Depression and Hypothalamic-Pituitary-Adrenal Activation: A Quantitative Summary of Four Decades of Research

Cinnamon Stetler, PhD, and Gregory E. Miller, PhD

Objectives: To summarize quantitatively the literature comparing hypothalamic-pituitary-adrenal (HPA) axis function between depressed and nondepressed individuals and to describe the important sources of variability in this literature. These sources include methodological differences between studies, as well as demographic or clinical differences between depressed samples. **Methods:** The current study used meta-analytic techniques to compare 671 effect sizes (cortisol, adrenocorticotropic hormone, or corticotropin-releasing hormone) across 361 studies, including 18,454 individuals. **Results:** Although depressed individuals tended to display increased cortisol ($d = 0.60$; 95% confidence interval [CI], 0.54–0.66) and adrenocorticotropic hormone levels ($d = 0.28$; 95% CI, 0.16–0.41), they did not display elevations in corticotropin-releasing hormone ($d = 0.02$; 95% CI, −0.47–0.51). The magnitude of the cortisol effect was reduced by almost half ($d = 0.33$; 95% CI, 0.21–0.45) when analyses were limited to studies that met minimal methodological standards. Gender did not significantly modify any HPA outcome. Studies that included older hospitalized individuals reported significantly greater cortisol differences between depressed and nondepressed groups compared with studies with younger outpatient samples. Important cortisol differences also emerged for atypical, endogenous, melancholic, and psychotic forms of depression. **Conclusions:** The current study suggests that the degree of HPA hyperactivity can vary considerably across patient groups. Results are consistent with HPA hyperactivity as a link between depression and increased risk for conditions, such as diabetes, dementia, coronary heart disease, and osteoporosis. Such a link is strongest among older inpatients who display melancholic or psychotic features of depression. **Key words:** depression, hypothalamic-pituitary-adrenal axis, cortisol, adrenocorticotropic hormone, corticotropin-releasing hormone, meta-analysis.

HPA = hypothalamic-pituitary-adrenal; **ACTH** = adrenocorticotropic hormone; **CRH** = corticotropin-releasing hormone; **dex** = dexamethasone; **CI** = confidence interval.

INTRODUCTION

Hyperactivity of the hypothalamic-pituitary-adrenal (HPA) axis during depression has been called one of the most reliable findings in all of biological psychiatry (1). Although hundreds of studies have examined HPA axis function during depression, our knowledge of the degree of hyperactivity and the clinical conditions under which it is seen remains incomplete. An estimated 20% to 80% of depressed individuals exhibit some form of HPA hyperactivation, but where in that range the actual figure lies and which demographic or clinical characteristics define the "affected" subgroup remains unclear (2). This is problematic, given the fact that HPA hyperactivity has been put forth as an important mechanism explaining both the pathophysiology of depression itself and its relationship with medical conditions like hypertension, heart disease, stroke, cognitive impairment, diabetes, obesity, and osteoporosis (3). Furthermore, depression is a heterogeneous illness that manifests in a variety of symptom sets, for a variety of reasons, at various points in life, lasts for a varying amount of time, and may or may not respond to treatment. A better understanding of when and where HPA function is disrupted may provide explanations for this variability and, at the same time, help to formulate better models of how depression influences the development and progression of medical illnesses. The primary goal of the current study is to go beyond the general maxim that equates depression and HPA hyperactivity to gain a more nuanced and accurate understanding of the precise magnitude and parameters of this association.

To do this, we have conducted a meta-analytic review that takes advantage of the copious literature on depression and the HPA axis. This type of quantitative synthesis will not only enable us to obtain an estimate of the magnitude of the overall association between depression and HPA axis activity but will also allow us to examine important methodological, demographic, or clinical factors that may moderate the association. We report a series of meta-analyses on hormonal products from all levels of the HPA axis, including corticotropin-releasing hormone (CRH), adrenocorticotropic hormone (ACTH), and cortisol, to obtain a comprehensive picture of the how the entire system is modulated by depression.

METHODS

The meta-analysis included studies that compared HPA axis function in a group of participants with depression to either a) a nondepressed control group; or b) another group with a specific subtype of depressive disorder. To identify relevant studies, computerized searches (on PubMed and PsychINFO) were performed covering all studies published in English through May 2009. A study had to meet several criteria to be included. In addition to making the appropriate group comparisons, participants must have been currently depressed (with depression as the primary diagnosis) at the time of the HPA axis assessment and free from nonpsychiatric medical conditions (e.g., postpartum or poststroke depression). Regarding the dependent variables, a study had to report either basal or postchallenge (with dexamethasone [dex] or dex plus CRH) HPA axis measures. Studies that only described participants as either suppressors or nonsuppressors post dex administration were not included. The final sample included 414 independent studies (Supplemental Digital Content 1 lists the studies included and extracted effects, available at http://links.lww.com/PSYMED/A24).

Coded Variables

For each study that met the inclusion criteria, we coded the number of participants in the depressed and control groups, as well as the mean age of each group. We also coded what percent of each group was female. When

From the Department of Psychology (C.S.), Furman University, Greenville, South Carolina; and the Department of Psychology (G.E.M.), University of British Columbia, Vancouver, British Columbia, Canada.

Address correspondence and reprint requests to Cinnamon Stetler, PhD, Department of Psychology, Furman University, 3300 Poinsett Hwy., Greenville, SC 29613. E-mail: cinnamon.stetler@furman.edu

Received for publication January 4, 2010; revision received November 1, 2010.

Supplemental digital content is available for this article. Direct URL citations appear in the printed text, and links to the digital files are provided in the HTML text of this article on the journal's Web site www.psychosomaticmedicine.org.

DOI: 10.1097/PSY.0b013e31820ad12b

0033-3174/11/7302-0114

reported by the authors, we coded the percent of the depressed group that was in their first depressive episode and the average episode length (in weeks), as a means of assessing chronicity. Depression severity was assessed in several ways. The mean score on any of several standardized clinical measures of depressive symptoms was coded. Other measures of severity included percentage of the depressed sample with a family history of mood disorders, and hospitalization status of the participants (percent inpatient). We also recorded what percentage of the depressed sample was classified as atypical, melancholic, endogenous, or psychotic.

We coded information about several important methodological characteristics of each study. If the method used to diagnose the depressed participants was reported, we coded whether it was a semistructured or structured interview, unstructured interview, or another method, and whether the study used standardized diagnostic criteria. We coded whether the depressed group had been off antidepressant medications for at least 2 weeks at the time of HPA axis assessment. Regarding assessment of the HPA axis, we coded whether it was measured in the morning, afternoon, at night, or continuously. If appropriate, we coded whether cortisol was measured in blood, saliva, urine, or cerebrospinal fluid (CSF), and/or whether CRH was measured in CSF or blood. We coded whether the HPA axis was measured at basal levels or after administration of dex or dex/CRH. Studies that administered another form of exogenous cortisol, ACTH, CRH, and/or metyrapone were included in a separate set of analyses. Finally, we coded whether the HPA axis value was based on a single assessment, or if a mean, total, area under the curve, or slope value was computed. The Supplemental Digital Content 1 lists each study that contributed an effect to the meta-analysis (available at http://links.lww.com/PSYMED/A24).

Calculating Effect Sizes and Analyzing Data

The standardized group mean difference statistic, *d,* was used as the estimate of effect size (4). Effect sizes were calculated such that a positive value represents a higher level in the depressed group compared with control (or depressive subtype) group and they were weighted by the total sample size of the study (4). Studies frequently reported multiple effect sizes from the same sample, introducing dependence into the data. Because effect sizes were nested within samples, we chose a multilevel, mixed-model approach (5). In addition to allowing us to retain the maximum amount of information from the literature by modeling (rather than avoiding) the dependence in our data, this strategy enabled us to specify random, rather than fixed, effects. Random effects models assume a distribution of effect sizes in the population and estimate the average effect size and the dispersion around it (6). We used HLM software (Scientific Software International, Lincolnwood, Illinois) (7) and restricted maximum likelihood estimation (except for model comparisons, when full maximum likelihood was used) to analyze the data.

Please refer to the online text (Supplemental Digital Content 2, available at http://links.lww.com/PSYMED/A25) for a complete description of our literature search, exclusion criteria, coding rationale, multilevel models and analytic strategy.

RESULTS

Cortisol

Descriptives

Three hundred fifty-four studies compared cortisol levels between individuals with major depression and healthy nondepressed individuals. These studies included a total of 18,374 individuals across both depressed and control groups. The average study's depressed sample included 27.6 individuals (standard deviation [SD, 40.7; range, 4–704) and was 60.3% female (SD, 25.1) and 41.8-year-old (SD, 11.6; range, 8.9–78). These studies yielded 535 effect size statistics on a variety of cortisol outcomes.

Overall Effect

The average effect size across all assessments from all studies comparing major depression to healthy controls was $d = 0.60$ (95% confidence interval [CI], 0.54–0.66), which is significantly different from zero, $t(353) = 17.20$, $p < .001$. This roughly corresponds to a medium effect size according to Cohen (8), who classified an effect size of 0.20 as small, 0.50 as medium, and 0.80 as large. The variance component ($\mu = 0.29$) was also significant, χ^2 (353) = 1210.66, $p < .001$, indicating that there was significant variation across studies in the magnitude of the effect size and that further analysis of moderating influences (e.g., methods, clinical features) was justified.

Predicting Effect Sizes: Methodological Factors

We tested two types of predictor variables: within-study variables that represented methodological factors and between-study variables that represented various characteristics of the depressed sample. We tested methodological factors first because these variables may serve as covariates in further analyses. These findings are summarized in the Supplemental Digital Content 3, available at http://links.lww.com/PSYMED/A26.

Time of Day

Cortisol levels in the body generally follow a well-characterized circadian rhythm, and some studies have shown that this rhythm is disrupted during major depression. To determine whether the time of day that the cortisol assessment was done is associated with the magnitude of the effect, we compared effect sizes from samples taken in the morning, afternoon, overnight, and continuously. When cortisol was measured in the morning, the average effect size was $d = 0.49$ (95% CI, 0.41–0.57), whereas samples taken in the afternoon yielded a significantly ($p = .001$) larger effect size of $d = 0.69$ (95% CI, 0.57–0.81). Samples taken at night yielded an average effect of $d = 0.54$ (95% CI, 0.42–0.66), which was not significantly different from morning or afternoon ($p = .49$). When cortisol was measured continuously (across more than one time period), the average effect size was $d = 0.74$ (95% CI, 0.56–0.96), a significantly larger effect compared with morning samples ($p = .006$) but not compared with afternoon or overnight samples ($p > .25$). Cortisol differences between depressed and healthy individuals are smallest in the morning and largest when assessed continuously throughout the day. Studies that assess cortisol levels continuously across several hours necessarily take several cortisol samples. This strategy increases the reliability of the cortisol measure and may, therefore, explain why the effect size for continuous sampling was larger than the others. Time of day accounted for only 2.2% of the within-studies variance.

Bodily Fluid

Studies that assess cortisol in saliva or urine may report different effect sizes compared with studies that measure it from blood or CSF for two primary reasons. First, saliva and urine represent noninvasive modes of assessment. By avoiding the stress-associated release of cortisol elicited during a vein or spinal puncture, assessments in saliva or urine may represent more reliable methods of assessing cortisol. Second,

saliva and urine also contain only unbound cortisol, whereas blood contains both bound and unbound cortisol. We tested this theory by comparing cortisol effects from samples taken in each of these fluids. Assessments done in blood yielded an average effect size of $d = 0.62$ (95% CI, 0.54–0.70). Assessments done in saliva produced an average effect size of $d = 0.33$ (95% CI, 0.15–0.51), whereas those done in urine averaged $d = 0.59$ (95% CI, 0.37–0.81). Effects in saliva were significantly smaller than effects in blood ($t(528) = -3.18$, $p < .01$). Effect sizes from urine samples were not significantly different from those taken from blood samples ($t(528) = -0.24$, $p = .8$). Only five studies assessed cortisol in CSF, and these produced an average effect size of $d = 1.02$ (95% CI, 0.61–1.43), which was significantly larger than the effect size from any other fluid ($p < .04$). Overall, differences in which fluid cortisol was assessed accounted for less than 1% of the variance in effect size. However, differences between depressed and control groups were much smaller when cortisol was measured in saliva compared with any other fluid.

Basal Versus Post Dex

Cortisol levels can also be assessed post administration of dex, a synthethic form of cortisol that is used to mimic the negative-feedback signals that cortisol exerts on the rest of the HPA axis. Researchers typically administer 0.5 mg to 1.5 mg of dex and then assess cortisol levels 12 hours to 24 hours later. Individuals with major depression are believed to exhibit higher levels of cortisol after dex administration compared with nondepressed individuals. This failure to suppress cortisol production after dex is thought to indicate an overactive HPA axis that does not respond to the normal shut-down signals. The size of any group differences in cortisol may vary according to whether the assessment was done on basal levels or after challenging the HPA axis with dex. Studies assessing basal levels of cortisol yielded an average effect size of $d = 0.58$ (95% CI, 0.50–0.60), whereas studies assessing cortisol after administering dex yielded a significantly larger average effect size of $d = 0.70$ (95% CI, 0.60–0.80; $t(530) = 2.24$, $p = .03$). This is similar to the effect size ($d = 0.57$; 95% CI, 0.22–0.93) yielded by the ten studies that also administered exogenous CRH post dex (the dex-CRH test). Thus, the elevations in cortisol during depression are greater when the HPA axis is artificially challenged compared to when it is not. The difference between basal and post dex assessment accounted for 1.6% of the within-study variance in effect size.

Although dex is by far the most frequent drug used to challenge the HPA axis, other synthetic forms of cortisol (e.g., predinsolone, hydrocortisone) have also been used in the literature ($n = 5$). The average effect reported by these studies was $d = 0.33$ (95% CI, −0.24–0.91). This was not reliably nonzero nor was it significantly different from the post dex effect. Studies that reported cortisol differences after stimulating the HPA axis with exogenous ACTH ($n = 9$) yielded a mean effect size of $d = 0.17$ (95% CI, −0.14–0.48). Studies that stimulated with HPA axis with exogenous CRH ($n = 14$) reported a mean cortisol effect of $d = 0$ (95% CI, −0.29–0.28). Seven studies used metyrapone, a drug that blocks cortisol synthesis, to challenge the HPA axis. These studies yielded a mean cortisol effect size of $d = 0.24$ (95% CI, 0.05–0.42). Taken together, these results suggest that, during depression, cortisol's ability to reduce further HPA activity is reliably disrupted.

Operational Definition

We examined whether single assessments yield different effect sizes compared with repeated assessments that were used to calculate either a mean, total, slope, or area under the curve value. Studies that assessed cortisol at a single point reported an average effect size of 0.56 (95% CI, 0.48–0.64). No other cortisol outcome's average effect size differed significantly from this ($p > .2$), except for those studies that calculated a mean value across multiple assessments. Those studies reported an average effect size of $d = 0.85$ (95% CI, 0.65–1.05; $t(526) = 2.79$, $p < .001$). We found a similar pattern of results when we restricted our analysis to only basal samples (single assessments: $d = 0.52$ versus multiple assessments: $d = 0.69$). Measuring cortisol multiple time tends to yield larger effect sizes. However, cortisol outcome accounted for less than 1% of the variance of effect size.

After a series of model comparisons (Supplemental Digital Content 4 provides a detailed description and results of the model comparisons, available at http://links.lww.com/PSYMED/A27), we retained the time of day that cortisol was assessed and whether or not it was assessed post dex administration as covariates in subsequent analyses.

Depressed Group Characteristics

We conducted regression analyses to test whether characteristics (e.g., demographics, severity, presence of depressive subtypes) of the depressed sample predicted effect size, controlling for the above methodological covariates. These findings are summarized in an online table (Supplemental Digital Content 3, available at http://links.lww.com/PSYMED/A26).

Demographics

Age

Cortisol levels have been thought to increase with age, particularly within depressed samples. However, investigations regarding age have usually compared adult samples, and cortisol differences between depressed and nondepressed children and/or adolescents are less established in the literature. To test whether age of the depressed sample predicts effect size, we first entered age as a linear predictor in the regression model. Twenty-seven studies did not report the mean age of their depressed sample and were not included in these analyses. Age did significantly predict effect size ($\gamma = 0\ 0.01$, standard error [SE] = 0.003, $t(323) = 4.02$, $p < .001$), suggesting that cortisol effects are larger as the average age of the depressed group increases. However, the effects of age may not be the same across the lifespan. We reran the model to test whether the quadratic effect of age was significant, and it was ($\gamma = 0.0001$, SE = 0.00004, $t(323) = 2.89$, $p = .004$).

When entered simultaneously in the regression model, only the linear effect remained significant, however ($\gamma = 0.02$, SE = 0.01, $t(322) = 2.43$, $p = .02$).

There were 23 studies that included only participants younger than 18 years. The average age of the depressed participants in these studies was 12.7 years (SD, 2.82; range, 8.9–17.9 years). These studies had an average effect size of $d = 0.21$ (95% CI, 0.07–0.35). On the other end of the age spectrum, 11 studies included only participants aged >65 years. The average age of the depressed group among these studies was 71.0 years (SD, 3.74; range, 65.1–78 years). The average effect size across these studies was $d = 0.71$ (95% CI, 0.28–1.14). Studies that examined adults between 19 years and 64 years had an average effect size of $d = 0.61$ (95% CI, 0.45–0.77). Age accounted for 7% of the between-studies variance. Cortisol differences in depression during childhood/adolescence were significantly smaller compared with cortisol differences during middle adulthood ($t(322) = 4.99$, $p < .001$) or during older adulthood ($t(322) = 2.26$, $p = .02$). Cortisol differences during middle and older adulthood were not significantly different. Among individuals with depression, cortisol levels seemed to increase with age, with the largest rate of increase coming between childhood/adolescence and adulthood.

Gender

To determine whether gender influenced the magnitude of the effect, we entered the percentage of females into the regression model. Twenty-five studies did not report the gender composition of their depressed sample and are not included in these analyses. The gender composition of the sample did not significantly predict effect size ($\gamma = -0.001$, SE = 0.002, $t(325) = -0.93$, $p = .35$). Studies that included all male participants (average $d = 0.60$) were not significantly different ($p = .49$) from studies that included all female participants (average $d = 0.52$). Depression seems to elevate cortisol levels equally in men and women.

Depression Severity

We coded several variables from each study that could be used to evaluate depression severity: intensity of symptoms; length of the depressive episode; chronicity of episodes; hospitalization status; and inclusion of minor depression. We evaluated each variable separately in a regression model to determine which factors influence effect size.

Symptoms

Just over 66% (k = 236) of the original studies administered a symptom severity measure to their depressed participants and reported group means. The vast majority of the studies used some version of the Hamilton Depression Scale; other measures employed include the Beck Depression Inventory, the Zung Depression Scale, the Center for Epidemiologic Studies-Depression Scale, and the Montgomery-Asberg Depression Rating Scale. Higher scores on each of these scales indicate more severe depressive symptoms. Means on each of these measures were standardized so that comparisons across studies could be made.

Symptom severity scores did significantly predict effect size ($\gamma = 0.09$, SE = 0.04, $t(232) = 2.12$, $p < .04$). For example, for every 4.58 points greater on the Hamilton Depression Scale or 5.76 points greater on the Beck Depression Inventory that the depressed group scored, a study reported a 0.09 larger effect size. However, severity scores only accounted for 1% of the between-studies variance in effect sizes.

Chronicity

Only 11% (k = 43) of the original studies reported the duration of the current depressive episode. The average duration (in weeks) was 71.98 (SD, 92.6; range, 4–458). Episode duration did not predict effect size ($\gamma = -0.0005$, SE = 0.002, $t(41) = -0.29$, $p = .78$), probably due to the large amount of variability within this measure. The percent of the depressed sample that was currently in its first depressive episode (as opposed to experiencing multiple episodes and a more chronic depression) was reported by 62 studies. The average percent of sample in a first depressive episode was 39.7% (SD, 28.5; range, 0% to 100%). This variable was not a significant predictor of effect size ($\gamma = -0.004$, SE = 0.003, $t(59) = -1.3$, $p = .19$). Only eight studies reported the percentage of their depressed sample that had a family history of clinical depression (mean %, 52.0; SD, 16; range, 17–68), too few studies to examine meaningfully any associations with effect size.

Hospitalization Status

We also recorded what percentage of the depressed sample was hospitalized. Approximately two thirds of the original studies (k = 227) reported these data. On average, these studies enrolled depressed samples that were 65.8% inpatient (SD, 44.5; range, 0–100). The patient status of the depressed sample was a significant predictor of effect size ($\gamma = 0.004$, SE = 0.001, $t(224) = 4.77$, $p < .001$). Studies that included no inpatients in their depressed samples reported an average effect size of $d = 0.32$ (95% CI, 0.18–0.46), compared with studies that included entirely hospitalized patients ($d = 0.74$; 95% CI, 0.56–0.92). These effect sizes were significantly different from one another ($t(186) = 4.72$, $p < .001$). Inpatient status explained 13.2% of the between-studies variance in effect size.

Because individuals with more severe depression symptoms are more likely to be hospitalized, we wondered if symptom severity could account for the association between cortisol effect size and hospitalization status. When symptom severity scores and hospitalization status were both entered into the model, severity scores were no longer significant ($\gamma = 0.04$, SE = 0.06, $t(151) = 0.87$, $p = .4$), whereas hospitalization status remained significant ($\gamma = 0.005$, SE = 0.001, $t(151) = 4.54$, $p < .001$). Being hospitalized due to depression is associated with increase in cortisol levels independent of symptom severity.

Similarly, studies using outpatients may be more likely to assess cortisol via saliva samples compared with inpatient

studies, which may be more likely to use a blood draw. Recall that effect sizes were smallest when cortisol was assessed in saliva. We wondered if the saliva/plasma difference could account for the effect of hospitalization status. When both bodily fluid and hospitalization status were entered into the model, the plasma/saliva difference was no longer significant ($\beta_1 = 0.47$, SE = 0.12, $t(316) = 1.15$, $p = .25$), whereas hospitalization status remained significant. Thus, the fact that hospitalization status is a significant predictor of cortisol effect size is not explained by the different methods often used to assess cortisol in inpatient versus outpatient settings.

Minor Depression

Another way to examine the effects of severity on effect size is to include studies that assessed cortisol in participants with minor depression. A diagnosis of minor depression (according to the *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition*) requires only two symptoms of depression to be present for more than 2 weeks, compared with the five symptoms needed for major depression. One of the two symptoms must be either anhedonia or depressed mood. Forty-five effect sizes were reported by 27 studies that compared participants with minor depression to nondepressed controls. These studies found an average effect size of $d = 0.29$ (95% CI, 0.07–0.51). Eighty-five effects were reported from 40 studies that compared individuals with major depression to individuals with minor depression. These studies reported an average effect size of $d = 0.31$ (95% CI, 0.17–0.45). These results suggest that individuals with minor depression have cortisol levels somewhere between those of individuals with major depression and those of healthy controls.

Methodological Quality

We found considerable variability in the methodology used across studies. The rigor with which the primary independent variable, depression, was assessed was not consistent. Some studies included only those participants who met the standardized (*Diagnostic and Statistical Manual of Mental Disorders*, Research Diagnostic Criteria, or International Statistical Classification of Diseases and Related Health Problems) criteria for depression ($n = 314$) as determined by a structured or semistructured interview ($n = 158$) to ensure diagnostic reliability. Antidepressants can influence the HPA axis, and many studies ($n = 162$) excluded participants who were taking these medications in the previous 2 weeks. As the above analyses reveal, age has significant effects on HPA axis activity. The majority of studies matched groups on age within 10 years ($n = 318$). Seventy-four studies did all of these things, thus meeting what we defined for this review as minimal standards for methodologic quality. The average effect size for these studies was $d = 0.33$ (95% CI, 0.21–0.45). This effect size is significantly smaller than the mean effect size of studies that did not meet all four of these criteria ($t(352) = -4.78$, $p < .01$), suggesting that diagnostic heterogeneity, antidepressant medications, and/or group differences in age could account for a sizable proportion of the cortisol differences, rather than depression itself.

We wondered if symptom severity or hospitalization status could account for these differences in effect size. A study that excluded participants on antidepressant medications may also include less severely depressed participants or fewer inpatients due to this exclusion criterion. However, we found that there was no significant difference in symptom severity between these studies and the other studies ($t(234) = 0.73$, $p = .47$). Thus, a difference in symptom severity does not seem to be an explanation for the smaller effect size among studies meeting our minimal standards for quality. However, studies that excluded participants on antidepressants also had significantly fewer inpatients compared with studies that did not exclude participants on antidepressants (55% versus 77%; $t(205) = 3.87$, $p < .01$). Given that having a greater percentage of inpatients in the sample is associated with a larger effect size, this variable could account for the difference between these 74 studies (meeting our minimal standards for quality) and the others.

Subtypes

Given the potential importance of diagnostic heterogeneity that these analyses reveal, we examined whether specific subtypes of depression were associated with differences in cortisol levels compared with healthy control groups.

Atypical

Twelve studies reported the percentage of their depressed sample that was classified as having atypical depression. The association between effect size and percent atypical was $\gamma = -0.006$, SE = 0.003, $t(10) = -1.78$, $p = .11$, indicating that when more individuals with atypical depression were included in the sample, effect sizes tended to be smaller. Although the association was nonsignificant due to the small number of studies that reported this information, the magnitude of the association is comparable to that of symptom severity and hospitalization status or the other subtypes of depression we examine. Five studies directly compared a group with atypical depression to a group whose depression was not atypical. The average effect size across these studies was $d = -0.34$ (95% CI, −0.59–0.09), indicating that individuals with atypical depression have cortisol levels approximately one third of an SD lower than individuals with nonatypical depression.

Melancholic

Eighty-six studies reported the percentage of their depressed sample that had melancholic depression. When this predictor (% melancholic) was entered into the regression model, it was significantly related to the magnitude of the effect size ($\gamma = 0.007$, SE = 0.001, $t(84) = 4.71$, $p < .001$). This indicates that, when more individuals with melancholic depression were included in the study, effect sizes tended to be larger. It remained a significant predictor when age was also entered in the model, so this relationship could not be explained by the fact that individuals with melancholic depression also tend to be older. The 38 studies that compared melancholic to nonmelancholic samples had an average effect

size of $d = 0.22$ (95% CI, 0.12–0.32; $t(37) = 4.86$, $p < .001$), suggesting that melancholia increases cortisol almost one quarter of an SD above regular depression.

Endogenous

Fifty-one studies reported the percentage of their depressed sample that had endogenous depression. When this predictor was entered into the regression model, it was positively associated with effect size ($\gamma = 0.008$, SE = 0.002, $t(49) = 3.39$, $p < .01$), indicating that more individuals with endogenous depression in the depressed group tended to produce larger effect sizes. This persisted when age was entered into the model as a covariate, suggesting that this association is not explained by age, even though endogenous depression is more common among older people. Eleven studies directly compared a group of endogenously depressed individuals to a group with nonendogenous depression and reported a mean effect size of $d = 0.29$ (95% CI, −0.04–0.62). Although this effect was not reliably nonzero ($p = .12$) due to the small number of studies included, it is consistent with the idea that endogenous depression is associated with even higher levels of cortisol than nonendogenous depression.

Psychotic

Sixty-six studies reported the percentage of their depressed sample that had psychotic depression. When this predictor was entered into the repression model, it was positively associated with effect size ($\gamma = 0.01$, SE = 0.004, $t(64) = 2.82$, $p < .01$). Seventeen studies directly compared individuals with psychotic depression to individuals with nonpsychotic depression. On average, these studies reported an effect size of $d = 0.47$ (95% CI, 0.18–0.76), indicating that the presence of psychotic features increase cortisol levels by nearly half an SD unit compared with depression without psychotic features.

Overlap With Hospitalization Status

Because individuals with melancholic, endogenous, or psychotic depression may be more likely to be hospitalized than individuals whose depression does not have these features, we wondered if the association between the subtypes and cortisol effects could be accounted for by hospitalization status. When the percentage of the sample that was melancholic/endogenous/ psychotic and the percentage of the sample that was inpatient were entered into the regression, both variables remained significant predictors of effect size. These results suggest that there is something unique about the melancholic/endogenous or psychotic nature of depression that contributes to cortisol hypersecretion, above and beyond inpatient status.

ACTH

Descriptives

Ninety-six studies compared ACTH levels between a group of depressed individuals and a group of healthy nondepressed controls. These studies included a total of 3,812 individuals across both depressed and control groups. The average study's depressed group included 20.7 individuals (SD, 14.7; range, 5–87) and was 42.2 (10.2) years old and was 60.8% (24.9) female. These studies yielded 118 effect size statistics on various ACTH outcomes.

Overall Effect

The average ACTH effect size across all assessments from all studies comparing major depression to healthy controls was $d = 0.28$ (95% CI, 0.16–0.41), which is significantly different from zero, $t(95) = 4.60$, $p < .001$. This suggests that, on average, individuals with major depression have blood ACTH levels over one quarter of an SD higher than healthy nondepressed individuals. The variance component (0.06) was not significant, χ^2 (95) = 115.7, $p = .07$.

Predicting Effect Sizes: Methodological Factors

Similar to what was done for cortisol outcomes, methodological factors were analyzed first because these may serve as covariates in future analyses. These results are summarized in an online table (Supplemental Digital Content 5, available at http://links.lww.com/PSYMED/A28). The relevant methodological factors for ACTH outcomes included: whether the HPA axis was assessed post dex administration or in a basal state; the time of day that the assessment was performed; and the way in which ACTH was operationalized. Each of these variables is defined in the same way as for cortisol outcomes in the previous section. Because ACTH is only measured via a blood draw, the fluid used for the assessment was not relevant to the ACTH analyses.

Time of Day

Studies that assessed ACTH levels in the morning had an average effect size of $d = 0.30$ (95% CI, 0.08–0.52). Performing the assessment in the afternoon ($d = 0.22$; 95% CI, −0.05–0.49), night ($d = 0.27$; 95% CI, −0.02–0.52), or continuously ($d = 0.46$; 95% CI, 0.05–0.85) did not significantly change the effect size ($p > 0.5$). Time of day accounted for just 1.4% of the variance in ACTH effect sizes. Thus, the magnitude of the difference in ACTH secretion between depressed and control groups seems to be relatively stable across the day and night.

Basal Versus Post Dex

The vast majority of effects (k = 103) described basal levels of ACTH, whereas 15 effects described the results for ACTH post administration of dex and 4 post dex/CRH. For basal levels, the average effect size was $d = 0.26$ (95% CI, 0.14–0.38), whereas for post dex levels, the average effect size was $d = 0.48$ (95% CI, 0.11–0.85). For post dex/CRH levels, the average effect size was $d = 0.38$ (95% CI, −0.56–1.32). The average effect size among these three groups of effects was not significantly different ($p = .25$). Whether ACTH was measured at rest or post challenge accounted for less than 1% of the variance.

A handful of studies examined whether ACTH levels differed post administration of exogenous cortisol (k = 3; $d = 0.08$; 95% CI, −0.27–0.43), ACTH (k = 3; $d = -0.12$; 95%

CI, −1.37–1.13), or CRH (k = 16; d = −0.63; 95% CI, −1.19–0.08). Only the CRH effect was reliably nonzero. The negative direction of this effect suggests that, among depressed individuals, the pituitary gland may be less sensitive to CRH signaling, resulting in reduced ACTH production compared with healthy individuals.

Operationalization

Finally, we explored whether the way in which ACTH levels were operationalized had a significant effect. On average, studies using single measurements found an effect size of d = 0.28 (95% CI, 0.14–0.42). Studies using area under the curve reported an average effect of d = 0.13 (95% CI, −0.24–0.50), whereas studies using a mean value reported an average effect of d = 0.36 (95% CI, −0.03–0.75). These effect sizes were not significantly different from that derived from a single measurement ($p > .4$). Operationalization accounted for 5.9% of the variance.

Model Comparison for Methodological Variables

In a similar fashion to cortisol outcomes, we explored whether a regression model that included any or all of the three methodological variables fit the data better than a model that predicted effect size from only the intercept, or average effect size. Our analyses revealed that a model with both basal versus post dex and method of operationalization method fit the data better than the null model (χ^2 (2) = 7.06, $p < .03$). Furthermore, that model fit the data just as well as the full model (with all three variables) but with fewer parameters. Thus, the way in which ACTH was operationalized and whether it was assessed pre or post dex will be retained as covariates in subsequent models.

Predicting Effect Sizes: Participant Variables

Demographic Variables

Age

To explore whether age was a significant predictor of ACTH effect size, we first entered age as a linear predictor in the regression model. Eleven studies did not report the age of their depressed sample and were excluded from this analysis. The average depressed group was 42.2 years old (SD, 10.25), but studies ranged from 9.9 years to 75 years. Age was significantly associated with ACTH effect size in both a linear (γ = 0.01, SE = 0.004, $t(83) = 2.57$, $p < .02$) and quadratic (γ = 0.0001, SE = 0.00004, $t(83) = 2.99$, $p < .05$) pattern. When entered simultaneously into the model, both the linear and quadratic effects of age remained significant. Age accounted for 16.7% of the between-studies variance. There were three studies that measured ACTH in children and/or adolescents (mean age, 10.2 years); the average effect size reported by these three studies was d = 0.18. Among studies of adults (age, 19–65 years), the average effect size was d = 0.25. On the other end of the continuum, three studies measured ACTH in older adults (mean age, 69.0 years). Two of the three studies reported multiple effect sizes, and these effect sizes were averaged within study. Among older adult samples, the mean effect size was d = 0.43. Taken together, these results suggest that the difference in ACTH levels between depressed and nondepressed individuals increases as participant age increases.

Gender

Gender composition (percent female) was not a significant predictor of ACTH effect size (γ = −0.0021, SE = 0.002, $t(86) = -1.05$, $p = .30$), and accounted for only 2.1% of the between-studies variance.

Depression Severity

Symptoms

Sixty-two studies reported the average symptom severity score among their depressed participants. Similar to the cortisol analyses, means on each type of symptom measure (Hamilton Depression Scale, Beck Depression Inventory, etc.) were standardized to permit comparisons across studies. Studies whose participants scored higher on these severity measures tended to report larger ACTH effects (γ = 0.20, SE = 0.08, $t(60) = 2.7$, $p < .02$). Severity accounted for 61.9% of the between-studies variance.

Because age and symptom severity were both significantly associated with effect size, we wondered if each variable had unique predictive value. When entered together in the model, both age and symptom severity were no longer significantly associated with effect size. This was not surprising, given the fact that age and symptom severity were significantly correlated with one another ($r = .29$, $p = .03$). Depressed individuals who are older also tend to have more severe symptoms, and together these factors are associated significantly with higher levels of ACTH.

Chronicity

Similar to the analyses for cortisol effects, we used length of current depressive episode and percentage of depressed sample in the first episode as indicators of chronicity. Among the 16 studies that reported episode length, it was not a significant predictor of effect size (γ = 0.0005, SE = 0.002, $t(14) = 0.24$, $p = .81$). The percentage of the depressed group that was experiencing the first depressive episode was reported by 26 studies. This variable was not a significant predictor of ACTH effect size (γ = 0.003, SE = 0.003, $t(24) = 1.18$, $p = .25$). Thus, samples with more chronic types of depression did not yield reliably greater ACTH effects.

Hospitalization Status

Fifty-nine studies reported whether their depressed participants were inpatients or outpatients at the time of the ACTH assessment. The average sample was 71.3% (SD, 43.7; range, 0–100) inpatient. Unlike the findings with cortisol outcomes, hospitalization status was not a significant predictor of effect size (γ = 0.006, SE = 0.001, $t(57) = 0.42$, $p = .67$) for ACTH outcomes. Studies that included no inpatients (n = 13) reported an average effect size of d = 0.21 (SE = 0.09), whereas studies that included only inpatients (n = 39) re-

ported an average effect size of $d = 0.29$ (SE = 0.13). These average effect sizes are not significantly different ($p = .57$).

Methodological Quality

We attempted to apply the same set of standards for minimal methodological quality to studies that reported ACTH effects as we did to studies of cortisol effects: Depressed participants met the standard diagnostic criteria as determined by a structured or semistructured interview; participants were antidepressant free for 2 weeks before assessment; an age-matched (within 10 years) control group was used. This resulted in 36 studies that reported 42 effects. The average effect size across all of these studies was $d = 0.27$ (95% CI, 0.00–0.54), which was not significantly different from zero or from the larger group of studies. Unlike cortisol outcomes, the magnitude of the effect size for ACTH outcomes was unchanged when only studies that met minimal standards for methodological quality were examined.

Subtypes

We examined whether the presence of atypical, endogenous, melancholic, or psychotic depression was associated with increases or decreases in the observed effect sizes. Only six studies reported the percentage of their depressed participants who displayed atypical symptoms, precluding any meaningful regression analyses. Neither endogenous ($n = 12$), melancholic ($n = 28$), or psychotic symptoms ($n = 12$) were significant predictors of effect size ($p > .20$). Although these null findings suggest that all types of depression result in comparable changes in ACTH levels, more studies that examine these subtypes specifically are needed before firm conclusions can be drawn.

CRH

Descriptives and Overall Effect Size

Sixteen studies examined CRH levels in depressed and nondepressed participants. These studies included a total of 888 individuals and reported 18 CRH effects. The average effect size across these studies was $d = -0.53$ (95% CI, −1.71–0.65). However, one study contributed one effect that was >3 SD from the mean. When this effect was dropped from the analysis, the mean effect size was $d = 0.02$ (95% CI, −0.47–0.51). Without the influence of this single study, these results indicate that CRH levels are not reliably different between depressed and nondepressed individuals. The significant amount of variance between studies ($\mu = 0.92$, $\chi^2(14) = 335.64$, $p < .001$) suggests that important moderators exist.

Predicting Effect Sizes: Methodological Factors

These results are summarized in an online table (Supplemental Digital Content 6, available at http://links.lww.com/PSYMED/A29). CRH levels can be assessed in either blood or CSF. Among studies that assessed CRH in blood (k = 6), depressed groups tended to have higher CRH levels than nondepressed groups ($d = 0.44$; 95% CI, −0.68–1.56). Among studies that assessed CRH in CSF (k = 11), depressed groups tended to have lower levels compared with controls ($d = -0.14$; 95% CI, −0.73–0.45). These effects were not significantly different from zero or one another ($p = .33$), even though they were in the opposite direction. However, these findings suggest that CRH levels in blood and CSF may reflect different physiological processes.

Fifteen studies reported group differences in basal CRH levels ($d = 0.01$; 95% CI, −0.56–0.58), whereas only two studies reported group differences in CRH levels post administration of dex ($d = 0.24$; 95% CI, −0.24–0.74). These effects were not significantly different ($p = .35$). More studies of CRH levels post administration of dex are needed before more reliable conclusions can be drawn.

Fifteen studies captured CRH levels at a single point in time ($d = 0.20$; 95% CI, −1.13–1.53), whereas two studies computed a mean CRH level across multiple assessments ($d = -1.17$; 95% CI, −2.40–0.06). These effects were not significantly different ($p = .06$).

Morning levels of CRH were reported by ten studies ($d = 0.32$; 95% CI, −0.44–1.08); seven studies reported CRH levels taken at other times of the day or night ($d = -0.55$; 95% CI, −1.18–0.08). These effects were significantly different ($t(15) = 2.27$, $p = .04$). CRH levels in the morning tended to be greater in depressed groups compared with controls, whereas the opposite tended to be true at other times of the day and night.

Single CRH assessments from blood samples taken in the morning or post dex administration tend to yield higher CRH levels among depressed groups, whereas assessment of mean basal levels of CRH taken from CSF during the afternoon or overnight tended to yield higher CRH levels among control groups. Given the small number of effects available for analysis, most of these comparisons were not statistically significant and should be interpreted cautiously until more data are available.

Predicting Effect Sizes: Participant Factors

Regarding study-level variables, sufficient data existed for analysis of age, gender, and hospitalization status as potential moderators.

Demographics

Age. Age of the sample was positively (but nonsignificantly) associated with effect size across the 16 studies ($\gamma = 0.05$, SE = 0.04) and accounted for approximately 10% of the between-studies variance.

Gender

The percentage of the depressed group that was female was reported by 15 studies. Gender was also positively (but nonsignificantly) associated with effect size ($\gamma = 0.01$, SE = 0.02) and accounted for approximately 6% of the variance. This finding suggests that the greater percentage of females among the depressed group, the greater the group differences in CRH.

Depression Severity

Hospitalization Status. Finally, 14 studies reported the percentage of the depressed group that was hospitalized at the time of the study. Hospitalization status (percent inpatient) was negatively associated with effect size ($\gamma = -0.02$, SE = 0.007), suggesting that when more inpatients were included in the depressed group, group differences in CRH tended to be smaller (closer to zero) or more negative. Because a negative effect size statistic indicates that the depressed group had lower levels of CRH compared with controls, inpatient status is associated with even lower levels of CRH compared with outpatients and healthy controls in some cases. Hospitalization status explained approximately 13.6% of the between-studies variance. Studies that included inpatients also tended to measure CRH in CSF (i.e., all seven studies that included only inpatients used CSF), so this may account for why CRH levels are lower among depressed inpatients.

Methodological Quality

We were unable to apply all of our criteria for methodological quality because only two studies reporting CRH effects used a semistructured interview, and all of them used standard diagnostic criteria. We did group the studies according to whether they excluded participants who had taken antidepressant medications in the 2 weeks before CRH collection ($n = 10$). We found that these studies had an average CRH effect size of $d = -0.30$, which was not significantly different from that of the studies that did not exclude participants on antidepressants ($d = 0.65$, $t(13) = -1.93$, $p = .08$). This difference accounted for 17% of study-level variance in effect size and suggests that the use of antidepressant medications may account for some of the group differences in CRH levels.

DISCUSSION

Although HPA axis function during depression has been addressed by hundreds of primary reports, this meta-analysis is the first quantitative review to describe the magnitude of the difference between depressed and nondepressed groups in cortisol, ACTH, and CRH. Across all studies, cortisol seems to be elevated by over half an SD unit ($d = 0.60$) among depressed individuals. In other words, approximately 73% of depressed individuals have cortisol values greater than the median cortisol value among nondepressed individuals (of which 50% are greater than the median, by definition). However, when studies that did not meet our minimal standards for methodological quality were excluded, the effect size was reduced by nearly half ($d = 0.33$). Here, only 64% of depressed individuals have cortisol values greater than the median cortisol value among nondepressed individuals. This reduction in effect size may be due to hospitalization status, in addition to diagnostic ambiguity, medication use, or age confounds. This smaller effect size is consistent with the group differences in cortisol ($d = 0.36$) at baseline reported in a meta-analysis of HPA reactivity to stress during depression (9). Across all studies, ACTH levels were elevated to a similar degree during depression ($d = 0.28$), and this effect was robust to controls for methodological quality. This is equivalent to 60% of depressed individuals who have ACTH levels greater than the median ACTH level among controls.

Surprisingly, CRH levels were not significantly elevated among depressed individuals ($d = 0.02$) across the entire pool of studies. Depressed individuals tended to have lower levels of CRH relative to controls ($d = -0.30$) when the influence of antidepressant medications was controlled. This is an important distinction because many antidepressants are thought to improve mood by altering HPA axis function (10,11). It may be the case that CRH is elevated in patients who are too clinically impaired to be off medication, but decreased in those whose impairments are milder or are improving enough for the medications to be discontinued. It is only by eliminating studies that include participants on antidepressant medications that we can disentangle the effects of the drug from the effects of the clinical condition itself. Although this result is based on a limited number of studies and should be interpreted with caution, it is markedly divergent from the widely accepted idea that depression is associated with CRH hypersecretion (12).

CRH effects were in the opposite direction, depending on whether it was measured in blood ($d = 0.44$) or CSF ($d = -0.14$). Unfortunately, neither measure is a good reflection of CRH within the HPA axis. Because CRH from the hypothalamus does not cross the blood-brain barrier, levels in blood are likely to reflect CRH production in places like the adrenal medulla itself, ovaries or testes, peripheral nerves, and lymphocytes (13). It is not clear how strongly peripheral CRH levels are correlated with levels within the central nervous system. Even within the brain and spinal cord, CRH is secreted from areas other than the hypothalamus, such as the prefrontal cortex and parts of the limbic system (13). Thus, CRH levels in the CSF may derive from a variety of sources, only some of which are within the HPA axis. In addition, extrahypothalamic areas increase CRH production in response to cortisol signaling, as opposed to the CRH downregulation that normally occurs in the hypothalamus (14). Other brain areas may alter secretion of CRH to compensate for alterations in CRH production (reflective of disturbed cortisol sensitivity) within the hypothalamus in a failed attempt to maintain homeostasis. Without a method to determine definitively the source of the CRH, it is difficult to determine exactly how CRH production within the HPA axis is altered during major depression.

Taken together, studies meeting minimal methodological standards show that depression is associated with small-to-moderate elevations in ACTH and cortisol and a reduction in CRH levels. Because there is no clinically meaningful cutoff point that separates normal from abnormal HPA axis function, it may be helpful to compare the results of the current study with those of other related meta-analyses. Quantitative reviews of the depression and immune function literature found reductions in lymphocyte activity ($d = -0.50$ to -1.04) (15) and elevations in inflammation ($d = 0.15$ to 0.35) (16). A recent meta-analysis of the change in HPA axis function post

treatment for depression revealed a cortisol effect of $d = 0.73$ and an ACTH effect of $d = 0.55$ (17). Furthermore, a meta-analysis of the chronic stress and HPA axis literature showed that post dex cortisol was significantly elevated in individuals who had developed depression in the context of a chronic stress ($d = 1.13$) (18). Thus, the difference in HPA function between depressed and nondepressed individuals seems to be comparable to the differences in immune function during depression. Furthermore, it is smaller in magnitude than the HPA differences between depressed and nondepressed groups in the context of chronic stress, and smaller than the within-person change in HPA function post treatment for depression.

The overall effect sizes should be interpreted with caution, given the considerable amount of heterogeneity that exists for each HPA axis outcome. In addition to allowing for the aggregation of results across a heterogeneous group of studies and individuals, the real value of this meta-analysis lies in our ability to describe important methodological and sample-level moderators that may explain some of the variance in the overall effect size.

Methodological Factors

The HPA axis can be assessed in a variety of ways, and our results suggest that how it is assessed can make a difference in the magnitude of the effect size observed.

Time of Day

First, time of day was an important moderator of effect size for both cortisol and CRH outcomes. CRH differences between depressed and nondepressed individuals were largest in the morning, whereas cortisol differences were largest in the afternoon (and smallest in the morning). Elevations in ACTH production were consistent throughout the day. Taken together, these findings suggest that cortisol production during depression is influenced by other factors besides ACTH and CRH, and the ACTH production is not solely a function of CRH levels. This idea of dysregulation within the HPA axis is consistent with others' conclusions (14,19) and suggests possible adrenal medullary- or immune-mediated stimulation of cortisol production in the adrenal cortex. Ample evidence supports the existence of these mechanisms (19), and researchers have suggested that they may play an increasingly important role in maintaining elevated cortisol production as ACTH production plateaus and chronic stress persists.

Challenge Tests

For all three hormones, effect sizes were larger after administration of dex (or dex/CRH) compared with basal levels. The administration of dex should result in reduced HPA axis activity, given the negative feedback loop that exists. Our results suggest that this process is less intact in depressed compared with nondepressed persons. Interestingly, stimulating the HPA axis with exogenous CRH seems to produce different results for cortisol and ACTH outcomes. After CRH administration, virtually no difference exists between depressed and nondepressed groups in cortisol production, whereas nondepressed individuals produce substantially more ACTH compared with depressed individuals in response to the same challenge. Although more studies are needed, this pattern of results suggests that the pituitary may become less sensitive to signals from the hypothalamus during depression.

Fluid

The bodily fluid in which the hormone was measured made a difference in the size of both CRH and cortisol effects. When cortisol levels were measured from blood or urine samples, group differences were significantly larger than saliva samples (but significantly smaller than CSF samples). This may reflect the fact that saliva samples are more likely to be taken from outpatients, who tend to show smaller elevations in cortisol. When both bodily fluid and hospitalization status were put in the model, bodily fluid was no longer significant, whereas hospitalization status remained a significant predictor of cortisol effect size.

When measured in CSF, CRH levels were slightly lower in depressed compared with nondepressed groups, whereas CRH levels were higher in depressed groups when measured in blood samples. These inconsistent results suggest that CRH levels in peripheral circulation may not reliably reflect levels in the hypothalamus. Researchers have cautioned against using plasma levels of CRH as a marker of HPA activity due to its indeterminate origins and variability due to assay procedures (14). However, CRH levels in CSF likely include CRH from a variety of extrahypothalamic sources (14). Group differences in CRH should be interpreted with extreme caution, given these inherent assessment limitations.

Participant Characteristics

Age

HPA differences between depressed and nondepressed individuals become larger with age. We found that participant age was significantly associated with cortisol and ACTH effect sizes, and accounted for 10% of the variance among CRH effects. We do not believe this is a function of illness chronicity, because this variable was not associated with effect size. However, to determine more definitively whether it is age or the accumulated impact of a lifetime of depression, future studies should include elderly participants with a significant history of depression and compare them with elderly individuals who are suffering from their first depressive episode.

Symptom Severity and Hospitalization Status

Although we found a positive association between symptom severity and cortisol effect size, this relationship was fully accounted for by hospitalization status. Inpatient samples yielded cortisol effect sizes more than double those from outpatient samples. Larger effect sizes among inpatient samples were also reported by reviews of depression and immune function (15,16). This suggests that at least some of the cortisol hypersecretion seen during depression is due to placement in a hospital setting, rather than the mental illness itself.

This conclusion is further supported by the fact that hospitalization status predicts cortisol effect size above and beyond symptom severity and melancholic, endogenous, or psychotic features. Hospitalization itself is a unique stressor that may disturb HPA function via multiple pathways. First, admission to a mental hospital is associated with suicidality but also with sudden exposure to a novel environment, submission to authority, social isolation, shame, and a loss of individual control, all of which have been associated with elevated cortisol levels independent of depression (20,21). Hospitalization also brings about a dramatic change in circadian rhythms due to institutional demands that can exacerbate cortisol dysregulation. In general, the stress of hospitalization may serve to further dysregulate cortisol production for depressed individuals with an already vulnerable HPA axis. Alternatively, those depressed individuals who have a disproportionately high level of cortisol may have characteristics that also make them more likely to be hospitalized, such as suicidal ideation.

Depressive Subtypes

Atypical Depression

Epidemiological research (22) has concluded that atypical depression is distinct from other forms of the depression. Atypical depression is marked by hypersomnia and fatigue, hyperphagia and weight gain, and emotional reactivity to interpersonal and external circumstances (23). It is present in "pure" form in approximately 15% to 30% of cases of depression (24). Our results suggest that atypical depression is associated with lower levels of cortisol compared with nonatypical depression. Cortisol levels among individuals with atypical depression may not be reliably higher than cortisol levels among healthy nondepressed persons. This is consistent with the notion that atypical depression is a unique form of mood disorder that responds to different medications and has different risk factors and a different clinical course (22). Inclusion of atypically depressed participants may explain why some studies find smaller than average cortisol effect sizes.

Melancholic and Endogenous Depression

In contrast to atypical depression, melancholic depression is characterized by anhedonia, insomnia, loss of appetite, diurnal mood variation, and feelings of excessive guilt and worthlessness (23). Approximately 25% to 30% of depressed individuals display classic melancholic features (24). Findings from the current meta-analysis show that melancholic depression is associated with even greater elevations in cortisol. Melancholic features are associated with 54% larger effect sizes compared with depression without melancholic features. This is consistent with studies that find a greater likelihood of cortisol nonsuppression after the dex suppression test among melancholic individuals (25). Our results also suggest that melancholia is not simply a marker of advanced age or inpatient status, because it remained a significant predictor even when these variables were in the regression model simultaneously. Considerable conceptual overlap exists between the melancholic and endogenous depressive subtypes. Many researchers seem to use these terms interchangeably, and our results suggest that the HPA axis alterations are similar as well.

Psychotic Depression

Approximately 14% to 25% of depressed individuals also report having concurrent delusions or hallucinations (26). Our results suggest that these individuals are more likely to display HPA hyperactivity; psychotic depression was associated with cortisol levels nearly half an SD higher than nonpsychotic depression, and thus >1 SD higher than nondepressed individuals. This finding is consistent with a previous meta-analysis that showed nonsupression of cortisol after the dex suppression test to be significantly more frequent during psychotic depression (27). This association is not surprising, given the connections that exist between the HPA axis and the subcortical dopamine system—that is, the area of the brain most closely associated with psychotic symptoms. In animal studies, cortisol release has been shown to stimulate dopamine production in the striatum in a dose-dependent manner (28). Due to the limits of cross-sectional data, these results cannot tell us whether higher cortisol levels produce delusions and hallucinations or if these psychotic symptoms lead to even higher cortisol levels.

Implications for Medical Illness

HPA axis hyperactivity has been implicated as a potential mechanism through which depression can increase the risk for physical disease and early mortality (29–31). One recent study (32) revealed that, among depressed individuals, cortisol secretion was associated prospectively and positively with risk for death from cardiovascular disease. Consistent with this research, our results suggest that HPA hyperactivity is most likely to be an important disease mechanism in older individuals who are hospitalized with melancholic or psychotic depression. Younger individuals with more moderate symptoms who are not hospitalized are less likely to show marked elevations in HPA axis function and, therefore, will face a smaller (although still elevated) risk for morbidity and mortality via this mechanism. Future studies should continue to examine what degree of HPA axis dysfunction, for how long, is necessary to confer increased diseased risk. Researchers who wish to examine HPA hyperactivity as a disease mechanism in depression may be best served by including older inpatients in their research. This group is more likely to display the dramatic cortisol elevations that are more likely to disrupt immune, metabolic, and neurocognitive functioning.

Our results are also consistent with the idea that different symptom clusters are associated with different patterns of HPA axis activity. This may have implications for which depressed individuals are at risk for which types of disease. Individuals displaying hypersomnia, hyperphagia, fatigue, and emotional reactivity may be less likely to suffer from hypercortisolemia and its medical consequences. Conversely, individuals with insomnia, loss of appetite, anhedonia, diurnal

mood variability, and/or delusions or hallucinations are more likely to display hypercortisolemia. These individuals are, therefore, more likely to be at risk for conditions where excessive cortisol is thought to play a pathogenic role, such as Type II diabetes, osteoporosis, or dementia (33,34).

Limitations and Future Directions

The sheer number of studies included in this meta-analysis serves as both its primary strength and weakness. The large number of studies creates a high degree of heterogeneity that is difficult to summarize meaningfully with a single effect size statistic. However, it is because we have so many studies that we can investigate important moderating variables with adequate statistical power. Although we found that several important moderators exist, it is still possible that we have left out other significant moderating factors. We are limited by the data reported in the individual studies.

Because the vast majority of our effect sizes come from studies published in peer-reviewed journals, it is possible that our results are biased due to the higher likelihood that published studies will contain significant results compared with unpublished reports (the file-drawer effect). However, the large number of studies included in the current meta-analysis also serves as a check; our analyses (not reported here but available on request) suggest that, in the vast majority of cases, it is very unlikely that our significant findings are due to the exclusion of unpublished reports with nonsignificant findings.

One important limitation of any meta-analysis is that it cannot exceed the design limitations of the studies on which it is based. Our report is based on cross-sectional studies (by design); thus, we are unable to draw any conclusions about causality. HPA disturbances may be a contributing factor, a correlate, or a consequence of depressive symptomatology. Although it has been shown that HPA axis dysregulation resolves after successful treatment for some depressed individuals, the current study is unable to address this issue (17). Second, we have attempted to apply a set of arbitrary criteria for minimal methodologic quality. We acknowledge that our set of standards may be idiosyncratic, although they are similar to those used in other meta-analyses of correlational studies (15). Our set of standards alone does not ensure high quality, and ideally other criteria, such as multiple HPA axis assessments or matching female participants on menstrual cycle phase, could also be applied. Our set of standards was the lowest threshold we could meaningfully set and still retain enough studies to analyze. Another limitation is that cortisol elevations, however large, are only one step in any eventual disease process. Without evidence of how target tissues are responding to the elevations in cortisol, we can only infer the metabolic, vascular, or immune implications of HPA dysregulation. This is important because it is these processes, rather than increased cortisol levels, that are more closely linked to disease outcomes (18,35). Future studies should take advantage of improvements in bioinformatic technologies and investigate the downstream consequences of cortisol elevations, such as changes in receptor levels, gene expression, or tissue remodeling.

Furthermore, most of the studies included in the current meta-analysis were underpowered. To detect a $d = 0.33$ with 0.80 power and $\alpha = 0.05$, one would need a total sample size of 237. Only three (of 355) studies had this number of participants. Future studies should attempt to include enough participants to have adequate statistical power. This can be a challenge when conducting research with a clinical population such as this. However, this challenge can be met by conducting longitudinal research that assesses HPA axis function in the same individual over time. Studies with a prospective design will not only have adequate statistical power with fewer subjects but they will add to the literature by addressing questions of directionality that cross-sectional studies (with which the literature is replete) cannot.

Reliable increases in ACTH and cortisol production during depression seem to exist, although these increases are not entirely consistent. ACTH production is greatest when depressed participants are older and have more severe symptoms. Cortisol production is greatest when it is assessed as a mean value in blood in the afternoon post HPA axis challenge or as total output in urine. Depressed participants who are older, hospitalized, have melancholic, endogenous, or psychotic features, and are taking antidepressant medications will also have the highest cortisol levels. Cortisol elevations, if present, are much smaller among individuals with atypical depression. Although CRH levels were not reliably elevated during depression, they were most likely to be elevated when CRH was measured from blood samples in older depressed women and in individuals taking antidepressant medications. Taken together, this quantitative analysis of >400 studies reveals a complex picture of HPA axis activity during depression. This heterogeneity within a single diagnostic label should be taken into account more fully as existing theories of disease are refined and new ones emerge in the field of psychosomatic medicine.

We thank Laurel Patterson, Matthew Hill, and Jane Woo for their valuable contributions to this project.

REFERENCES

1. Pariante CM, Lightman SL. The HPA axis in major depression: classical theories and new developments. Trends Neurosci 2008;31:464–8.
2. Thase ME, Jindal R, Howland RH. Biological aspects of depression. In: Gotlieb IH, Hammen CL, editors. Handbook of Depression. New York: Guilford Press; 2002.
3. Brown ES, Varghese FP, McEwen BS. Association of depression with medical illness: does cortisol play a role? Biol Psychiatry 2004;55:1–9.
4. Hedges LV, Olkin I. Statistical Methods for Meta-Analysis. Orlando, FL: Academic Press; 1985.
5. Kalaian HA, Raudenbush SW. A multivariate mixed linear model for meta-analysis. Psychol Methods 1996;1:227–35.
6. Hedges LV, Vevea JL. Fixed and random-effects models in meta-analysis. Psychol Methods 1998;3:486–504.
7. Raudenbush S, Bryk A, Congdon R. HLM 6: Hierarchical Linear and Nonlinear Modeling [computer software]. Chicago: Scientific Software International; 2000.
8. Cohen J. Statistical Power Analysis for the Behavioral Sciences. 2nd ed. Hillsdale, NJ: Erlbaum; 1988.
9. Burke HM, Davis MC, Otte C, Mohr DC. Depression and cortisol

responses to psychological stress: a meta-analysis. Psychoneuroendocrinology 2005;30:846–56.
10. Schule C. Neuroendocrinological mechanisms of actions of antidepressant drugs. J Neuroendocrinol 2007;19:213–26.
11. Schule C, Baghai TC, Eser D, Rupprecht R. Hypothalamic-pituitary-adrenocortical system dysregulation and new treatment strategies in depression. Expert Rev Neurother 2009;9:1005–19.
12. Binneman B, Feltner D, Kolluri S, Shi Y, Qiu R, Stiger T. A 6-week randomized, placebo-controlled trial of CP-316,311 (a selective CRH1 antagonist) in the treatment of major depression. Am J Psychiatry 2008; 165:617–20.
13. Kaltsas GA, Chrousos G. The neuroendocrinology of stress. In: Cacioppo JT, Tassinary LG, Bersani G, editors. Handbook of Psychophysiology. 3rd ed. Cambridge: Cambridge University Press; 2007.
14. Owens MJ, Nemeroff CB. Physiology and pharmacology of corticotropin-releasing factor. Pharmacol Rev 1991;43:425–73.
15. Herbert TB, Cohen S. Depression and immunity: a meta-analytic review. Psychol Bull 1993;113:472–86.
16. Howren MB, Lamkin DM, Suls J. Associations of depression with C-reactive protein, IL-1, and IL-6: a meta-analysis. Psychosom Med 2009;71:171–86.
17. McKay MS, Zakzanis KK. The impact of treatment on HPA axis activity in unipolar major depression. J Psychiatr Res 2009.
18. Miller GE, Chen E, Zhou ES. If it goes up, must it come down? Chronic stress and the hypothalamic-pituitary-adrenocortical axis in humans. Psychol Bull 2007;133:25–45.
19. Bornstein SR, Chrousos GP. Clinical review 104: Adrenocorticotropin (ACTH)- and non-ACTH-mediated regulation of the adrenal cortex: neural and immune inputs. J Clin Endocrinol Metab 1999;84:1729–36.
20. Dickerson SS, Kemeny ME. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. Psychol Bull 2004;130:355–91.
21. Taylor SE. Hospital patient behavior: reactance, helplessness, or control? In: Friedman HS, DiMatteo MR, editors. Interpersonal Issues in Health Care. New York: Academic Press; 1982.
22. Matza LS, Revicki DA, Davidson JR, Stewart JW. Depression with atypical features in the national comorbidity survey: classification, description, and consequences. Arch Gen Psychiatry 2003;60:817–26.
23. American Psychiatric Association. Diagnostic and Statistical Manual of Mental Disorders. 4th ed. Washington DC: American Psychiatric Association; 1994.
24. Gold PW, Chrousos GP. Organization of the stress system and its dysregulation in melancholic and atypical depression: high vs low CRH/NE states. Mol Psychiatry 2002;7:254–75.
25. Rush AJ, Weissenburger JE. Melancholic symptom features and DSM-IV. Am J Psychiatry 1994;151:489–98.
26. Ohayon MM, Schatzberg AF. Prevalence of depressive episodes with psychotic features in the general population. Am J Psychiatry 2002;159: 1855–61.
27. Nelson JC, Davis CL. DST studies in psychotic depression: a meta-analysis. Am J Psychiatry 1997;154:1497–503.
28. Fleming SK, Blasey C, Schatzberg AF. Neuropsychological correlates of psychotic features in major depressive disorders: a review and meta-analysis. J Psychiatr Res 2004;38:27–35.
29. Frasure-Smith N, Lesperance F. Reflections on depression as a cardiac risk factor. Psychosom Med 2005;67 Suppl 1:S19–S25.
30. Glassman AH, Shapiro PA. Depression and the course of coronary artery disease. Am J Psychiatry 1998;155:4–11.
31. Evans DL, Charney DS, Lewis L, Golden RN, Gorman JM, Krishnan KR, Nemeroff CB, Bremner JD, Carney RM, Coyne JC, Delong MR, Frasure-Smith N, Glassman AH, Gold PW, Grant I, Gwyther L, Ironson G, Johnson RL, Kanner AM, Katon WJ, Kaufmann PG, Keefe FJ, Ketter T, Laughren TP, Leserman J, Lyketsos CG, McDonald WM, McEwen BS, Miller AH, Musselman D, O'Connor C, Petitto JM, Pollock BG, Robinson RG, Roose SP, Rowland J, Sheline Y, Sheps DS, Simon G, Spiegel D, Stunkard A, Sunderland T, Tibbits P Jr, Valvo WJ. Mood disorders in the medically ill: scientific review and recommendations. Biol Psychiatry 2005;58:175–89.
32. Jokinen J, Nordström P. HPA axis hyperactivity and cardiovascular mortality in mood disorder inpatients. J Affect Disord 2009;116:88–92.
33. Rosmond R, Bjorntorp P. The hypothalamic-pituitary-adrenal axis activity as a predictor of cardiovascular disease, type 2 diabetes and stroke. J Intern Med 2000;247:188–97.
34. Lupien SJ, de Leon M, de Santi S, Convit A, Tarshish C, Nair NP, Thakur M, McEwen BS, Hauger RL, Meaney MJ. Cortisol levels during human aging predict hippocampal atrophy and memory deficits. Nat Neurosci 1998;1:69–73.
35. Miller G, Chen E, Cole SW. Health psychology: developing biologically plausible models linking the social world and physical health. Annu Rev Psychol 2009;60:501–24.