.
- Stirling J. Jr, Amaya-Jackson L., Amaya-Jackson L. (2008) Understanding the behavioral and emotional consequences of child abuse. *Pediatrics*, **122**(3), 667–673.
- Tajima E.A., Herrenkohl T.I., Huang B., Whitney S.D. (2004) Measuring child maltreatment: A comparison of prospective parent reports and retrospective adolescent reports. *American Journal of Orthopsychiatry*, **74**(4), 424–435.
- White H.R., Widom C.S., Chen P.H. (2007) Congruence between adolescents' self-reports and their adult retrospective reports regarding parental discipline practices during their adolescence. *Psychological Reports*, **101**(3, part 2), 1079–1094.
- Widom C.S., Morris S. (1997) Accuracy of adult recollections of childhood victimization, Part 2: Childhood sexual abuse. *Psychological Assessment*, **9**(1), 34–46.
- Widom C.S., Raphael K.G., DuMont K.A. (2004) The case for prospective longitudinal studies in child maltreatment research: Commentary on Dube, Williamson, Thompson, Felitti, and Anda. *Child Abuse and Neglect*, **28**(7), 715–722.
- Widom C.S., Shepard R.L. (1996) Accuracy of adult recollections of childhood victimization: Part 1. Childhood physical abuse. *Psychological Assessment*, **8**(4), 412–421.
- Widom C.S., Weiler B.L., Cottler L.B. (1999) Childhood victimization and drug abuse: A comparison of prospective and retrospective findings. *Journal of Consulting and Clinical Psychology*, **67**(6), 867–880.
- Williams L.M. (1994) Recall of childhood trauma: A prospective study of women's memories of child sexual abuse. *Journal of Consulting and Clinical Psychology*, **62**(6), 1167–1176.
- Wissow L.S. (1995) Child abuse and neglect. *The New England Journal of Medicine*, **332**(21), 1425–1431.
- World Health Organization. (1993) *Composite International Diagnostic Interview (CIDI)*, Geneva, World Health Organization.