Psychological Medicine, 2004, **34**, 1475–1482. © 2004 Cambridge University Press
DOI: 10.1017/S003329170400265X Printed in the United Kingdom

Childhood sexual abuse, stressful life events and risk for major depression in women

KENNETH S. KENDLER*, JONATHAN W. KUHN AND CAROL A. PRESCOTT

Department of Psychiatry, Medical College of Virginia/Virginia Commonwealth University, Richmond, VA, USA; Department of Human Genetics, Medical College of Virginia/Virginia Commonwealth University, Richmond, VA, USA; Virginia Institute for Psychiatric and Behavioral Genetics, Richmond, VA, USA

ABSTRACT

Background. In animals, early trauma can produce long-lasting changes in sensitivity to the pathogenic effects of stress. To explore whether similar processes occur in humans, we examine whether childhood sexual abuse (CSA) in women alters sensitivity in adulthood to the depressogenic effects of stressful life events (SLEs).

Method. A history of CSA was obtained from a population-based sample of 1404 female adult twins. Cox Proportional hazard models were used to predict onsets of episodes of DSM-III-R major depression (MD) in the past year from previously assessed levels of neuroticism (N), CSA and past-year SLEs scored on long-term contextual threat.

Results. In the best-fit model, onset of MD was predicted by CSA, SLEs and N. Individuals with CSA (and especially with severe CSA) had both an overall increased risk for MD and a substantially increased sensitivity to the depressogenic effects of SLEs. A 'dose–response' relationship between severity of CSA and sensitivity to SLEs was clearer in those with low to average levels of N than in those with high levels of N.

Conclusion. As documented with physiological responses to a standardized laboratory stressor, CSA increases stress sensitivity in women in a more naturalistic setting. Both genetic and early environmental risk factors can produce long-term increase in the sensitivity of individuals to depressogenic life experiences.

INTRODUCTION

Sensitivity to the pathogenic effects of stress constitutes a critical risk factor for a number of psychiatric disorders including major depression (MD) (Kessler, 1997). Previous studies in our population-based twin sample of women have demonstrated that genetic risk factors (Kendler *et al*. 1995) and the personality trait of neuroticism (N) (Kendler *et al*. 2004) alter sensitivity to the depressogenic effects of stressful life events (SLEs). In rodents and non-human primates, exposure to certain stressors early in life increases the sensitivity of the organism to later adversities (Heim & Nemeroff, 2001; Sanchez *et al*. 2001). It is of both practical and theoretical interest to determine whether similar effects can also be demonstrated in humans.

Childhood sexual abuse (CSA) is a plausible candidate for such an early stressor. Women with a history of CSA are at increased risk for MD in adulthood (Burnam *et al*. 1988; Bifulco *et al*. 1991; Mullen *et al*. 1993; Fergusson *et al*. 1996; Fergusson & Mullen, 1999; Kendler *et al*. 2000), and this relationship is likely to be largely causal (Kendler *et al*. 2000; Nelson *et al*. 2002). Furthermore, CSA in women is associated with an increased autonomic and hypothalamic–pituitary–adrenal (HPA) response to controlled laboratory stress (Heim *et al*. 2000). In this report, we explore whether the hyper-responsiveness

* Address for correspondence: Dr K. S. Kendler, Medical College of Virginia, Box 980126, Richmond, VA 23298-0126, USA.

to stress of women with CSA is detectable in a more naturalistic setting involving exposure to SLEs and onset of depressive episodes.

METHOD

Sample

Participating twins came from on-going investigations of female–female twin pairs from the population-based Virginia Twin Registry, the details of which have been outlined elsewhere (Kendler *et al.* 1992; Kendler & Prescott, 1998, 1999). The pairs derive from two related samples, born from 1934 to 1974, which became eligible to participate if both members responded to a mailed questionnaire, the response rate to which was ~64%. Eighty-eight per cent of our sample was first interviewed face to face during 1987–1989 [at which time their mean age (S.D.) and range was 30·1 ± 7·6 (17–55) years] and has been the subject of three additional telephone interview waves. The last of these waves was completed during 1995–1997, an average (± S.D.) of 92 ± 7 months after their first assessment. The rest were first interviewed face to face during 1992–1994 and assessed a second time (with the same interview given to the rest of the sample on the fourth wave) by telephone during 1996–1997. As approved by the VCU IRB, written informed consent was obtained in this study for all face-to-face interviews, while verbal consent was obtained for telephone interviews.

Measures

During each interview wave, we assessed the occurrence over the last year of 14 individual symptoms representing the disaggregated nine 'A Criteria' for MD in DSM-III-R (APA, 1987) (e.g. two items for criterion A4 to assess insomnia *versus* hypersomnia). For each reported symptom, interviewers probed to ensure that it was due neither to physical illness nor medication. The respondents and interviewers then aggregated last year's symptoms reported into syndromes formed of co-occurring symptoms. If depressive syndromes occurred, respondents were asked for the months of their onset and offset of all episodes in the last year. The diagnosis of MD was made by computer algorithm incorporating the DSM-III-R criteria, except criterion B2 (which excludes depressive syndromes considered to be 'uncomplicated bereavement'). In 375 twins who were interviewed twice by different interviewers with a mean (S.D.) inter-interview interval of 30 (9) days, the inter-interview reliability of the diagnosis of MD in the last year was good: κ (Cohen, 1960) = +0·68 (95% CIs 0·57–0·80), tetrachoric correlation = +0·92 (0·86–0·98).

N was measured using the 12-item scale from the shortened EPQ (Eysenck *et al.* 1985). In this study, we take the levels of N reported from wave 1 interviews. Our interviews assessed the occurrence, to the nearest month, of 11 '*personal*' *stressful life events*, (events occurring primarily to the informant): 'assault' (assault, rape or mugging), 'divorce/separation' (divorce, marital separation, broken engagement or break up of other romantic relationship), 'major financial problem', 'serious housing problems', 'serious illness or injury', 'job loss' (laid off from a job or fired), 'legal problems' (trouble with the police or other legal trouble), 'loss of confidant' (separation from other loved one or close friend other than spouse/partner), 'serious marital problems' (involving a marital or marriage-like intimate, cohabiting relationship), 'robbed' and 'serious difficulties at work'. We also assessed four classes of '*network*' *events*, meaning events that occur primarily to, or in interaction with, an individual in the respondent's social network. These event classes consisted of (i) 'getting along with' – serious trouble getting along with an individual in the network, (ii) 'crisis' – a serious personal crisis of someone in the network, (iii) 'death' – death of an individual in the network and (iv) 'illness' – serious illness of someone in the network.

Each reported SLE in waves 3 and 4 was rated by the interviewer on the level of long-term contextual threat (LTCT). Using the criteria proposed by Brown (1989), we trained our interviewers to rate

> what most people would be expected to feel about an event in a particular set of circumstances and biography, taking no account either of what the respondent says about his or her reaction or about any psychiatric or physical symptoms that followed it (Brown, 1989, p. 24).

Interviewers were instructed to ask several structured probes after each reported SLE. For example, if a respondent reported a serious accident in the last year, the script for the

https://doi.org/10.1017/S003329170400265X Published online by Cambridge University Press

interviewers read

Briefly, tell me what happened … In what way did you life change because of this serious accident? Did the accident cause any financial problems for you? Were there any other ways in which you expected the accident to affect you?

The interviewers were instructed to continue inquiring until they had enough information to rate LTCT. After Brown (Brown & Harris, 1978; Brown, 1989), LTCT was rated on a 4-point scale: minor, low moderate, high moderate and severe. While our rating of LTCT adopted the general approach of Brown and coworkers (Brown, 1989), we did not adopt his method of rating threat levels by a blinded panel of expert judges, but instead had the interviewer perform the assessment. The sample size of this study made the panel-based method of assessment unfeasible. Reliability of our ratings of LTCT was determined by both an inter-rater and test–rest design. Inter-rater reliability was assessed by having experienced interviewers, who were blind to all other aspects of the assessment, review tape recordings of the interview sections in which 92 randomly selected individual SLEs were evaluated. Test–retest reliability was obtained by repeating the interview with 191 respondents at a mean interval of 4 weeks. We obtained 173 scored life events that were reported to have occurred within 1 month of one another and we assumed represented the same event. We assessed reliability on these 4-point scales by Spearman correlation (r_s) and weighted kappa (κ) (Fleiss *et al*. 1969). The test–retest reliability for LTCT was $r_s = +0{\cdot}60$ and $\kappa = +0{\cdot}41$, while inter-rater reliability was $r_s = +0{\cdot}69$ and $\kappa = +0{\cdot}67$. We also obtained ratings of dependence/independence of all reported SLEs but did not use these ratings in these analyses.

During our second wave interview, we examined the willingness of twins to answer questions about CSA and their preferred method of assessment. Most preferred a mailed questionnaire, which was employed using items developed by Mullen and colleagues (Martin *et al*. 1993). Our initial item was:

Before you were 16, did any adult, or any other person older than yourself, involve you in any unwanted incidents like (i) inviting or requesting you to do something sexual, (ii) kissing or hugging you in a sexual way, (iii) touching or fondling your private parts, (iv) showing their sex organs to you, (v) making you touch them in a sexual way or (vi) attempting or having sexual intercourse.

With respect to CSA exposure, we divided our sample into four exclusive, hierarchical categories: (1) no CSA (69·6%), (2) *non-genital* CSA [numbers (i), (ii) and (iv); 7·8%], (3) *genital* CSA (genital contact but no intercourse) [numbers (iii) and (v); 14·1%] and (4) *intercourse* [number (vi); 8·4%]. In this paper, we refer to non-genital, genital and intercourse CSA as respectively, mild, moderate and severe. We also asked twins to report on CSA in their co-twin. While the association between self and co-twin reported CSA was highly significant in this sample ($\chi^2 = 300{\cdot}1$, df = 9, $p < 0{\cdot}0001$), the level of agreement was modest [contingency coefficient $= +0{\cdot}50$, weighted $\kappa = +0{\cdot}40$ (95% CI $+0{\cdot}33$ to $+0{\cdot}47$)].

Statistical methods

In this sample of women, CSA is associated with higher levels of N (Kendler *et al*. 2002) and N is associated with greater sensitivity to the depressogenic effects of SLEs (Kendler *et al*. in press). N was therefore included in our model to (i) permit us to discriminate between a direct effect of CSA on stress sensitivity *versus* a more indirect effect mediated through personality and (ii) to determine if the relationship between CSA, SLE and risk for depressive onsets was modified by levels of N.

Using person-months as the unit of analysis, these analyses were conducted with a Cox proportional hazards model operationalized in the SAS procedure PHREG (Cox, 1972, 1995). In our full model, three predictor variables were used: CSA, N and LTCT. When multiple events occurred in the same month, LTCT was coded as the highest threat level of any recorded event. The dependent variable was the onset of a depressive episode.

The final model was developed based on four strata. Each stratum consisted of data for subjects with a given number of prior onsets – specifically zero or one prior onset in the past 13 months – for subjects from waves 3 and 4. There were too few subjects with two or more onsets to include such data. This stratification is a conservative way to deal with within-subject correlation. An 18-strata model was also developed in which twins were randomly assigned

https://doi.org/10.1017/S003329170400265X Published online by Cambridge University Press

into two separate groups to conservatively evaluate the impact of familial correlations. Coding was done based on the 'Conditional A' model proposed by Hosmer & Lemeshow (1999, pp. 308–317).

LTCT was coded so that 0 meant no SLE occurrence in the month and 1–4 meant the occurrence of a SLE with minor, low moderate, high moderate and severe LTCT. To incorporate the ordinal structure and simplify interpretation of the interaction, LTCT was coded as follows: Four dummy variables X1, X2, X3 and X4 were used. If there was no life event, all four were coded as zero. If there was a significant life event with LTCT 1 or higher, X1 was coded as 1. If LTCT was $\geqslant 2$, X2 was also coded as 1. If LTCT was $\geqslant 3$, then X3 was coded as 1. For an event with LTCT = 4, all four dummy variables were coded as 1. Thus, the coding for a month with an event with a LTCT = 2 was: X1 = 1, X2 = 1, X3 = 0, X4 = 0. This method of dummy variable coding is often referred to as thermometer coding (Masters, 1993). Finally, these codings were incorporated as time-dependent covariates with a linear decay which abated after 3 months. This was done by multiplying the indicator by 4/3 for the month of a stressful life event, by 2/3 for the month after the stressful life event and 0 for 3 or more months after the event.

Thermometer coding was also used for CSA. The four levels, coded into three indicator variables, were for no, mild, moderate and severe CSA. N was standardized to have a mean of zero and a standard deviation of 1, allowing easy interpretation and a meaningful quadratic term. The final model was found by elimination of non-significant terms starting with all three-way interactions.

In these analyses, the dependent variable – a depressive episode – is dichotomous. Our goal is to clarify the nature of the interaction between CSA and SLEs in the prediction of depression and *from a statistical perspective, any interaction is scale-dependent*. The use of a Cox regression model has numerous advantages in these analyses but instead of predicting the probability of a depressive onset, the model predicts a logarithmic transformation of this probability. By including this logarithmic function in the dependent variable, the nature of what an interaction means has changed. What is a multiplicative interaction in a probability model becomes additive in the Cox model, while what is additive in the probability model becomes a negative interaction in the Cox model.

Information for these analyses comes from a total of 1404 individuals who participated in waves 3 and 4 of the study of female–female twin pairs. These individuals reported a total of 265 onsets of MD and 2811 periods of observation where each period of observation either begins at the start of a 1-year prevalence window or the month following a recovery from an episode and ends either at the conclusion of that 1-year window or at the time of an onset of a depressive episode. The number of these onsets that occurred with zero and one prior episode in the 13-month time period were respectively: 185 (69·8%), and 80 (30·2%).

We started with a Cox model with 36 terms: four LTCT indicator variables, three CSA indicator variables, sex, N, N^2, and all two-way interactions between them. Non-significant terms (i.e. $p > 0{\cdot}05$) were removed one at a time (eliminating interactions first, and then main effects not involved in interactions) and the model rerun. After 29 eliminations, the final model, with seven terms, was reached. The elimination process was repeated, beginning with several other arbitrarily chosen terms to verify that the resulting seven-term model was the only all-significant term subset of the full 36-term model.

RESULTS

Results from the best-fit Cox model for predicting onset of MD are seen in Table 1 and include strong and significant main effects for N, CSA and LTCT. The model detected no significant difference in risk between those with minor *versus* low moderate LTCT or between those with mild *versus* moderate CSA.

In this model, each increase of a standard deviation of N carries a hazard ratio (HR) for a depressive onset of 1·37. The two meaningful levels of severity of CSA carried further risks for an onset of MD (assessed as a HR) with similar HRs of $\sim 1{\cdot}6$. Since these ratios multiply, that compared to those with no history of CSA, the HR for a depressive onset associated with mild or moderate CSA is 1·62 and with severe CSA is 2·66 (i.e. $1{\cdot}62 \times 1{\cdot}64$). The three increasing levels of LTCT each carried an additional risk for a

https://doi.org/10.1017/S003329170400265X Published online by Cambridge University Press

Table 1. *Parameter estimates for best fit model predicting risk of onset of an episode of major depression in women as a function of history of childhood sexual abuse (CSA), stressful life event exposure (quantified using level of long-term contextual threat [LTCT]) and level of the personality trait of neuroticism*

Variable	Hazard ratio	95% CI	χ^2 (df = 1)	*p* value
Neuroticism	1·37	1·21–1·56	24·04	< 0·0001
Mild and moderate CSA	1·62	1·23–2·15	11·36	0·0008
Severe CSA	1·64	1·13–2·39	6·72	0·01
Mild or low moderate LTCT	1·89	1·32–2·71	11·90	0·0006
High moderate LTCT	2·05	1·32–3·19	10·25	0·001
Severe LTCT	2·44	1·44–4·15	10·85	0·001
Neuroticism × severe CSA	0·67	0·50–0·91	6·64	0·01

depressive onset with HRs ranging from 1·89 to 2·44. Compared to months with no life-event exposure, the HR for a depressive onset associated with a SLE with a LTCT rating of minor or low moderate, high moderate and severe can be estimated to be equal to 1·89, 3·87 (i.e. 1·89 × 2·05) and 9·45 respectively.

In addition to these main effects, the final model contains one interaction term – a negative interaction between N and severe CSA. This interaction means that the joint impact on risk for MD of severe CSA and high levels of N is less than predicted by their main effects. No negative interactions, however, were seen between CSA and LTCT.

The predictions of the best-fit model are illustrated in Fig. 1. We assign, as a point of reference, an HR of unity to a woman with a mean level of N, no history of CSA and no exposure to a SLE (meaning an LTCT value of zero). The figure depicts the relationship between a history of CSA, exposure to LTCT and risk for a depressive onset in women with low levels of N (–1 S.D. from the mean), average levels of N (mean N), and high levels of N (+1 S.D. from the mean). The slope of the lines seen in the figure – which reflects the gain in risk for MD with increasing levels of LTCT – index directly *stress sensitivity*. The *steeper* this curve is, the greater is the sensitivity to the depressogenic effects of SLEs.

The figure clearly illustrates the main effects of N, CSA and LTCT in the prediction of depressive onsets. That is, the risk for MD becomes

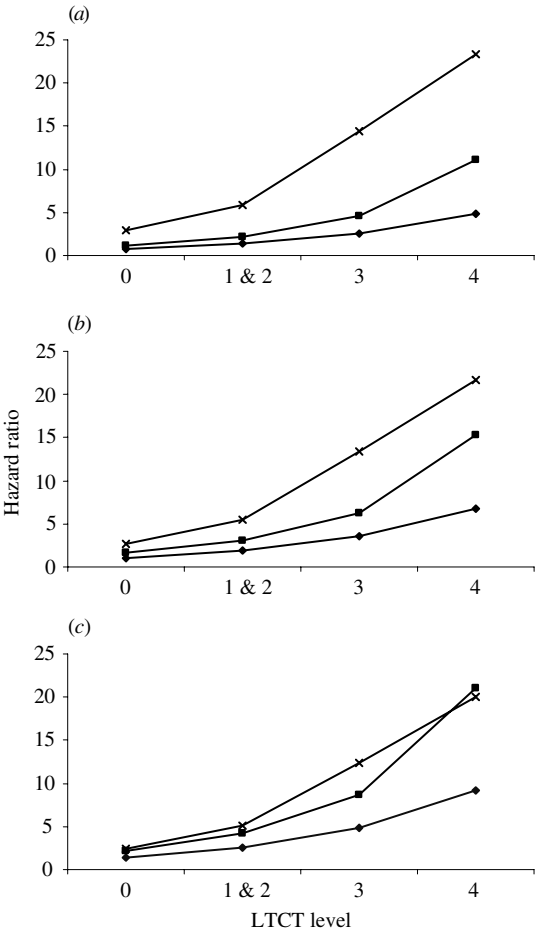

FIG. 1. The hazard ratio for onset of an episode of major depression, defined by DSM-III-R (APA, 1987), as a function of level of the personality trait neuroticism, exposure to a stressful life event – rated by level of long-term contextual threat (LTCT) – and prior history of childhood sexual abuse (CSA). CSA 1 and 2 mean a history of mild or moderate CSA (see text for definition) and CSA 3 means a history of severe CSA. (*a*) The relationship for those with low neuroticism (defined as 1 S.D. below the mean). (*b*) The relationship for those with average neuroticism. (*c*) The relationship for those with high neuroticism (defined as 1 S.D. above the mean). The hazard ratio is standardized so that a value of unity reflects the risk of depressive onset for an individual of average neuroticism, with no CSA and without exposure to any SLE – that is an LTCT rating of 0. –◆–, No CSA; –■–, CSA 1 and 2; –×–, CSA 3.

greater as the level of LTCT or N increase and is substantially higher in those with *versus* those without a history of CSA. The figure also demonstrates – at least for those with no, or mild CSA – increasing stress sensitivity with increasing levels of N. For example, the slope of the curve for those with no CSA is considerably

https://doi.org/10.1017/S003329170400265X Published online by Cambridge University Press

steeper for those with high N than for those with low levels of N.

More interestingly, however, is that, in addition to all these effects, the model presents clear evidence that the slope of the lines are substantially greater in those with *versus* those without a history of CSA. In women with average or low levels of N, the 'dose–response' relationship is particularly clear. The level of stress sensitivity is relatively modest in those with no history of CSA, intermediate in those with mild or moderate CSA and high in those with exposure to severe CSA. In those with high levels of N, by contrast, the results indicate that the level of stress sensitivity is approximately equally elevated in those with mild/moderate and severe CSA.

COMMENT

The goal of this paper was to determine whether adult women with a history of sexual abuse in childhood demonstrated increased stress sensitivity in a naturalistic setting. Our best-fitting Cox proportional hazards model demonstrated that woman with CSA were substantially more sensitive to the depressogenic effects of SLEs than were women with no such history. Furthermore, the greatest increase in stress sensitivity were seen in women exposed to the most severe form of abuse.

We included, as a control variable, the personality trait of N which was significantly predicted by CSA (Kendler *et al*. 2002). Furthermore, in accord with prior findings (Kendler *et al*. in press), our best-fit model demonstrated that N also influenced the sensitivity to the depression-inducing effects of adversity. By including N in the model, we can conclude that the observed impact of CSA on stress sensitivity is *not* being mediated through this major personality dimension of neuroticism/emotionality.

One unanticipated result emerged from our analyses. The 'dose–response' relationship between severity of CSA and stress sensitivity, while clearly seen in those with average and low levels of N, was obscured in those with high N. This occurred because of a relative reduction in the impact of severe CSA on risk for MD in those who are highly 'neurotic'. We are uncertain about the replicability of this finding and so do not wish to speculate extensively about its significance. Two plausible interpretations are (i) a 'ceiling effect' in that the high risk for MD with elevated levels of N has less 'room' for the large impact of severe CSA and/or (ii) an increasing sensitivity to the effects of mild and moderate CSA in those with higher levels of N.

It is of interest to view these results in the context of several prior findings. First, when subjected to a standard laboratory psychosocial stressor of public speaking and mental arithmetic before an audience, women with a history of childhood physical or sexual abuse displayed an augmented ACTH, cortisol and heart-rate response (Heim *et al*. 2000). Second, women with a history of childhood sexual or physical abuse demonstrate hippocampal shrinkage (Bremner *et al*. 1997; Stein *et al*. 1997; Vythilingam *et al*. 2002). While these changes might result directly from trauma, they could also arise from chronic HPA over-activation. Third, a large body of work in rodents and non-human primates shows that certain early environmental stressors augment later behavioral and HPA axis sensitivity to stressors (Heim & Nemeroff, 2001). Fourth, other forms of early childhood stress, especially parental loss, may also increase sensitivity to the depressogenic effects of SLEs (Brown & Harris, 1978; Bifulco *et al*. 1987).

Along with these prior findings, the current results suggest that across mammalian species, certain early adversities may produce long-lasting psychobiological changes which alter the sensitivity of the organism to stressors. While not addressed in the current study, the HPA axis and associated brain CRF systems may play a critical role in this process.

It might seem contradictory to claim that CSA alters sensitivity to the effects of SLEs when our best-fit model indicates main effects for CSA and SLEs with no CSA–SLE interaction. This occurs because we are fitting a Cox model which assumes a multiplicative relationship between variables. So what is additive in the Cox model is multiplicative on the scale of probability (or HR) of risk. Had there been no interaction between CSA and SLE in the prediction of MD on the probability scale, the Cox model would have produced a *negative* interaction between the two variables, evidence for which was not found in our analyses.

We have previously demonstrated for MD what we have termed *genetic control of sensitivity*

https://doi.org/10.1017/S003329170400265X Published online by Cambridge University Press

to the environment (Kendler *et al*. 1995). That is, genetic risk factors for MD appear to *both* increase average risk for onset and render individuals more sensitive to the pathogenic effects of SLEs. Our current study suggests that a pathogenic early environmental event – CSA – can act similarly. Just as the overall risk for MD may arise from a complex developmental interplay of genetic and environmental risk factors (Kendler *et al*. 2002), so sensitivity to the depressogenic effects of SLEs may be influenced by both genes and prior environmental experiences.

Limitations

These results should be interpreted in the context of four potentially significant methodological limitations. First, these findings were based on a single sample of white female twins from one geographical region. Although the rates in our sample of both CSA (Fergusson & Mullen, 1999) and 1-year prevalence for MD (Kessler *et al*. 1994) are similar to those reported elsewhere, our findings may or may not extrapolate to other groups. Second, our analyses assumed that when SLEs occurred in the same month as depressive onsets, the SLE preceded the onset. We have explored this issue in two prior analyses (Kendler *et al*. 1995, 1998) both of which support the validity of this assumption. Third, CSA is often associated with a range of other risk factors for subsequent psychiatric illness such as parental psychopathology, physical abuse and parent–child conflict (Fergusson & Mullen, 1999). In our prior study (Kendler *et al*. 2000), we found little change in the association between CSA and *lifetime* MD in this sample when detailed measures of these potentially associated risk factors were added as covariates. We added these covariates (including measures of parent–child relationship, family environment and financial status, church attendance and parental disciplinary practices; see Kendler *et al*. 2000 for details) to our current model and this produced no substantial change in findings. Fourth, as we have shown elsewhere (Kendler *et al*. 2002), the impact of CSA on risk for MD is probably mediated through a range of personal and psychosocial variables. For the sake of simplicity, these potential mediators were not included in these analyses.

ACKNOWLEDGEMENTS

This work was supported by NIH grant MH/AA/DA-49492. We acknowledge the contribution of the Virginia Twin Registry, now part of the Mid-Atlantic Twin Registry (MATR), to ascertainment of subjects for this study. The MATR, directed by Dr L. Corey and Dr L. Eaves, has received support from the National Institutes of Health, the Carman Trust and the WM Keck, John Templeton and Robert Wood Johnson Foundations. We thank Patsy Waring, Frank Butera, Sarah Woltz, Barbara Brooke and Lisa Halberstadt for supervising data collection, and Indrani Ray and Steven Aggen for database assistance.

DECLARATION OF INTEREST

None.

REFERENCES

APA (1987). *Diagnostic and Statistical Manual of Mental Disorders* (revised 3rd edn). American Psychiatric Association: Washington, DC.

Bifulco, A., Brown, G. B. & Adler, Z. (1991). Early sexual abuse and clinical depression in adult life. *British Journal of Psychiatry* **159**, 115–122.

Bifulco, A. T., Brown, G. W. & Harris, T. O. (1987). Childhood loss of parent, lack of adequate parental care and adult depression: a replication. *Journal of Affective Disorders* **12**, 115–128.

Bremner, J. D., Randall, P., Vermetten, E., Staib, L., Bronen, R. A., Mazure, C., Capelli, S., McCarthy, G., Innis, R. B. & Charney, D. S. (1997). Magnetic resonance imaging-based measurement of hippocampal volume in posttraumatic stress disorder related to childhood physical and sexual abuse – a preliminary report. *Biological Psychiatry* **41**, 23–32.

Brown, G. W. (1989). Life events and measurement. In *Life Events and Illness* (ed. G. W. Brown and T. O. Harris), pp. 3–45. The Guilford Press: New York.

Brown, G. W. & Harris, T. O. (1978). *Social Origins of Depression: A Study of Psychiatric Disorder in Women*. Tavistock: London.

Burnam, M. A., Stein, J. A., Golding, J. M., Siegel, J. M., Sorenson, S. B., Forsythe, A. B. & Telles, C. A. (1988). Sexual assault and mental disorders in a community population. *Journal of Consulting and Clinical Psychology* **56**, 843–850.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* **20**, 37–46.

Cox, D. R. (1972). Regression models and life tables (with discussion). *Journal of the Royal Statistical Society* (series B) **34**, 187–220.

Cox, D. R. (1995). *Survival Analysis Using The SAS System: A Practical Guide* (2nd edn) (ed. P. D. Allison). SAS Institute Inc.: Cary, NC.

Eysenck, S. B. G., Eysenck, H. J. & Barrett, P. (1985). A revised version of the psychoticism scale. *Personality and Individual Differences* **6**, 21–29.

Fergusson, D. M., Horwood, L. J. & Lynskey, M. T. (1996). 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**, 1365–1374.

Fergusson, D. M. & Mullen, P. E. (1999). *Childhood Sexual Abuse: An Evidence Based Perspective*. Sage Publications, Inc.: Thousand Oaks, CA.

Fleiss, J. S., Cohen, J. & Everett, B. S. (1969). Large sample standard errors of kappa and weighted kappa. *Psychological Bulletin* **72**, 323–327.

Heim, C. & Nemeroff, C. B. (2001). The role of childhood trauma in the neurobiology of mood and anxiety disorders: preclinical and clinical studies. *Biological Psychiatry* **49**, 1023–1039.

Heim, C., Newport, D. J., Heit, S., Graham, Y. P., Wilcox, M., Bonsall, R., Miller, A. H. & Nemeroff, C. B. (2000). Pituitary-adrenal and autonomic responses to stress in women after sexual and physical abuse in childhood. *Journal of the American Medical Association* **284**, 592–597.

Hosmer, D. W. & Lemeshow, S. (1999). *Applied Survival Analysis: Regression Modeling of Time to Event Data*. John Wiley & Sons: New York.

Kendler, K. S., Bulik, C. M., Silberg, J. L., Hettema, J. M., Myers, J. & Prescott, C. A. (2000). Childhood sexual abuse and adult psychiatric and substance use disorders in women: an epidemiological and Cotwin control analysis. *Archives of General Psychiatry* **57**, 953–959.

Kendler, K. S., Gardner, C. O. & Prescott, C. A. (2002). Toward a comprehensive developmental model for major depression in women. *American Journal of Psychiatry* **159**, 1133–1145.

Kendler, K. S., Karkowski, L. & Prescott, C. A. (1998). Stressful life events and major depression: risk period, long-term contextual threat and diagnostic specificity. *Journal of Nervous and Mental Disease* **186**, 661–669.

Kendler, K. S., Kessler, R. C., Walters, E. E., MacLean, C. J., Sham, P. C., Neale, M. C., Heath, A. C. & Eaves, L. J. (1995). Stressful life events, genetic liability and onset of an episode of major depression in women. *American Journal of Psychiatry* **152**, 833–842.

Kendler, K. S., Kuhn, J. & Prescott, C. A. (2004). Interrelationship of neuroticism, sex and stressful life events in the prediction of episodes of major depression. *American Journal of Psychiatry* **161**, 631–636.

Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C. & Eaves, L. J. (1992). A population-based twin study of major depression in women: the impact of varying definitions of illness. *Archives of General Psychiatry* **49**, 257–266.

Kendler, K. S. & Prescott, C. A. (1998). Cannabis use, abuse and dependence in a population-based sample of female twins. *American Journal of Psychiatry* **155**, 1016–1022.

Kendler, K. S. & Prescott, C. A. (1999). A population-based twin study of lifetime major depression in men and women. *Archives of General Psychiatry* **56**, 39–44.

Kessler, R. C. (1997). The effects of stressful life events on depression. *Annual Review of Psychology* **48**, 191–214.

Kessler, R. C., McGonagle, K. A., Zhao, S., Nelson, C. B., Hughes, M., Eshleman, S., Wittchen, H.-U. & Kendler, K. S. (1994). Lifetime and 12-month prevalence of DSM-III-R psychiatric disorders in the United States: results from the National Comorbidity Survey. *Archives of General Psychiatry* **51**, 8–19.

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 & Neglect* **17**, 383–392.

Masters, T. (1993). *Practical Neural Network Recipes in C++*. Academic Press: Boston, MA.

Mullen, P. E., Martin, J. L., Anderson, J. C., Romans, S. E. & Herbison, G. P. (1993). Childhood sexual abuse and mental health in adult life. *British Journal of Psychiatry* **163**, 721–732.

Nelson, E. C., Heath, A. C., Madden, P. A. F., Cooper, M. L., Dinwiddie, S. H., Bucholz, K. K., Glowinski, A., McLaughlin, T., Dunne, M. P., Statham, D. J. & Martin, N. G. (2002). Association between self-reported childhood sexual abuse and adverse psychosocial outcomes: results from a twin study. *Archives of General Psychiatry* **59**, 139–145.

Sanchez, M. M., Ladd, C. O. & Plotsky, P. M. (2001). Early adverse experience as a developmental risk factor for later psychopathology: evidence from rodent and primate models. *Developmental Psychopathology* **13**, 419–449.

Stein, M. B., Koverola, C., Hanna, C., Torchia, M. G. & McClarty, B. (1997). Hippocampal volume in women victimized by childhood sexual abuse. *Psychological Medicine* **27**, 951–959.

Vythilingam, M., Heim, C., Newport, J., Miller, A. H., Anderson, E., Bronen, R., Brummer, M., Staib, L., Vermetten, E., Charney, D. S., Nemeroff, C. B. & Bremner, J. D. (2002). Childhood trauma associated with smaller hippocampal volume in women with major depression. *American Journal of Psychiatry* **159**, 2072–2080.