Cortisol awakening response and subsequent depression: prospective longitudinal study

Rebecca Carnegie, Ricardo Araya, Yoav Ben-Shlomo, Vivette Glover, Thomas G. O’Connor, Kieran J. O’Donnell, Rebecca Pearson and Glyn Lewis

Background
Some studies have found an association between elevated cortisol and subsequent depression, but findings are inconsistent. The cortisol awakening response may be a more stable measure of hypothalamic–pituitary–adrenal function and potentially of stress reactivity.

Aims
To investigate whether salivary cortisol, particularly the cortisol awakening response, is associated with subsequent depression in a large population cohort.

Method
Young people (aged 15 years, n = 841) from the Avon Longitudinal Study of Parents and Children (ALSPAC) collected salivary cortisol at four time points for 3 school days. Logistic regression was used to calculate odds ratios for developing depression meeting ICD-10 criteria at 18 years.

Results
We found no evidence for an association between salivary cortisol and subsequent depression. Odds ratios for the cortisol awakening response were 1.24 per standard deviation (95% CI 0.93–1.66, \(P = 0.14\)) before and 1.12 (95% CI 0.73–1.72, \(P = 0.61\)) after adjustment for confounding factors. There was no evidence that the other cortisol measures, including cortisol at each time point, diurnal drop and area under the curve, were associated with subsequent depression.

Conclusions
Our findings do not support the hypothesis that elevated salivary cortisol increases the short-term risk of subsequent depressive illness. The results suggest that if an association does exist, it is small and unlikely to be of clinical significance.

Declaration of interest
None.
cohort. Continuing follow-up of these participants found decreasing evidence for an association between CAR and subsequent depression over time.26 It is possible that the abnormalities may be early biological changes of an emerging or subclinical depressive process, rather than pre-existing markers observable in healthy individuals. Further studies using longer follow-up periods would help to clarify the association between CAR and subsequent depression. In our study we investigated whether CAR was associated with subsequent depression in a longitudinal study of a population cohort of British adolescents. We also examined other salivary cortisol measures used in previous HPA research. To our knowledge, this is the largest longitudinal study using a comprehensive cortisol protocol to date. Using an established population birth cohort has the advantage of being able to account for confounding factors without recall bias, as well as maximising on the proven collaboration of participants, which have both been suggested as limitations in previous cortisol research.19,27

Method

Participants

The Avon Longitudinal Study of Parents and Children (ALSPAC) is an ongoing prospective longitudinal study of children born in the former county of Avon, UK.28,29 Pregnant women with expected delivery dates between 1 April 1991 and 31 December 1992 were recruited, with catch-up recruitment at age 7. The complete ALSPAC sample consists of 14701 children alive at 1 year. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the local research ethics committees.

Participants were recruited from attendees of a research clinic at 15 years, which assessed 5501 of the ALSPAC participants (online Fig. DS1). Around 3020 participants were invited at random to participate in the cortisol data collection, 1175 declined. Of the 1845 participants agreeing to enter the study, 1035 returned their saliva sampling packs. However, 75 participants provided empty cortisol samples and 119 provided insufficient data for analysis of CAR. Of the remaining 841 participants, 173 participants had no outcome data for depression at 18, either because of non-attendance at the follow-up research clinic (n = 122) or non-completion of the Revised Clinical Interview Schedule (CIS-R)26 diagnostic tool at the follow-up clinic (n = 51).

This left 668 participants (22.1% of invited) with both exposure and outcome measure at 18, on which the primary analysis was undertaken. For 208 participants there were one or more covariates missing from the ALSPAC data-set, leaving 460 complete cases. Sociodemographic characteristics of active participants were compared with the invited participants, using chi-squared tests to evaluate any selective attrition. Non-participants were defined as those randomly invited to participate at the baseline clinic but either declining to participate or failing to provide sufficient data.

Cortisol sampling

As described in detail elsewhere,31 each participant agreeing to take part in the study was given a cortisol sampling pack during their clinic visit. Each pack contained detailed sampling instructions, 12 Salivette collection devices (Sarstedt, Germany), a sample collection diary sheet and a pre-paid envelope in which to return the samples to the laboratory. Participants were instructed to provide four samples per day over 3 consecutive school days, and to record the time of the sample. The samples were collected immediately on waking, at 30 min after waking, after school and before bed. To avoid contamination, participants were instructed not to eat or brush their teeth for at least 30 min prior to collection. On completion, samples were returned to the laboratory and stored at —20°C. Salivary cortisol was determined using a commercially available enzyme immune assay. Inter-assay and intra-assay variation were 7.9 and 8.9 respectively.

- Cortisol samples were excluded if they were greater than 82 nmol/L (the upper limit of the cortisol assay) or greater than 4 standard deviations above the mean. Individual samples were also excluded if the timing deviated from the specified protocol. For the CAR, samples were excluded if taking the initial sample after 10.00 h or taking the second sample less than 20 or more than 60 min after the first. For the mean levels at each time point, the second sample was excluded if taken after 11.00 h, the ‘after school’ samples were excluded if before 15.00 h or after 18.00 h, and the final ‘before bed sample’ was excluded if before 19.00 h. Although previous literature suggested an absent CAR in a proportion of the population, a secondary analysis excluded individuals with an absent or negative CAR, in case this was the result of measurement error or non-adherence.

- The primary cortisol measure of interest was the CAR, derived using the difference between the two morning cortisol samples. Additionally, the mean cortisol level at each time point, the estimated daily secretion as given by the area under the curve and the diurnal drop in cortisol, and estimated total morning cortisol were analysed. The diurnal drop in cortisol was taken as the difference between mean morning and bedtime cortisol levels. Total morning cortisol was calculated using the sum of the two morning samples. Each cortisol variable was calculated as a mean of the available samples over the study period.

Confounding factors

Potential confounders were decided a priori based on the available literature. Confounders available from assessment at 15 years included current depressive symptoms (using bands of the Development and Well-Being Assessment, DAWBA)32 body fat percentage, oral contraceptive use, smoking and alcohol use. Historical assessments also provided potentially confounding background characteristics such as gender, social class, birth weight, gestation, prior depressive symptoms (using continuous scores of the Short Mood and Feelings Questionnaire (SMFQ))33 assessed at 10, 13 and 14 years), maternal education, maternal history of depression and exposure to postnatal depression (assessed using the Edinburgh Postnatal Depression Scale, EPDS).34

Following the development of a full set of covariates, certain variables were dropped from further consideration as they were neither associated with the cortisol measures nor had an impact on the effect size. This also helped to minimise numbers lost through incomplete data. This subset included anxiety symptoms at 15 years, neuroticism, cannabis use, friendships and social support, early significant life events, hours of sleep and income. Additionally, to avoid overadjustment and preserve sample size, groups of similar potential confounders were reduced to a single representative variable, with the others dropped from the final model. This subset included baseline fitness level and exercise, which was replaced by adiposity alone and antenatal depressive and anxiety symptoms, which was replaced by postnatal depressive symptoms alone.

Finally, because of the composition of the sample, other variables with negligible prevalence were dropped as covariates. First, as none of the individuals in the sample population were pre-pubertal (taken as a Tanner stage ≤2), pubertal stage was dropped as a confounder. Similarly, as only three participants reported steroid use prior to the study period, none of whom developed depression on follow-up, this was also dropped from the final model.
Outcome measures

The main outcome measure was ICD-10 diagnosis of depression, which was assessed using the CIS-R at a research clinic attended by participants at a mean age of 17.7 years. Responses to the questions about 14 symptom groups, as well as their onset and duration, were given via computer self-assessment at a clinic visit. The five symptom groups used to assess depression were sleep problems, poor concentration, fatigue and depression (which were scored on a scale of 0–4) and depressive thoughts (scored from 0 to 5). Responses to questions were aggregated by the specified algorithm to make a binary variable of ICD-10 diagnosis of depression.

A secondary analysis was undertaken using a binary cut off of ‘depressive symptoms’ to account for subclinical symptoms of depression. A ‘depressive symptoms’ score was derived from the sum of the five depressive symptom groups in the CIS-R. This score ranged from 0 to 21. A score of ≥8 depressive symptoms was chosen, aiming for a prevalence of 15%. Secondary analysis also looked at depressive symptoms after 1 year using a cut-off of 5 on the SMFQ, assessed by postal questionnaire at a mean age of 16.7 years.

Statistical analysis

Statistical analysis was performed using Stata 12 on Windows. Logistic regression was used to calculate odds ratios between each cortisol variable and the subsequent development of depression. As CAR measured on a single day has been suggested to be related to current circumstances rather than a stable interpersonal state marker, the number of days used to calculate a mean CAR was included as a covariate. The waking cortisol level was also included as a covariate as this has been shown to affect the size of the CAR.

The possibility of a non-linear relationship between cortisol and depression was investigated using a quadratic term. We postulated a priori that there may be an interaction between gender and CAR and tested this using an appropriate interaction term in the regression model. An interaction term in the regression model was also used to investigate whether the association between cortisol and subsequent depression differed for participants with previous depression, baseline depressive symptoms or those with a positive family history of depression.

Table 1 Differences in characteristics of participants with exposure and outcome data compared with those invited to participate from baseline assessment clinic

<table>
<thead>
<tr>
<th></th>
<th>Non-participants (n = 2363)</th>
<th>Participants with exposure and outcome (n = 668)</th>
<th>χ²</th>
<th>t-test</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1159 (49.1)</td>
<td>301 (45.1)</td>
<td>3.3</td>
<td>0.07</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>1204 (51.0)</td>
<td>367 (54.9)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Maternal education, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;O-level</td>
<td>426 (19.4)</td>
<td>99 (15.5)</td>
<td>7.2</td>
<td>0.03</td>
<td></td>
</tr>
<tr>
<td>O-level</td>
<td>747 (34.0)</td>
<td>247 (38.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A-level or higher</td>
<td>1022 (46.6)</td>
<td>293 (45.8)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>168</td>
<td>29</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Social class,a n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>75 (3.6)</td>
<td>30 (4.8)</td>
<td>4.7</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>563 (26.7)</td>
<td>177 (28.4)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>1133 (53.8)</td>
<td>327 (52.5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>281 (13.3)</td>
<td>70 (11.2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>54 (2.6)</td>
<td>19 (3.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>257</td>
<td>45</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ethnicity of child, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>2075 (96.1)</td>
<td>605 (96.2)</td>
<td>0.0</td>
<td>0.93</td>
<td></td>
</tr>
<tr>
<td>Black and minority ethnic</td>
<td>84 (3.9)</td>
<td>24 (3.8)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>204</td>
<td>39</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression at 15 years, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>222 (9.5)</td>
<td>42 (6.3)</td>
<td>6.6</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>2118 (90.5)</td>
<td>625 (93.7)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>23</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Previous depression, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>345 (14.8)</td>
<td>82 (12.4)</td>
<td>2.5</td>
<td>0.12</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1980 (85.2)</td>
<td>579 (87.6)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>38</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Depression at 18 years, n (%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td>108 (7.3)</td>
<td>46 (6.9)</td>
<td>0.1</td>
<td>0.74</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>1376 (92.7)</td>
<td>622 (93.1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>879</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Body fat, %: mean (s.d.)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22.2 (10.3)</td>
<td></td>
<td>23.2 (10.3)</td>
<td>−2.2</td>
<td>0.03</td>
<td></td>
</tr>
</tbody>
</table>

a. I, professional occupations; II, managerial and technical; III, skilled occupations (includes manual and non-manual); IV, partly-skilled occupations; V, unskilled occupations.
collected values of the missing variables and background characteristics of the sample previously shown to affect missing data were included in the imputation model (Fig. D3S1). Non-normally distributed variables (such as adiposity) were log-transformed to allow imputation. Highly skewed variables such as maternal measures of depression and previous depressive symptom scores were imputed using prediction matching. Categories with very small numbers (such as DAWBA bands 4 and 5, more than three siblings, household income bands, and frequency of exercise, cannabis and alcohol use) were collapsed. Monte Carlo errors were < 10% of standard errors and the maximum fraction of missing information value was 0.43.38

### Results

#### Descriptive statistics

The characteristics of participants actively taking part in the cortisol study are shown in Table 1. Active study participants were more likely to be female, have higher maternal education and a lower prevalence of depression at baseline than non-participants. Distribution of social class, ethnicity, previous depression scores and prevalence of depression at 18 (where this was known) differed little between the groups.

Descriptive statistics for cortisol are given in Table 2. Around 75% of participants showed a rise in their morning cortisol, with

### Table 2

Cortisol values (nmols/l) in all participants

<table>
<thead>
<tr>
<th>Cortisol awakening response</th>
<th>All participants</th>
<th>With depression</th>
<th>Without depression</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>841</td>
<td>46</td>
<td>622</td>
</tr>
<tr>
<td>Mean</td>
<td>3.84</td>
<td>5.07</td>
<td>3.87</td>
</tr>
<tr>
<td>s.d.</td>
<td>4.71</td>
<td>4.61</td>
<td>5.52</td>
</tr>
<tr>
<td>Range</td>
<td>14.60 to 26.11</td>
<td>12.34 to 26.11</td>
<td>4.77 to 24.53</td>
</tr>
<tr>
<td>Interquartile range</td>
<td>1.04 to 6.40</td>
<td>1.28 to 7.93</td>
<td>1.13 to 6.40</td>
</tr>
<tr>
<td>Mean sample time</td>
<td>0 h 35 min</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Waking cortisol

<table>
<thead>
<tr>
<th>All participants</th>
<th>With depression</th>
<th>Without depression</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>887</td>
<td>49</td>
</tr>
<tr>
<td>Mean</td>
<td>7.62</td>
<td>7.11</td>
</tr>
<tr>
<td>s.d.</td>
<td>3.72</td>
<td>3.03</td>
</tr>
<tr>
<td>Range</td>
<td>0.03 to 26.74</td>
<td>2.38 to 17.53</td>
</tr>
<tr>
<td>Interquartile range</td>
<td>5.10 to 9.22</td>
<td>4.75 to 9.22</td>
</tr>
<tr>
<td>Mean sample time</td>
<td>07.03 h</td>
<td></td>
</tr>
</tbody>
</table>

Cortisol sample 2

<table>
<thead>
<tr>
<th>All participants</th>
<th>With depression</th>
<th>Without depression</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>891</td>
<td>49</td>
</tr>
<tr>
<td>Mean</td>
<td>11.30</td>
<td>11.85</td>
</tr>
<tr>
<td>s.d.</td>
<td>5.20</td>
<td>5.70</td>
</tr>
<tr>
<td>Range</td>
<td>0.44 to 33.75</td>
<td>2.79 to 33.06</td>
</tr>
<tr>
<td>Interquartile range</td>
<td>7.81 to 14.25</td>
<td>7.82 to 15.10</td>
</tr>
<tr>
<td>Mean sample time</td>
<td>07.41 h</td>
<td></td>
</tr>
</tbody>
</table>

After school cortisol

<table>
<thead>
<tr>
<th>All participants</th>
<th>With depression</th>
<th>Without depression</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>868</td>
<td>45</td>
</tr>
<tr>
<td>Mean</td>
<td>2.40</td>
<td>2.16</td>
</tr>
<tr>
<td>s.d.</td>
<td>1.60</td>
<td>1.33</td>
</tr>
<tr>
<td>Range</td>
<td>0.00 to 11.20</td>
<td>0.39 to 7.61</td>
</tr>
<tr>
<td>Interquartile range</td>
<td>1.40 to 2.95</td>
<td>1.48 to 2.39</td>
</tr>
<tr>
<td>Mean sample time</td>
<td>16.30 h</td>
<td></td>
</tr>
</tbody>
</table>

Bedtime cortisol

<table>
<thead>
<tr>
<th>All participants</th>
<th>With depression</th>
<th>Without depression</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>930</td>
<td>49</td>
</tr>
<tr>
<td>Mean</td>
<td>1.19</td>
<td>0.92</td>
</tr>
<tr>
<td>s.d.</td>
<td>1.38</td>
<td>0.81</td>
</tr>
<tr>
<td>Range</td>
<td>0.00 to 12.22</td>
<td>0.00 to 4.35</td>
</tr>
<tr>
<td>Interquartile range</td>
<td>0.46 to 1.33</td>
<td>0.39 to 1.16</td>
</tr>
<tr>
<td>Mean sample time</td>
<td>21.49 h</td>
<td></td>
</tr>
</tbody>
</table>

Diurnal drop

<table>
<thead>
<tr>
<th>All participants</th>
<th>With depression</th>
<th>Without depression</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>897</td>
<td>49</td>
</tr>
<tr>
<td>Mean</td>
<td>8.41</td>
<td>8.57</td>
</tr>
<tr>
<td>s.d.</td>
<td>3.79</td>
<td>3.37</td>
</tr>
<tr>
<td>Range</td>
<td>−2.30 to 28.46</td>
<td>2.12 to 17.55</td>
</tr>
<tr>
<td>Interquartile range</td>
<td>5.95 to 10.90</td>
<td>6.65 to 10.85</td>
</tr>
<tr>
<td>Mean sample time</td>
<td>16 h 5 min</td>
<td></td>
</tr>
</tbody>
</table>

Total daily cortisol (area under the curve, g)

<table>
<thead>
<tr>
<th>All participants</th>
<th>With depression</th>
<th>Without depression</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>721</td>
<td>39</td>
</tr>
<tr>
<td>Mean</td>
<td>12.27</td>
<td>11.38</td>
</tr>
<tr>
<td>s.d.</td>
<td>4581</td>
<td>4068</td>
</tr>
<tr>
<td>Range</td>
<td>3712 to 26.646</td>
<td>8813 to 14.005</td>
</tr>
<tr>
<td>Interquartile range</td>
<td>9060 to 14.885</td>
<td>8976 to 14.811</td>
</tr>
<tr>
<td>Mean sample time</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Table 3

Characteristics of participants by quartile of cortisol awakening response (CAR) (n = 841)

<table>
<thead>
<tr>
<th>Attribute</th>
<th>CAR quartile, n (%)</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Baseline and prior depression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>With depression at baseline</td>
<td></td>
<td></td>
</tr>
<tr>
<td>History of depression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>n = 839</td>
<td>8.1</td>
<td>3.8</td>
</tr>
<tr>
<td>832</td>
<td>11.1</td>
<td>12.9</td>
</tr>
<tr>
<td>Background characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td></td>
<td></td>
</tr>
<tr>
<td>841</td>
<td>43.3</td>
<td>42.9</td>
</tr>
<tr>
<td>Social class 4 or 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>774</td>
<td>15.3</td>
<td>12.6</td>
</tr>
<tr>
<td>Post O-level maternal education</td>
<td></td>
<td></td>
</tr>
<tr>
<td>798</td>
<td>42.0</td>
<td>49.5</td>
</tr>
<tr>
<td>Maternal history of depression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>791</td>
<td>7.7</td>
<td>7.0</td>
</tr>
<tr>
<td>Exposed to postnatal depression</td>
<td></td>
<td></td>
</tr>
<tr>
<td>834</td>
<td>16.7</td>
<td>15.9</td>
</tr>
<tr>
<td>Low birth weight</td>
<td></td>
<td></td>
</tr>
<tr>
<td>929</td>
<td>3.0</td>
<td>5.9</td>
</tr>
<tr>
<td>Preterm birth (&gt;32 weeks)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>841</td>
<td>0.5</td>
<td>1.0</td>
</tr>
<tr>
<td>Lifestyle factors</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Highest body fat quartile</td>
<td></td>
<td></td>
</tr>
<tr>
<td>837</td>
<td>22.0</td>
<td>22.6</td>
</tr>
<tr>
<td>Taking oral contraceptive pill</td>
<td></td>
<td></td>
</tr>
<tr>
<td>811</td>
<td>4.4</td>
<td>3.9</td>
</tr>
<tr>
<td>Current daily smoker</td>
<td></td>
<td></td>
</tr>
<tr>
<td>828</td>
<td>3.5</td>
<td>4.8</td>
</tr>
<tr>
<td>Alcohol &gt;2 days per week</td>
<td></td>
<td></td>
</tr>
<tr>
<td>704</td>
<td>2.9</td>
<td>2.2</td>
</tr>
</tbody>
</table>

a. Table includes all participants with a CAR, including those with missing outcome data.
the remainder showing no response or a negative response after waking. The four daily samples displayed the expected diurnal variation. The CAR was normally distributed with very slight right deviation (skewness 0.58).

### Confounding factors

Table 3 shows the characteristics of participants in each quartile group of CAR. The biggest differences were seen for gender, body fat, alcohol and maternal history of depression. There was a higher proportion of baseline depression in the top quartile, however this quartile also had a higher proportion of women.

### Association with depression

Of the participants, 46 (6.89%) were diagnosed with depression at 18 years, 33 of whom were female (71.74%), and 34 of whom (73.91%) were considered ‘new onset’ as they had no previous recorded episodes.

Results of logistic regression for the main cortisol parameters of interest are provided in Table 4. There was no statistical evidence for an association between any of the cortisol measures and depression at follow-up, either before or after adjustment for confounding factors. The unadjusted odds ratio for depression at 18 was 1.24 per standard deviation of CAR (95% CI 0.93–1.66, P = 0.14). Adjusting for confounding factors and lifestyle factors decreased the effect size to 1.12 (95% CI 0.73–1.72, P = 0.61).

There was still no evidence for an association using the first day sample only (which gave the highest value of CAR) (OR = 1.09, 95% CI 0.80–1.48, P = 0.58). A similar result was found when CIS-R ‘depressive symptoms’ was used as a binary outcome (OR = 1.07, 95% CI 0.78–1.48, P = 0.67). Although there was weak evidence for an association between CAR and depressive symptoms at 16 years (OR = 1.22, 95% CI 0.99–1.51, P = 0.06) this was markedly attenuated by confounding factors (online Table DS1). Adjusting for gender alone reduced the odds ratio to 1.11 (95% CI 0.89–1.37, P = 0.36).

There was no evidence that any of the other cortisol parameters including the area under the curve, diurnal drop or total morning cortisol secretion were associated with subsequent depression. There was no evidence that the association between CAR and depression varied by gender (P = 0.69), previous depression (P = 0.44) or maternal history of depression (P = 0.67). There was some evidence for a possible interaction between baseline depression score and CAR (P = 0.06). However, the apparent interaction disappeared after excluding the 42 people with depression at baseline (P = 0.87), suggesting a possible association between CAR and depression resolution or recurrence, but not with depression incidence.

Table 5 shows the characteristics of the cortisol awakening response divided into quartiles. The proportion with depression appears higher in the top and bottom quartiles, suggesting a U-shaped distribution. However, there was no statistical evidence for the possibility when investigated using a quadratic term (P = 0.38). The relationships were attenuated following adjustment for gender, and for adiposity and alcohol use, which were also most common in the top quartile. These factors also provided the biggest attenuation of the relationship between cortisol and depression in the regression analyses.

### Missing data

Imputation of missing confounders increased the final adjusted model numbers to 668. Imputation of missing outcome data increased the sample size to 841, with depression prevalence of 7.61%. Results of logistic regression using both imputed data-sets are provided in Table 4. Despite gains in statistical power, imputation of the missing data did not alter our findings.
The findings from this large prospective cohort study do not support the hypothesis that the CAR, or any other basal salivary cortisol measures including cortisol at each time point, diurnal drop and area under the curve, increase the risk of developing a depressive illness. The small unadjusted association with depression at 16 years in this study may initially appear to corroborate previous suggestions of a time-limited association between the CAR and depression.26 However, as adjusting for confounding factors nullifies this association, we find no evidence for a unique association between cortisol and depression at either time point.

Strengths and limitations

This study has several strengths including a large sample size, which is to our knowledge the largest prospective study using CAR to date. However, this study may still have lacked sufficient statistical power. Although the confidence intervals in this study do overlap those in previous research, the point estimate of the previous study (OR = 2.96) lies well outside our estimates here.17 The upper limit of our confidence interval (OR = 1.7) suggests that if a true difference in CAR does exist, it is small.

Recruitment from the ALSPAC cohort has enabled this study to consider a wealth of potential confounders. The inability of previous studies to account for these without inducing recall bias has been suggested as a potential factor underlying some of the inconsistency among previous studies.19 In our study there was some evidence that adjusting for certain confounding factors, such as alcohol use, would attenuate an apparent relationship with subsequent depression.

It is important to consider both selection and loss to follow-up bias as only a subset of participants provided cortisol samples and there was some attrition over time. This raises the concern that participants who are susceptible to depression, may be less likely to participate in or to complete the study protocol. However, there was a similar prevalence of depression at follow-up in this sample to the rest of the ALSPAC cohort. Additionally, controlling for potentially confounding factors such as social class and baseline depression (which also related to whether data were missing) would have minimised this bias. Finally, we investigated any influence missing data might have made using multiple imputation, and this had no overall impact on our results.

A further strength of this study is the comprehensive protocol for cortisol assessment, which exceeded a ‘minimum protocol’ for salivary cortisol collection in several ways.23 It is possible that a two-sample protocol for CAR calculation may lose some sensitivity compared with a four-sample protocol.23,36 However, a two-sample protocol is well accepted in the literature, and any calculation of CAR in a prospective study has advantages over much of the literature to date. Nevertheless, measurement error is likely to have occurred because of inaccurate reporting of sample collection times, potentially reducing the strength of an association.27 This problem has led to some researchers starting to use automatic timing devices for salivary cortisol collection. However, there is no reason that poor adherence would occur more frequently than in previous studies, especially given the proven levels of cooperation of ALSPAC participants over the years. Measurement bias may also have been introduced by choosing point prevalences of depression at 16 and 18 years, although we do not think this would have introduced a major bias especially given the surge in depressive prevalence in this age group.39

Implications

The findings from this large prospective study do not support the hypothesis that salivary cortisol is associated with the subsequent onset of depression at a population level. Our results suggest that if a true association exists, it is smaller than estimated by previous studies, is markedly attenuated by confounding factors and is unlikely to have any clinical significance. There are a number of reasons why this result appears to contradict previous study findings, which have implications for future research.

It is possible that abnormal HPA function may have a minor influence on the complex aetiology of depression, making the overall population difference small, or that HPA abnormalities are only found in a small subset of depression, which is no doubt a heterogeneous disorder.40 Future prospective studies will need adequate power, requiring large sample sizes to detect small differences, or to investigate distinct subgroups.

The biological complexities of the HPA axis may also account for research inconsistencies. Salivary cortisol alone may be too crude a measure for prospective prediction and a more comprehensive approach may be required.41 Using laboratory-based tests of stress responsiveness, although impractical for studies of this size, may provide different parameters with which the association with depression becomes clearer.

At present a lack of clear hypothesis about which cortisol parameters are associated with future depressive illness may be causing spurious differences in cortisol being identified through multiple testing and selective reporting. Future research should publish both negative as well as positive results of all cortisol measures collected, to ensure a balanced view in literature and avoid publication bias. Furthermore, research to clarify any possible biological link between the HPA axis and depression, including mechanisms and potential parameters for measurement, will facilitate the improved design of any future epidemiological studies.

**Discriminant Validity**

The UK Medical Research Council (Grant reference: 74882), the Wellcome Trust (Grant reference: 07426/1) and the University of Bristol provide core support for ALSPAC. This research was supported by NIH grant R01 MH673842 to T.G.O’C., which funded the CIS-R; and the Wellcome Trust Institutional Strategic Support grant 097825.

**Funding**

**Table 5** Cortisol (nmols/l) and depressive symptoms score given by quartile of cortisol awakening response (CAR)

<table>
<thead>
<tr>
<th>CAR quartile</th>
<th>n</th>
<th>Cortisol mean (s.d.)</th>
<th>Depressive symptoms score, mean (s.d.)</th>
<th>% casesa</th>
<th>OR (95% CI)</th>
<th>P</th>
<th>OR (95% CI)</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>210</td>
<td>−1.48 (2.59)</td>
<td>2.97 (3.69)</td>
<td>6.88</td>
<td>1.00</td>
<td>1.00</td>
<td>0.578</td>
<td>0.324</td>
</tr>
<tr>
<td>2</td>
<td>210</td>
<td>2.17 (0.64)</td>
<td>3.10 (3.62)</td>
<td>5.95</td>
<td>0.77 (0.31–1.90)</td>
<td>0.578</td>
<td>0.47 (0.14–1.58)</td>
<td>0.224</td>
</tr>
<tr>
<td>3</td>
<td>211</td>
<td>4.75 (3.84)</td>
<td>3.22 (3.81)</td>
<td>5.33</td>
<td>0.68 (0.27–1.72)</td>
<td>0.578</td>
<td>0.57 (0.18–1.75)</td>
<td>0.324</td>
</tr>
<tr>
<td>4</td>
<td>211</td>
<td>9.88 (3.53)</td>
<td>3.31 (3.71)</td>
<td>9.36</td>
<td>1.27 (0.31–1.91)</td>
<td>0.564</td>
<td>0.85 (0.30–2.47)</td>
<td>0.772</td>
</tr>
</tbody>
</table>

CIS-R: Revised Clinical Interview Schedule.

a. Percentage of individuals with an ICD-10 diagnosis of depression based on the specified algorithm for the CIS-R.
We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses.

Rebecca Carnegie, BSc, MBChB, Ricardo Araya, MResPsych, PhD, Academic Unit of Psychiatry, School of Social and Community Medicine, University of Bristol, UK; Yoav Ben-Shlomo, BSc, MBBCh, MSc, PhD, FFPH, School of Social and Community Medicine, University of Bristol, UK; Vivette Glover, MA, PhD, DSc, Institute of Reproductive and Developmental Biology, Imperial College London, UK; Thomas G. O’Connor, PhD, Wynne Centre for Family Research, Department of Psychiatry, University of Rochester Medical Center, Rochester, New York, USA; Kieran J. O’Donnell, PhD, Douglas Mental Health University Institute, Department of Psychiatry, McGill University, Canada; Rebecca Pearson, PhD, Academic Unit of Psychiatry, School of Social and Community Medicine, University of Bristol, UK; Glyn Lewis, PhD, Mental Health Sciences Unit, University College London, UK.

Correspondence: Rebecca Carnegie, University of Bristol, Oakfield House, Oakfield Grove, Bristol BS8 2BN, UK. Email: rebecca.carnegie@bristol.ac.uk

First received 14 Jan 2013, final revision 1 Aug 2013, accepted 5 Sep 2013

References