Life stress, 5-HTTLPR and mental disorder: findings from a 30-year longitudinal study

David M. Fergusson, L. John Horwood, Allison L. Miller and Martin A. Kennedy

Background

Recent meta-analyses have raised concerns about the replicability of gene \times \text{environment} interactions involving the serotonin transporter gene (5-HTTLPR) in moderating the associations between adverse life events and mental disorders.

Aims

To use data gathered over the course of a 30-year longitudinal study of a New Zealand birth cohort to test the hypothesis that the presence of short (‘s’)/alleles of 5-HTTLPR are associated with an increased response to life stress.

Method

Participants were 893 individuals from the Christchurch Health and Development Study who had complete data on: the 5-HTTLPR genotype; psychiatric disorders up to the age of 30; and exposure to childhood and adult adverse life events.

Results

A series of 104 regression models were fitted to four mental health outcomes (depressive symptoms, major depression, anxiety disorder and suicidal ideation) observed at ages 18, 21, 25 and 30 using 13 measures of life-course stress that spanned childhood and adult stressors. Both multiplicative and additive models were fitted to the data. No evidence was found that would support the hypothesis that ‘s’ alleles of 5-HTTLPR are associated with increased responsiveness to life stressors.

Conclusions

The present findings add to the evidence suggesting that it is unlikely that there is a stable gene \times \text{environment} interaction involving 5-HTTLPR, life stress and mental disorders.

Declaration of interest

None.

In 2003 Caspi and colleagues published a landmark paper on the role of the serotonin transporter gene (SLC6A4; previously known as 5-HTT or SERT) in moderating the association between stressful life events and the risk of depression.\(^1\) This research was motivated by findings that suggested that a length polymorphism (5-HTTLPR) in the promoter region of SLC6A4 affects transcription of the gene\(^2,3\) associated with anxiety-related traits, including neuroticism and depression.\(^3\) The major conclusion of the study was that 5-HTTLPR moderated this association with individuals homozygous for the longer repeat allele (‘l’) being less sensitive to stressful life events than those of ‘l/s’ or ‘s/s’ genotype. This paper proved highly influential and was widely cited as evidence of the ways in which genes may interact with the environment to influence risks of mental disorders.\(^4,6\) with many attempts to replicate and extend the original findings reported by Caspi et al.\(^1\)

Recently, concerns have been raised about the evidence of a gene \times \text{environment} interaction between 5-HTTLPR and stressful life events in determining risks of depression and related disorders. Specifically, two major systematic reviews of the evidence have failed to find consistent and replicable gene \times \text{environment} effects.\(^5,6\) The first analysis, conducted by Munafò et al.\(^5\) reported a systematic review of 15 studies and a meta-analysis of 5 studies. The principal finding of this review was that the main effect of the 5-HTTLPR genotype and the interaction effect between 5-HTTLPR and significant life events on risk of depression were negligible. In a further paper, Risch et al.\(^6\) applied meta-analytic methods to a series of 14 studies that had examined gene \times \text{environment} effects involving 5-HTTLPR. This review found that when data were pooled across studies including the original Caspi et al\(^6\) study, there was no evidence of a gene \times \text{environment} interaction. There have been a number of responses to the meta-analyses reported by Munafò et al and Risch et al.\(^5,6\) First, it has been noted that these analyses ignore the basic scientific research involving the serotonin transporter gene.\(^4,7\) This evidence includes the role of 5-HTTLPR in drug responses and also findings from animal models and human experimental studies involving gene \times \text{environment} interactions.\(^8–11\) Although these arguments may have some foundation they fail to explain the recurrent failure of studies to replicate the 5-HTTLPR gene \times \text{environment} hypothesis. A further series of criticisms has focused on the selective nature of the studies reviewed by Munafò et al\(^5\) and Risch et al\(^6\) and it has been argued that when the totality of research evidence is taken into account the evidence in support of the 5-HTTLPR gene \times \text{environment} hypothesis is far stronger than suggested by the results of meta-analysis.\(^4,12\) A further observation has been that the failures to replicate the original findings may be the result of shortcomings in measurement. Specifically both Uher & McGuffin\(^12,13\) and Caspi et al\(^4\) have argued that studies that use checklist measures of life events may provide inadequate measures of adversity and that better measures of adversity, including in-depth interviews and the assessment of specific life events, may be preferable to checklist life events.

Inspection of the contemporary literature on this topic suggests a substantial division of conclusions based on meta-analysis and the conclusions drawn on the basis of conceptually based arguments. On the one hand the defenders of the 5-HTTLPR gene \times \text{environment} hypothesis reject the conclusions of recent meta-analyses because of the limitations they see in these methods for resolving the scientific and conceptual issues that underlie this hypothesis. On the other hand the critics of this hypothesis point to the need for stable evidence of a replicable association before the hypothesis can be accepted. Perhaps the only thing on which the divided camps agree is the need for further research to clarify issues in this area. Given the rejection of meta-analytic findings, a further approach that may clarify some of the issues raised may be to conduct or locate research that closely corresponds to the...
methodology used in the original Caspi et al. paper and use this research to directly replicate and cross-validate their findings. This would allow an independent test of the 5-HTTLPR gene × environment hypothesis that may largely overcome the problems of between-study heterogeneity that are claimed to compromise the conclusions of the meta-analyses.5,6 It turns out that such a study exists. This study is the Christchurch Health and Development Study (CHDS), which is a longitudinal study of a birth cohort of 1265 children born in the Christchurch (New Zealand) urban region who have been studied from birth to the age of 30. This study has a number of close parallels to the Dunedin Study on which the Caspi et al. analysis was based. These parallels include the fact that both studies have been conducted in the same geographic region (the South Island of New Zealand) over a similar historical period (1970–2010); both have gathered repeated-measures data on multiple sources of stress and adversity over the life course including: stressful life events, child abuse and trauma, exposure to inter-parental conflict, unemployment, violence victimisation and similar measures; and both have gathered measures of mental disorders using DSM criteria from adolescence into adulthood.

This paper summarises the findings of a series of analyses using the CHDS in which we have attempted to replicate and extend the findings of Caspi et al. on the role of 5-HTTLPR in moderating the relationships between life-course adversity and risks of depression, anxiety and similar outcomes. Given recent concerns about the role of measurement choice and quality in determining the replicability of gene × environment findings, the study uses a series of different measures of life-course adversity collected in different ways and at different times. The measures include: observational measures of parent–child interaction; prospectively collected measures of family changes in childhood; retrospective reports of child abuse and family violence; measures of adverse life events using a 30-item life-event checklist; measures of adult adversity based on in-depth interviews related to employment history, welfare dependence, poverty, exposure to inter-partner violence and unwanted pregnancy/abortion. The aims of the analysis were two-fold: to determine the extent to which the findings reported by Caspi et al. could be replicated using data from the CHDS and a wide range of measures of life-course adversity; and to examine the extent to which findings about the 5-HTTLPR gene × environment hypothesis varied with the ways in which life-course adversity was measured.

**Method**

**Sample**

The data were gathered during the course of the Christchurch Health and Development Study (CHDS). In this study a birth cohort of 1265 children born in the Christchurch (New Zealand) urban region in mid-1977 has been studied at birth, 4 months, 1 year and annually to age 16 years, and again at ages 18, 21, 25 and 30 years.14,15 Sample retention rates were high throughout the study and at age 30 the study was still able to assess over 80% of the surviving cohort. All phases of the study were subject to ethical approval from the Canterbury Regional Health and Disability Ethics Committee, and all forms of data collection were subject to the signed consent of study participants. The present analysis is based on a sample of 893 cohort members (85% White, 15% New Zealand Chinese/Pacific Island ethnicity) who were assessed on mental health outcomes in adulthood (ages 18–30 years) and who were successfully genotyped for 5-HTTLPR. This sample represented 72% of the surviving cohort. Analysis of sample attrition showed no significant relationships between measures of mental health and rates of sample loss.

**Deoxyribonucleic acid (DNA) preparation**

Between the ages of 28 and 30, participants were asked to provide a peripheral blood sample for DNA analysis: 916 agreed, with most (91.4%) providing a blood sample from which DNA was extracted using a sodium chloride precipitation procedure. For the remaining participants (8.6%), saliva was collected using Oragene collection kits (DNA Genotek, Ottawa, Canada) and DNA was extracted according to the supplier’s instructions. See the online supplement to this paper for further details on DNA preparation.

**5-HTTLPR genotyping**

Polymerase chain reaction (PCR) followed by agarose gel electrophoresis was used for genotyping 5-HTTLPR, essentially as described by Gelernter et al.,16 and as used by Caspi et al.1 Primer sequences were (5′-ATGGGAGGACCTAAACCGCTTAATG-3′) and (5′-GGACCAGAGGTGGGCGGGA-3′). The predicted amplimers are 419bp for the 16 repeat (‘l’) allele and 375bp for the 14 repeat (‘s’) allele. Cycling conditions were an initial 15 min denaturing step at 95°C, followed by 35 cycles of 94°C for 30 s, 66°C for 30 s and 72°C for 40 s, and a final extension phase of 72°C for 15 min. Reactions were performed in 10 μl volumes using 0.5 U of Fisher Biotech Taq-Ti (Wembley WA, Australia) in the supplier’s buffer containing 1.5 mM MgCl2, 500 nM each primer, 200 μM dNTPs and approximately 50 ng of genomic DNA. All reactions were set up in 384 well plates using a Janus (Perkin Elmer) automated workstation. Polymerase chain reaction products were resolved on 2% agarose TBE gels supplemented with ethidium bromide and visualised by ultraviolet transillumination. Agarose gel banding patterns were independently read and double entered into a spreadsheet by two operators masked to each other’s calls. The entries were compared, and wherever an inconsistency was observed (17/893 or approximately 2% of samples), that sample was genotyped again to resolve the discrepancy. Once all genotypes were obtained, 75 samples (approximately 8%) were drawn randomly from across the cohort for repeat PCR analysis. These genotype calls were in complete concordance with the initial analysis. The distribution of 5-HTTLPR genotype across samples was found to be in Hardy–Weinberg equilibrium (χ²(2) = 0.70, P = 0.74). There was no significant difference in genotype frequencies between genders (χ²(2) = 1.20, P = 0.55).

**Measures**

The measures used in the present analysis are described briefly below. A more detailed description of each measure is provided in the online supplement.

**Childhood adversity/maltreatment (0–16 years)**

The following measures were used to assess the extent of exposure to stress/adversity during childhood. Exposure to:

(a) childhood sexual abuse was assessed on the basis of retrospective reports of unwanted sexual experiences in childhood (<16 years) obtained at ages 18 and 21 years;17,18

(b) childhood physical abuse was assessed on the basis of retrospective reports obtained at ages 18 and 21 years of the extent to which the participant’s parent(s) was reported to have used methods of physical punishment during childhood (<16 years);15,19

(c) inter-parental violence was assessed on the basis of retrospective reports at age 18 of the extent to which the participant had witnessed incidents of inter-parental conflict and physical violence during childhood (<16 years).20
Inter-partner violence victimisation was assessed at ages 21, 25 and 30 using items from the Conflict Tactics Scale to obtain participant reports of physical assault or serious threats of physical assault by a partner in the past 12 months.

Punitive parenting behaviour during early childhood (age 3–5 years) was assessed using prospectively collected data on changes of parental management strategies. These indicators were: the individual reported being physically punished from a parent; the individual fell into the top decile on the measure of childhood inter-parental conflict; the individual experienced more than three changes of parents during childhood; the individual was observed to be exposed to three or more punitive parenting behaviours in early childhood.

Adolescent/adult stressful life events

At each assessment from age 15 to 30 participants were interviewed using a life-event checklist that examined the range and severity of adverse life events experienced for each 12 month period since the previous assessment. For each event the participant was asked to report whether they had experienced the event in each 12-month period since the previous assessment; and if they had experienced the event, to rate the extent to which the event caused them to become upset/distressed on a four-point scale. Using these data, two measures were constructed to reflect the overall extent or severity of exposure to stressful life events in the preceding 3 years: a total life-event score based on a count of the total number of life events reported over the 3-year period and a weighted life-event score based on a sum of the total life events reported over the 3-year period with each life event weighted by the severity of reported upset/distress.

Other measures of adult adversity (16–30 years)

The above life-event report data were supplemented by the following additional measures of adult life-course stress/adversity, gathered in the context of in-depth questioning of the participant on aspects of their life experience during assessments conducted from age 18 to 30.

(a) Prolonged unemployment was assessed on the basis of participant reports of having been unemployed and seeking work for a period of 12 months or longer after the previous assessment.

(b) Welfare dependency was assessed from detailed questioning on receipt of government welfare benefits at any time since the previous assessment.

(c) Poverty/depressed living standards was assessed at age 30 using the Short Form Economic Living Standards Index (ELSI) to examine the participant’s current material/economic circumstances.

(d) Inter-partner violence victimisation was assessed at ages 21, 25 and 30 using items from the Conflict Tactics Scale to obtain participant reports of physical assault or serious threats of physical assault by a partner in the past 12 months.

(e) Abortion/unwanted pregnancy (females only). As part of this study a detailed pregnancy history was obtained for each woman, including information on the woman’s history of unwanted pregnancy resulting in abortion, or unwanted pregnancy coming to term where the woman reported severe distress or other adverse reaction to the pregnancy.

Mental health outcomes

At ages 18, 21, 25 and 30, participants were administered a comprehensive mental health interview that assessed aspects of the individual’s mental health and psychosocial adjustment over the period since the previous assessment. As part of this interview, participants were assessed on DSM–IV symptom criteria for major depression and a range of anxiety disorders (generalised anxiety disorder, panic disorders, agoraphobia, social phobia, specific phobia). Questioning was based on the relevant sections of the Composite International Diagnostic Interview (CIDI). Participants were also asked about the extent and timing of any suicidal thoughts since the previous assessment. These data were used to construct four outcome measures at each age. These measures were:

(a) Symptoms of major depression: the number of DSM–IV major depression symptom criteria reported for the previous 12 months;

(b) Major depression: a dichotomous measure reflecting whether the participant met diagnostic criteria for a major depressive episode in the previous 12 months;

(c) Anxiety disorder: a dichotomous measure reflecting whether the participant met diagnostic criteria for any anxiety disorder in the past 12 months;

(d) Suicidal ideation: a dichotomous measure reflecting whether the participant reported suicidal thoughts in the past 12 months.

Statistical analysis

A common analysis framework was applied to all analyses in which a generalised estimating equation (GEE) approach was used to fit population-averaged models to the repeated-measures data for each mental health outcome. To illustrate the general approach used, a Poisson regression model was fitted to model the rate of depressive symptoms in the past 12 months as a function of 5-HTTLPR genotype, exposure to life events and their interaction (Tables 1 and 2). This model was of the form:

$$\log(Y_i) = B_0 + B_1(5-HTTLPR) + B_2(\text{stress}) + B_3(5-HTTLPR \times \text{stress})$$

where $Y_i$ was the rate of depressive symptoms for the $i$-th participant at the $t$-th time of observation (18, 21, 25, 30 years); 5-HTTLPR was the measure of genotype coded to reflect the number of short alleles ($'T''T''T'$; 0; $'T''s''s'$; 1; $'s''s''s'$; 2). Stress was the total number of life events reported for the previous 3 years scored as 0, 0–1 event; 1, 2–4 events; 2, 5–7 events; 3, 8–10 events; 4, 11+ events; and 5-HTTLPR × stress was an interaction term given by the product of the number of short alleles × number of life events. In this model the coefficient $B_1$ represents the main effect of genotype; $B_2$, the main effect of life stress; and $B_3$, the change in the effect of stress attributable to having a given number of short alleles. The intercept term $B_0$ was permitted to vary with time of observation to allow for age-related changes in the rate of depressive symptoms. With the model formulated in this way a positive $B_3$ coefficient would be consistent with the Caspi et al. hypothesis of greater responsivity to stress with an increasing number of $'s'$ alleles. The test of significance of the interaction effect was based on the standard $Z$-test given by the ratio of the regression parameter $B_3$ to its standard error.
The above approach was extended to fit a series of models for each outcome (depressive symptoms, major depression, anxiety disorder, suicidal ideation) varying the measure of stress over the array of measures of childhood adversity and adolescent/adult life stressors described above. For depressive symptoms Poisson regression models were fitted, whereas for dichotomous outcomes logistic regression models were used. For each model the value of the Z-test of significance of the interaction parameter $B_3$ was plotted for each outcome across the multiple measures of stress and adversity (Fig. 1). The advantage of using the Z-values in this way is that it provides a measure of effect size that is consistent across multiple different outcomes analysed in different ways, and across multiple different measures of stress/adversity. In addition the sign of the Z-test indicates the direction of the gene × environment interaction effect.

For Poisson and logistic regression the effect of each additional ‘s’ allele is to modify the slope of the association between number of life events and the rate of the outcome in a multiplicative fashion. To examine the robustness of the findings a second series of GEE models was fitted to test a simple ‘additive’ interaction assumption. This was achieved by directly modelling either the mean number of depressive symptoms or the probability of the assumption. This was achieved by directly modelling either the mean number of depressive symptoms or the probability of the disorder as a linear function of genotype, stress and their interaction. All GEE models were fitted assuming an unstructured disorder as a linear function of genotype, stress and their interaction. All GEE models were fitted assuming an unstructured correlation matrix of the repeated measures of each outcome within individuals over time, and using robust estimates of standard errors to allow for possible overdispersion in the multiplicative models and failure to conform to standard normality assumptions for the additive models.

### Results

**5-HTTLPR, life events and depressive symptoms**

Table 1 shows the mean rate of depressive symptoms in the previous 12 months at ages 18, 21, 25 and 30 years, cross-classified by the total number of adverse life events reported over the preceding 3 years. Table 1 is further stratified by the 5-HTTLPR genotype into ‘l/l’; ‘l/s’ and ‘s/s’ groups. For each stratum and each level of the life-event measure, Table 1 also provides population-averaged estimates of mean depressive symptoms pooled over the period from 18 to 30 years. Inspection of Table 1 shows a clear tendency for rates of depressive symptoms to increase with increasing life-event exposure but no clear evidence of a gene × environment interaction in which the slope of the relationship between life events and depressive symptoms varies with genotype. These conclusions were confirmed by fitting a Poisson regression model to the repeated-measures data in Table 1 in which the rate of depressive symptoms was modelled as a function of the main effects of number of adverse life events and number of 5-HTTLPR short alleles and the multiplicative interaction of life events by number of short alleles (see Method). Table 2 summarises this analysis and shows that there was:

(a) evidence of a highly significant linear association between the number of life events reported and rates of depressive symptoms ($P<0.0001$);

(b) no evidence of a main effect of 5-HTTLPR on rates of depressive symptoms ($P=0.75$);

(c) no evidence of a significant gene × environment interaction involving number of 5-HTTLPR short alleles and adverse life events ($P=0.84$).

### Extension to multiple measures of stress and disorder

The analysis reported in Tables 1 and 2 was extended to consider multiple measures of stress/adversity and multiple outcome measures. Measures of stress included: childhood events (exposure to childhood sexual abuse, childhood physical abuse, inter-parental violence, changes of parents, punitive parenting); measures of adolescent/adult life events constructed using both an unweighted life-event count and scores weighted by the severity of distress caused; and other measures of adult adversity derived from in-depth questioning about specific sources of stress (unemployment, welfare dependence, poverty/depressed living standards, inter-partner violence victimisation, abortion/unwanted pregnancy). The outcomes considered included: depressive symptoms (as in Table 1); major depression; anxiety disorder; and suicidal ideation. All outcomes were assessed for the previous 12 months at ages 18, 21, 25 and 30 years. In total 13 stress measures and 4 outcomes were considered, giving a total of 52 analyses testing for multiplicative gene × environment interaction.

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Mean depressive symptoms in past 12 months (at ages 18, 21, 25, 30 and pooled 18–30) by number of life events (past 3 years) and 5-HTTLPR genotype a</th>
<th>Life events, n</th>
<th>0–1</th>
<th>2–4</th>
<th>5–7</th>
<th>8–10</th>
<th>11+</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Long/long alleles</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 years</td>
<td>0.2 (42)</td>
<td>1.3 (98)</td>
<td>2.3 (73)</td>
<td>2.5 (35)</td>
<td>3.7 (25)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21 years</td>
<td>0.9 (32)</td>
<td>1.2 (89)</td>
<td>1.6 (69)</td>
<td>2.5 (42)</td>
<td>3.6 (45)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 years</td>
<td>0.3 (44)</td>
<td>1.0 (119)</td>
<td>1.7 (75)</td>
<td>2.2 (25)</td>
<td>2.9 (16)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 years</td>
<td>0.9 (74)</td>
<td>1.5 (136)</td>
<td>2.0 (55)</td>
<td>2.1 (8)</td>
<td>6.5 (4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pooled age: 18–30 years</td>
<td>0.6 (194)</td>
<td>1.3 (442)</td>
<td>1.9 (272)</td>
<td>2.4 (110)</td>
<td>3.6 (90)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Long/short alleles</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 years</td>
<td>0.2 (69)</td>
<td>1.3 (141)</td>
<td>2.0 (102)</td>
<td>3.9 (62)</td>
<td>2.5 (43)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21 years</td>
<td>0.4 (59)</td>
<td>1.1 (133)</td>
<td>1.8 (88)</td>
<td>2.6 (64)</td>
<td>3.7 (69)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 years</td>
<td>0.8 (70)</td>
<td>0.8 (176)</td>
<td>1.8 (106)</td>
<td>3.2 (39)</td>
<td>5.1 (35)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 years</td>
<td>1.2 (103)</td>
<td>1.2 (198)</td>
<td>1.5 (68)</td>
<td>2.1 (22)</td>
<td>7.6 (14)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pooled age: 18–30 years</td>
<td>0.7 (318)</td>
<td>1.1 (648)</td>
<td>1.8 (364)</td>
<td>3.1 (187)</td>
<td>4.0 (161)</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Short/short alleles</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18 years</td>
<td>1.2 (17)</td>
<td>1.4 (73)</td>
<td>2.2 (52)</td>
<td>0.9 (22)</td>
<td>3.1 (16)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>21 years</td>
<td>0.7 (18)</td>
<td>1.8 (62)</td>
<td>2.2 (46)</td>
<td>2.4 (25)</td>
<td>3.0 (27)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25 years</td>
<td>0.5 (25)</td>
<td>1.4 (88)</td>
<td>2.0 (46)</td>
<td>2.9 (11)</td>
<td>3.4 (9)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>30 years</td>
<td>0.7 (40)</td>
<td>1.0 (87)</td>
<td>2.0 (39)</td>
<td>2.4 (10)</td>
<td>2.3 (3)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pooled age: 18–30 years</td>
<td>0.8 (100)</td>
<td>1.4 (310)</td>
<td>2.1 (183)</td>
<td>2.0 (68)</td>
<td>3.1 (55)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

a. Numbers in parentheses represent the number of observations for each cell of the table.
interaction effects between number of 5-HTTLPR short alleles, exposure to stress and outcomes (see Method). The results of these tests are summarised in Fig. 1, which gives plots of the values of the Z-test of significance for each interaction. The findings lead to the following conclusions.

(a) Depressive symptoms: of the 13 tests of interaction only one, involving childhood sexual abuse, was significant ($P < 0.05$). The sign of this test was negative, indicating that an increasing number of ‘s’ alleles was associated with reduced responsivity to stress. Overall, only 4 of the 13 tests were positive in sign and hence consistent with the original Caspi et al hypothesis, and none of these was significant.

(b) Major depression: of the 13 tests only one, involving childhood sexual abuse, was significant ($P < 0.05$). The sign of this test was negative. Overall, 5 of the 13 tests were positive in sign.

(c) Anxiety disorder: of the 13 tests, 3 (childhood sexual abuse, childhood adversity score and prolonged unemployment) reached statistical significance ($P < 0.05$), with all of these tests suggesting that an increasing number of ‘s’ alleles was associated with reduced responsivity to stress. Eleven of the 13 results were negative in sign, suggesting that there was a general tendency for increasing numbers of ‘s’ alleles to be associated with a reduction of the effects of stressors on anxiety disorders.

(d) Suicidal ideation: none of the 13 tests of interaction was statistically significant. Only 4 of the 13 results suggested an effect in a positive direction.

The overall impression conveyed by the results in Fig. 1 was that for most measures of stress there was no evidence to suggest that 5-HTTLPR moderated their effects on depression, anxiety or suicidality. Only 5 of the 52 results were statistically significant. It is noteworthy that all significant tests of gene × environment interactions suggested that increasing numbers of ‘s’ alleles led to reduced sensitivity to stressful events. Given the large number of comparisons made, these results could well be a result of chance. One way of addressing this is to apply a Bonferroni corrected significance value to the results in Fig. 1. This test showed that none of the comparisons in Table 1 reached the Bonferroni corrected value of $P = 0.001$.

The analyses summarised in Fig. 1 were replicated using linear regression models to test for additive interaction effects (see Method). The findings from these analyses were virtually identical to the findings of the multiplicative models. The overall findings from the 104 models (52 multiplicative, 52 additive models) fitted to the data are summarised in Fig. 2, which shows the distribution of Z-test values for gene × environment interaction for each series of analyses. The figure shows that for both series the majority (two-thirds) of Z-values were negative, indicating a general tendency for reduced responsivity to stress/adversity with increasing number of ‘s’ alleles. This trend is reflected in the overall mean Z-value for each distribution ($−0.58$ for multiplicative models, $−0.57$ for additive models). It also shows that for both series of analyses the only significant tests of gene × environment interaction were negative in sign (five tests for the multiplicative models, six for the additive models).

Further analyses
The analyses reported in Fig. 2 were extended in a number of ways to examine the sensitivity of these findings to further approaches to analysis. These extensions involved examination of: different lags (1, 2, 3, 4 years) for the assessment of adult life-event measures; alternative ways of classifying 5-HTTLPR genotype (‘l’/‘l’, ‘l’/‘s’, ‘s’/‘s’ and ‘l’/‘l’, ‘l’/‘s’, ‘s’/‘s’); the implications of

---

**Table 2** Tests of gene, life event and gene × life event interaction effects from Poisson regression model fitted to the data in Table 1

<table>
<thead>
<tr>
<th>Effect</th>
<th>B (s.e.)</th>
<th>$P$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main effects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of 5-HTTLPR short alleles</td>
<td>0.031 (0.096)</td>
<td>0.75</td>
</tr>
<tr>
<td>Number of life events</td>
<td>0.340 (0.039)</td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Interaction</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of short alleles × number of life events</td>
<td>−0.007 (0.035)</td>
<td>0.84</td>
</tr>
</tbody>
</table>

---

**Fig. 1** Z-test values for tests of significance of number of 5-HTTLPR short alleles gene × life events interaction from fitted multiplicative models for varying mental health outcomes and varying measures of adverse life events.
A number of recent articles have argued that the recurrent failure to replicate the 5-HTTLPR gene × environment hypothesis is a result of shortcomings in the ways in which environmental adversity has been assessed, with studies using life-events checklists being associated with increased rates of stress. For this reason, the present analysis was stratified by gender to examine gene × environment interactions separately for males and females. None of these elaborations altered any of the conclusions drawn above.

Main findings
In this analysis we have used data gathered over the course of a 30-year longitudinal study to conduct a replication and extension of the findings of Caspi et al using a study that has strong similarities to the Dunedin Multidisciplinary Study. Despite extensive analyses that involved 13 measures of childhood and adult adversity and 4 outcome measures, modelled using both multiplicative and additive models, we have been unable to find any evidence that supports the hypothesis that 5-HTTLPR moderates the relationships between exposure to adversity and mental disorder. Although the analysis found a number of significant gene × environment interactions these were no greater than would be expected by chance. Furthermore in all cases where significant gene × environment effects were found these were opposite in direction to that hypothesised, with increasing numbers of ‘s’ alleles being associated with reduced responses to stress.

A number of recent articles have argued that the recurrent failure of studies to replicate the 5-HTTLPR gene × environment hypothesis is a result of shortcomings in the ways in which environmental adversity has been assessed, with studies using life-events checklists being associated with increased rates of failure to replicate.4,12,13 For this reason, the present analysis focused on examining the extent to which variations in the ways in which adversity is measured are associated with different findings. Specifically the present analysis used five different approaches to measuring adversity. These were: direct observation of parent–child interaction; prospectively assessed measures of family change during childhood; retrospectively reported child abuse and inter-parental violence; life-event checklist measures; and interview-based assessments of unemployment, welfare dependence, poverty/depressed living standards, inter-partner violence and unwanted pregnancy/abortion. Despite the array of measures used in the analysis we were unable to find any statistically significant gene × environment interactions in the expected direction. These findings suggest that it is unlikely that between-study failures to replicate findings for 5-HTTLPR can be explained away simply as being because of shortcomings in measurement approaches.

The accumulated evidence on the 5-HTTLPR hypothesis has led to a highly divided and controversial literature that has produced enigmatic findings. Consideration of the history of research in this area suggests that three factors may have combined to produce this situation.

Factors involved
Overinterpretation of the evidence
As a number of commentators have noted6,27 there has been an unfortunate tendency for the original Caspi et al findings to be overinterpreted as an exemplar of the ways in which genes and environment may interact in the development of complex mental disorders. This led to what has proved to be a highly unstable interaction being used to draw strong conclusions about the interplay of nature and nurture. Irrespective of debates about the existence of a gene × environment interaction involving 5-HTTLPR it is now clear that the evidence in this area is currently far too weak and inconsistent to be used to draw strong inferences about the role of nature and nurture in mental disorder.

Weak interaction effects
Inspection of the original findings reported by Caspi et al1 shows that the purported gene × environment interactions involved relatively minor differences in the slopes of the relationships between measures of adversity and rates of mental disorder. These differences were most evident at the extremes of the distribution. For this reason it could be confidently predicted that the original findings for the 5-HTTLPR gene × environment interaction would prove to be unstable and difficult to replicate, as has proved to be the case.

Complexity of measures of adversity
As the present paper clearly demonstrates, variation in exposure to life-course adversity can be assessed in a large number of ways. Furthermore, for any given measurement there are often a number of choices about the ways in which variables are scaled and represented. This gives researchers seeking gene × environment interactions the latitude to explore a range of measures assessed in different ways. In turn, this leads to the possibility of increased
risks of type I statistical errors as a result of selective reporting of promising findings. One way to address these issues is for studies in this area to report a range of results in order to examine the robustness of findings across a range of measures (as has been attempted in this paper).

Implications

Consideration of these problems highlights the difficulties that arise in locating stable gene × environment interactions. Most of the problems can be traced back to the small size of interactions involving a single gene and a specific measure of environmental adversity. One way of moving beyond these problems is for greater investment in studies that examine the ways in which multiple genetic factors interact with multiple sources of environmental adversity. This approach may overcome some of the problems of replication that have beset studies of gene × environment interaction to date.

References

Online supplement

DNA preparation

Genomic deoxyribonucleic acid (DNA) was obtained from consenting participants between the ages of 28 and 30. In most cases (91.4%), DNA was extracted from peripheral blood using a salting-out extraction procedure. To control for laboratory processing errors, three 5 ml EDTA blood tubes were drawn from the majority of participants and two tubes were independently processed for DNA extraction on separate days, with the third unopened tube retained frozen as a back-up. The resulting duplicate DNA preparations were stored in two parallel collections. Sample labelling and handling was rigorously controlled and audited throughout all laboratory procedures. Once the entire cohort was recruited, gender-typing polymerase chain reaction (PCR) analysis was carried out on both collections, using an in-house XY multiplex PCR (protocol available from the author on request) and the results compared. Of the 793 duplicate samples extracted from blood, there was one gender disagreement (0.13% error rate). This mismatch was resolved by disposal of both samples, and extraction of the third stored blood sample from that participant. For those participants not wishing or not able to undergo phlebotomy (8.6% of participants), saliva was collected using Oragene™ collection kits (DNA Genotek, Ottawa, Canada) and DNA was extracted according to the supplier’s instructions. Duplicate samples were not obtained for these individuals, although gender-typing PCR was carried out as described above. Finally, an additional quality control comes from some 40 internal duplicates, that is, DNA samples that resulted from independent collection on different occasions (such as blood samples and Oragene™ saliva samples provided at different times). Genotypes from these samples were cross-checked to ensure they yielded identical results (100% concurrence).

Description of measures

Childhood adversity/maltreatment (0–16 years)

The following measures were used to assess the extent of exposure to stress/adversity during childhood.

Childhood sexual abuse

Exposure to childhood sexual abuse was assessed on the basis of retrospective reports obtained at ages 18 and 21 years. Participants were questioned about their experience of a range of 15 abusive experiences prior to age 16, and for each incident reported further detail was gathered on the nature and context of the abuse. On the basis of this questioning participants were classified into four groups reflecting the most severe form of abuse reported at either age: no childhood sexual abuse; non-contact childhood sexual abuse (for example indecent exposure, lewd or threatening sexual comments); contact childhood sexual abuse involving inappropriate touching of genital areas; attempted/completed sexual penetration.

Childhood physical abuse

Exposure to childhood physical abuse was assessed on the basis of retrospective reports obtained at ages 18 and 21 of the extent to which the participant’s parent(s) were reported to have used methods of physical punishment during childhood (<16 years). Separate reports were obtained for use of physical punishment by each parent. For the purposes of the present analysis participants were classified into three groups reflecting the most severe form of physical punishment by either parent reported at either age. These groups were: parents never or rarely used physical punishment; at least one parent regularly used physical punishment; at least one parent used frequent, severe or harsh physical punishment.

Inter-parental violence

This was assessed at age 18 using selected items from the Conflict Tactics Scale to assess the extent to which the participant had witnessed incidents of inter-parental conflict and physical violence during childhood. Separate reports were obtained for father-initiated and mother-initiated violence. These items were combined to form a scale measure reflecting the extent of inter-parental violence. For analysis purposes participants were classified into three groups ranging from those reporting no inter-parental violence (50% of the sample) to those whose scores placed them in the top 10% of the distribution of inter-parental violence.

Changes of parents

As part of the annual assessments from age 1 to 16 years information was obtained on changes of parents since the previous assessment. An overall measure of family stability during childhood was developed based on a count of the number of changes of parents experienced by the child from birth to age 16 years. This count included all changes due to parental separation/divorce, reconciliation, remarriage/cohabitation, parental death, fostering and other changes of custodial parents. For analysis purposes the sample was classified into four groups (0, 1–2, 3–4, 5+ changes) based on the distribution of number of changes of parents by age 16.

Parental punitive behaviour

As part of the assessments when the study children were aged 3, 4 and 5 years, interviewer observations of the quality of mother–child interaction were obtained using the Avoidance of Restriction and Punishment subscale of the HOME Inventory. As part of the assessments when the study children were aged 3, 4 and 5 years, interviewer observations of the quality of mother–child interaction were obtained using the Avoidance of Restriction and Punishment subscale of the HOME Inventory. This scale comprised a total of eight items reflecting such behaviours as: shouting at the child, expressing annoyance and hostility to the child, spanking or slapping the child, scolding or putting down (demeaning) the child. Scale scores were calculated for each year using an unweighted sum of the items. Scores were then averaged over the three assessments to provide an overall measure of the extent of punitive parenting behaviours in early childhood.

Childhood adversity score

The above measures were used to construct an overall index of the extent of stress/adversity experienced by the child to age 16 years based on a sum of five dichotomous indicators of adverse childhood experiences. These indicators were whether the individual:

(a) reported childhood sexual abuse involving genital contact or attempted/completed intercourse;

(b) reported experiencing frequent, severe or harsh physical punishment from a parent;

(c) fell into the top decile on the measure of childhood inter-parental conflict;

(d) had experienced more than three changes of parents during childhood;

(e) was observed to be exposed to three or more punitive parenting behaviours in early childhood.
Overall, 69% of the sample was classified as experiencing no adversity; 20% as having one adverse experience; and 11% as experiencing two or more adversities.

**Adolescent/adult stressful life events**

As part of the participant interview at each assessment from age 15 onwards participants were questioned using a life-event checklist that examined the range and severity of adverse life events experienced for each 12-month period since the previous assessment. At ages 15 and 16 this checklist was based on the Feeling Bad Scale. From age 18 onwards this scale was updated to include elements from adult life-events checklists, including the Social Readjustment Rating Scale and the life-events scale reported by Henderson et al. The full scale comprised a list of 30 adverse life events spanning such domains as: employment problems; serious financial difficulties; partner relationship problems; serious relationship problems with friends or other family members; serious illness or accident of the participant, other close friends or family members; sexual or physical assault victimisation; pregnancy, pregnancy loss and parenthood. For each event the participant was asked to report whether they had experienced the event in each 12-month period since the previous assessment; and if they had experienced the event, to rate the extent to which the event caused them to become upset/distressed on a four-point scale ranging from not upset/distressed to very upset/distressed. Using these data, two measures were constructed to reflect the overall extent or severity of exposure to stressful life events in the preceding 3 years at ages 18, 21, 25 and 30 years. The first measure was the total life-event score and this measure was based on an unweighted sum of the total number of life events that the individual reported over the 3-year period. For analysis purposes the life-event score at each age was classified into a series of class intervals as follows: 0, 0–1 event; 1, 2–4 events; 3, 5–7 events; 4, 8–10 events; 5, 11+ events. The second measure was the weighted life-event score. A disadvantage of the previous measure was that it did not take into account the extent of distress that each event caused for the individual. To address this issue a weighted life-event score was constructed by summing the total life events reported, with each event weighted by the level of distress that it caused, to provide a measure of the overall severity of life stress. For analysis purposes the weighted life-event score at each age was classified into four class intervals ranging from scores that placed individuals in the lowest 35–40% of the distribution of overall life stress to scores that placed them in the highest 10% of the sample.

**Other measures of adult adversity (16–30 years)**

The above life-event report data were supplemented by the following additional measures of life-course stress/adversity over the period from age 16 to 30 years. In each case the measure was obtained as part of in-depth questioning of the participant on aspects of their life experience.

**Prolonged unemployment**

At each assessment from age 18 to 30 years participants were questioned in detail about their employment history since the previous assessment. As part of this questioning, information was obtained on the extent and duration of any periods of unemployment. For the purposes of this analysis participants were classified as having experienced prolonged unemployment if they reported having been unemployed and seeking work for a period of 12 months or longer since the previous assessment.

**Welfare dependence**

At each assessment participants were also questioned about sources of income since the previous assessment. As part of this questioning, information was obtained on receipt of government welfare support, including unemployment, sickness, invalid’s, domestic purposes and related benefits. Participants were classified as welfare dependent if they reported receiving welfare payments at any time since the previous assessment.

**Poverty/depressed living standards (age 30 only)**

At age 30, participants were questioned using the short form Economic Living Standards Index (ELSI) to assess the extent of current material/economic deprivation. The ELSI scale combines a mix of items reflecting social and material deprivation, economic hardship with self-report ratings of living standards and income adequacy to derive an overall scale score reflecting the overall quality of material living standards experienced by the individual. For the purposes of the present analysis participants were classified into four groups on the basis of their ELSI scores, ranging from those whose scores placed them in the 50% of the sample with the highest living standards (group 1) to those whose scores placed them in the 10% of the sample with the lowest living standards (group 4).

**Inter-partner violence victimisation**

At ages 21, 25 and 30, participants were questioned concerning their partner relationship experiences over the past 12 months, and in particular the extent and nature of any incidents of inter-partner conflict and violence. Questioning was based on the Conflict Tactics Scale. For the purposes of the present analysis participants were classified as being a victim of inter-partner violence if they reported any incident of physical assault or serious threats of physical assault by a partner in the past 12 months.

**Abortion/unwanted pregnancy (females only)**

At each assessment from age 16 to 30 years female participants were questioned in detail concerning any pregnancies occurring since the previous assessment and the outcome of each pregnancy. At age 30, women were further questioned about their lifetime pregnancy history, with additional information being obtained on planning of pregnancy, the women's reaction to each pregnancy and whether the pregnancy was wanted. Using these data it was possible to construct a detailed history for each woman on the timing, planning, wantedness and outcome of each pregnancy. For the purposes of the present analysis a dichotomous measure was constructed at each age based on the woman's history of unwanted pregnancy resulting in abortion, or unwanted pregnancy coming to term where the woman reported severe distress or other adverse reaction to the pregnancy.

**Mental health outcomes**

At ages 18, 21, 25 and 30, participants were administered a comprehensive mental health interview that assessed aspects of the individual's mental health and psychosocial adjustment over the period since the previous assessment. As part of this interview, participants were assessed on DSM–IV symptom criteria for major depression and a range of anxiety disorders (generalised anxiety disorder, panic disorders, agoraphobia, social phobia, specific phobia). Questioning was based on the relevant sections of the Composite International Diagnostic Interview (CIDI). Participants were also asked about the extent and timing of any...
suicidal thoughts since the previous assessment. These data were used to construct four outcome measures at each age. These measures were:

(a) symptoms of major depression: the number of DSM–IV major depression symptom criteria reported for the previous 12 months;

(b) major depression: a dichotomous measure reflecting whether the participant met diagnostic criteria for a major depressive episode in the previous 12 months;

(c) anxiety disorder: a dichotomous measure reflecting whether the participant met diagnostic criteria for any anxiety disorder in the past 12 months;

(d) suicidal ideation: a dichotomous measure reflecting whether the participant reported suicidal thoughts in the past 12 months.

In addition, at ages 18, 21 and 25, interviews were conducted with a significant other nominated by the participant. As part of this interview the significant other was questioned about the participant’s mental health, including symptoms of major depression occurring in the past 12 months. These data were used to construct a separate informant report of number of depressive symptoms at each age.

Additional references


Life stress, 5-HTTLPR and mental disorder: findings from a 30-year longitudinal study
David M. Fergusson, L. John Horwood, Allison L. Miller and Martin A. Kennedy
Access the most recent version at DOI: 10.1192/bjp.bp.110.085993