Psychoneuroendocrinology 80 (2017) 26–35

Contents lists available at ScienceDirect

Psychoneuroendocrinology

journal homepage: www.elsevier.com/locate/psyneuen

Review

Meta-analytical assessment of the effects of protocol variations on cortisol responses to the Trier Social Stress Test

William K. Goodman[a], Johanna Janson[b], Jutta M. Wolf[a,*]

[a] Brandeis University, Psychology Department, 415 South St. MS062, Waltham, MA 02454 USA
[b] Friedrich-Alexander-Universität Erlangen-Nürnberg, Nägelsbachstr. 49a, 91052 Erlangen, Germany

ARTICLE INFO

Article history:
Received 8 December 2016
Received in revised form 2 February 2017
Accepted 27 February 2017

Keywords:
Cortisol
Stress
TSST
Confounds
Protocol variations
Research methodology

ABSTRACT

Background: The Trier Social Stress Test (TSST) is one of the most widely used laboratory stress tests. Exposure to this psychosocial stressor has been shown to stimulate an acute cortisol stress response in the majority of healthy individuals, while deviations from the typical pattern, i.e., cortisol reactivity dysfunctions have been linked to an ever-increasing number of negative health outcomes. However, significant variability between labs exists in strength of observed cortisol responses in healthy individuals. This variability raises the question of how to distinguish across labs between cortisol stress response patterns that reflect health risk from those that are due to methodological differences. Thus, we propose a systematic review and meta-analysis that aims at quantifying the effects of methodological variation in study and TSST protocol elements on cortisol stress responses in healthy individuals.
Methods: Literature searches were conducted using standard databases for English language with key words including Trier Social Stress Test, TSST, Cortisol, and Laboratory Stressor among others. 186 studies met our inclusion criteria of healthy human participants without systemic immunological or endocrine dysfunction and provided sufficient information to compute a total of 237 sub-sample effect sizes.
Results and discussion: With regard to study protocol variations that may risk confounding baseline cortisol values and thus influence subsequent reactivity measures, meta-analytical examination revealed that acclimation periods pre-TSST below 30 or perhaps even 15 min may suffice, at least as long as no interfering activities, i.e., questionnaires, are taking place during that timeframe. Assessing the effects of TSST protocol variations on cortisol response strength, several observations are noteworthy. First, shortening speech preparation time did not change cortisol responses in any way, nor did including questionnaires during that period show an effect. As such, our findings suggest that speech preparation time is one TSST element that can be used to reduce the burden for participants as well as laboratory logistics. Secondly, having an all female panel and instructing panel members to show negative instead of neutral behavior towards the participants both were associated with considerably reduced cortisol stress response strengths. Thirdly, several variables of interest, such as content of the speech task or gender match between active panel member and participant, were problematic to evaluate due to the large number of studies not reporting those details. This calls for future studies to report more details regarding potentially relevant protocol specifications.

Contents

* Corresponding author.
E-mail addresses: wgoodman@brandeis.edu (W.K. Goodman), johanna.janson@fau.de (J. Janson), jmw@brandeis.edu (J.M. Wolf).

http://dx.doi.org/10.1016/j.psyneuen.2017.02.030
0306-4530/

1. Introduction

Stress is a ubiquitous experience for human beings, and is associated with activation of stress systems, most prominently the hypothalamus-pituitary-adrenal (HPA) axis (Kirschbaum et al., 1993). Although today, typical stressors no longer present as physical threats but most frequently are of social evaluative nature (Dickerson et al., 2009), associated HPA axis dysfunctions are still linked to a host of negative health consequences (Johnson et al., 1992). This stands in contrast to the idea that evolutionarily, acute HPA axis responses to environmental threats are regarded adequate and beneficial (de Kloet et al., 1999).

The need to increase our understanding of the mechanism linking social-evaluative stressors to mental and physical well-being has led to a shift in research efforts to focus on beneficial and dysfunctional physiological processes activated by social evaluative stressors (Allen et al., 2014; Dickerson and Kemeny, 2004). To study health-relevant physiological stress mechanisms, it became imperative to develop a laboratory protocol that would reliably activate the HPA axis as assessed by increases in its end hormone cortisol, in response to a psychosocial stressor. In 1993, Kirschbaum and colleagues introduced the Trier Social Stress Test (TSST), which was developed with this exact aim in mind. Subsequent studies (Campbell and Ehlert, 2012; Foley and Kirschbaum, 2010; Het et al., 2009) and also meta-analyses (Dickerson and Kemeny, 2004) of TSST studies confirmed strong and robust cortisol stress responses. Not surprisingly, the TSST has become one of the most widely used laboratory stress tests with numerous studies utilizing this stress inducing paradigm to evaluate questions related to stress effects on learning (Joëls et al., 2006), memory (Cornelisse et al., 2011; Luethi et al., 2009; Wiemers and Wolf, 2015), mental health (Gerber et al., 2013) and physical health (Kemeny, 2003; Miller et al., 2009).

The TSST as originally specified (Kirschbaum et al., 1993) consists of a ten minute speech preparation period, followed by a ten minute testing period, during which the participant is asked to give a five minute job application speech followed by five minutes of mental arithmetic. Participants are informed that their performance will be video and audio recorded for later analyses of verbal and non-verbal behavior. During the testing portion, the participant is asked to stand in front of three judges who provide neutral feedback in response to the participant's speech and arithmetic accuracy. The development of the exact test protocol was closely informed by prominent stress theories (Kirschbaum et al., 1993). That is, it incorporates stressor elements that speak to the Mason factors (Mason, 1975) novelty, unpredictability, uncertainty, ambiguity, and ego-involvement, and it acknowledges Lazarus' stress theory (Lazarus, 1993) emphasizing cognitive appraisal processes and predicting that situations in which the demands exceed one's recourses are perceived as stressful. The strong social-evaluative component of the TSST further aligns with predictions of theories emphasizing social identity threat and self-preservation processes (Dickerson et al., 2004; Townsend et al., 2011).

Despite the TSST protocol being theoretically specified and outlined minutely, significant variability between labs exists in the strength of the observed cortisol responses in healthy individuals (Kudielka et al., 2007). This variability raises the question of how to distinguish between differences in response patterns reflecting individual differences in stress experience and consequences such as health risks, from those that are due to lab-dependent differences in TSST protocols. A brief look at the TSST literature reveals differences between laboratories with regard to both cortisol sampling (Bellingrath and Kudielka, 2008; Buchanan et al., 2014) protocols (number of samples, sampling frequency) as well as differences in the exact implementation of the TSST protocol (von Känel et al., 2006). The latter entails not only the 20-min TSST protocol itself but its embedding into the overall study protocol as well. Importantly, both types of protocol variations may influence the strength of the assessed cortisol stress response.

1.1. Cortisol sampling protocol

As mentioned above, cortisol is widely used as a measure of HPA axis activity and reactivity. The cascade-like nature of the HPA axis with two hormones – CRH and ACTH – preceding cortisol release as well as relatively slow signal transport by blood results in cortisol increases occurring 20–30 min after stress onset (Kirschbaum et al., 1993). Depending on the exact length of the TSST protocol, this translates into maximum cortisol increases being observed 0–30 min post-TSST. However, in addition to effects related to TSST length, sampling protocol variations risk confounding the differences observed between studies in cortisol stress responses. Specifically, samples may be taken too infrequently during this important time frame and thus studies may miss capturing the full extent of the cortisol stress response. The first aim of the current review was thus to examine cortisol sampling protocol variations in terms of frequency and timing and their effects on capturing maximum cortisol increases.

1.2. Study and TSST protocol

Further examination of studies using the TSST reveals the introduction of numerous deviations from the original protocol specifications as well as differences between labs with regard to the more general study setting leading up to the TSST. With regard to the latter, protocol variations that may influence cortisol stress response strength due to the previously described time dynamics of cortisol release may be particularly interesting to consider. These include differences between study protocols in how much time is allotted to allow participants to recover from potential pre-study stress and to acclimate to the study setting (Dickerson and Kemeny, 2004). Closely related to these considerations are activities taking place during the pre-TSST period that may interfere with pre-stress recovery and acclimation, such as answering questionnaires, or that are themselves stressful, such as receiving an intravenous catheter. Lastly, as basal cortisol levels show a well-described circadian rhythm (Clow et al., 2004), considerable pre-TSST baseline differences are expected depending on the time of day a TSST study takes place. Thus, the second aim of the current review was to examine the effects of physiologically relevant study protocol variations on maximum cortisol stress response strength.

TSST protocol modifications, on the other hand, may influence the strength of cortisol stress responses by changing the perception of the stressfulness of the TSST. For example, receiving neutral feedback is rather unusual in social interactions and from a stress theory perspective, represents not only a novel and ambiguous situation, but the lack of experience minimizes the resources a participant can draw on to deal with it (Dykman et al., 1989). This would be less true in instances in which study protocols have their TSST panel members give negative feedback to the participant (Swann et al., 1992). Other relevant variations may include the length of the various TSST components (Engert et al., 2013), such that a lower limit may exist to allow for a measurable cortisol change to occur, and content differences (Al'Absi et al., 1997), such as changing the free speech period to a question and answer period or using arithmetic tasks that vary in difficulty (Strahler et al., 2010) from the original serial subtraction task (Kirschbaum et al., 1993). Although frequently implemented, to date the effects of these TSST protocol modifications have not been systematically evaluated. Hence, the third aim of the current review was to examine the effects of psychologically relevant TSST protocol changes on maximum cortisol stress responses.

1.3. Study aims

In summary, while many studies have demonstrated the reliability of the TSST in terms of inducing significant increases in cortisol, no systematic review exists on the effects of study protocol variations on cortisol stress response strength for TSST studies only. The purpose of this review was thus composed of three main goals: (1) A systematic examination of cortisol sampling protocols in terms of frequency and timing with the aim of developing a recommendation for best practices to capture maximum cortisol increases. (2) Using meta-analytic statistical techniques to examine effect size differences due to physiologically relevant study protocol variations, including time of day the TSST is taking place, pre-TSST acclimation period duration, and activities during that period. (3) Using meta-analytic statistical techniques to examine effect size differences due to psychologically relevant variations in TSST protocol implementation with a focus on composition of TSST elements, such as speech preparation duration and difficulty of presentation and arithmetic tasks, and factors relating to the social aspect, i.e., the TSST panel, such as type of feedback given to the participant, size of panel, or panel gender composition.

2. Methods

2.1. Search strategy

A detailed search of electronic databases was performed (articles published by early 2016) for original research articles published since 1993 in both English and German languages. Databases included: Academic Search Premier, PsychInfo, PsychArticles, Psyndex. Searches included the Authors' own bibliographic libraries. The following search terms were utilized singularly and in combination to identify articles of interest: TSST, Trier Social Stress Test, Cortisol, HPA axis, Modified TSST/Trier Social Stress Test, HHNA (Hypothalamus-Hypophysen-Nebennierenrinden-Achse [English translation: HPA axis]). After removal of duplicate studies (n = 630) and studies reporting results from TSST-1 toxic shock syndrome toxin (n = 324), a total of 757 articles were assessed for potentially fulfilling our selection criteria.

2.2. Inclusion/exclusion criteria

This meta-analysis included original research articles assessed in previous systematic reviews (Dickerson and Kemeny, 2004) and new studies published subsequent to those reviews. Studies were required to have participants exposed to the TSST or a slightly modified version of the TSST without additional stressor or treatment exposure prior to the TSST (e.g., memory testing, oxytocin administration; n = 94 excluded). Studies identified as review papers/secondary data analysis studies (n = 9) were excluded as well.

Adult participants were the primary focus of this meta-analysis due to the confounding effects of TSST protocol modification for younger participants (Stroud et al., 2009), leading to the exclusion of n = 55 studies solely focused on children. For adult participants, only non-clinical healthy samples were included (n = 202 excluded; note: participants in controls groups in clinical studies free from any disorders influencing cortisol reactivity were included). This approach was chosen to aid in answering Aim 1, specifically, to avoid complex effects of disease status (i.e., varying by diagnosis, definition, and extent of interaction between physiological disease mechanisms with HPA axis activity and/or reactivity) to influence cortisol stress response patterns and thus confounding effects of cortisol sampling protocols. This resulted in 397 studies to be considered for further review.

2.3. Analysis of study methodology/outcome measures

All studies in this meta-analysis were required to have at least one cortisol (salivary or plasma) assessment before and after the TSST. Studies were examined for suitability across multiple methodological variables including: participant's age, sample size, number of cortisol measurements, and degree to which the TSST was modified (e.g. performing mental arithmetic only). If study protocols included multiple TSST exposures, only data pertaining to the first exposure were extracted. This meta-analysis examined the statistical relationship between methodological changes to the TSST and magnitude of cortisol increases. Methodological variations of interest were a) study protocol variations affecting pre-TSST cortisol values: Time of day the TSST was conducted, Time spent in lab before the TSST, and pre-TSST activities, and b) TSST protocol variations affecting HPA reactivity: speech preparation duration, questionnaires administered during speech preparation, speech composition, arithmetic difficulty, size of jury, gender composition of jury panel, gender of juror providing instructions to the participant, and juror feedback.

Effect sizes were computed following standard guidelines (Cumming, 2013) with cortisol increases defined as standardized

mean differences (SMD) between the author-identified baseline cortisol assessment, and the maximum cortisol concentration subsequent to the TSST. On the backdrop of described cortisol kinetics, this approach allowed for comparison across studies independent of sample collection methods (e.g., plasma versus saliva) and the respectively reported cortisol fractions (i.e., total versus free). For all cases the SMD was calculated using raw data provided by the study authors or graphical/tabular data reported in published articles and was calculated by $d = \frac{Cortisol_{Max} - Cortisol_{Baseline}}{(SD_{Max} + SD_{Baseline})/2}$ and meta-analytic effect sizes denoted as *d'*. Cortisol collection times were standardized to the start of the TSST preparation period (as time point 0). As we did not have access to the original data for each study, we chose a random effect meta-analysis excluding covariates (e.g., gender or age) to avoid aggregation bias limiting interpretation of results (Thompson and Higgins, 2002).

2.4. Data collection

Data for this study were collected by two primary methods: online surveys and extraction of information from published articles. Data extraction was originally conducted on the 397 published articles and focused on TSST protocols and cortisol responses, but did not always yield satisfactory results. In an effort to obtain the most comprehensive dataset, the corresponding authors were contacted and asked to complete an online survey (hosted by Qualtrics) designed to evaluate both the published and unpublished details of relevant manuscripts. The authors of 85 studies graciously provided information. Qualtrics was configured to ask authors at which time of day they conducted their TSST (in one hour increments), how long participants were in the lab before the TSST (min), whether participants completed questionnaires or had an IV placed during their pre-TSST time, how long participants were given to prepare for the speech portion of the TSST (min), whether participant completed questionnaires during their speech preparation time, the composition of the speech portion of the TSST (e.g. job talks, question and answers, other/please specify), starting numbers and subtraction numbers used in the arithmetic tasks, and jury composition (number of jurors, genders, ages, instructions) and behaviors (neutral or negative feedback). Guided by these responses and the information coded for, the following groupings were extracted: time of day the TSST was conducted (morning <12p.m., lunchtime 12p.m.–2p.m., afternoon >2p.m., and mixed TSST starting times) and time in lab pre-TSST (0–15 min, 16–30 min, 31–60 min, and greater than 1 h). Speech preparation times (3 min, 5 min, & 10 min), jury numbers (2 or 3 jurors) and arithmetic subtraction number (13, or 17) were chosen for their universal nature with few studies (<10%) choosing other options.

Minor publication biases favoring stronger effects were evident in diagnostic plots, with slight asymmetry in funnel plots (observed outcome by standard errors). However, concerns of overestimation of overall effects were alleviated to an extent by the consistency in findings between the current meta analysis and previous studies establishing the robust cortisol response to the TSST (Dickerson and Kemeny, 2004).

Cortisol values were obtained from reported values, or estimated from graphical results (400% enlarged graph, proportional distance measures [in pixels] from x-axis, scaled to y-axis cortisol concentration values). Studies where authors were unable to be contacted and published articles did not provide extractable cortisol data were excluded (n = 210). In total 186 (Qualtrics n = 85, coded n = 101, 8592 total participants) of the originally identified 397 studies were included in the final analysis (see Table 1), and contributed a total of 237 sub-samples (see Supplemental Tables 1–3). Meta-analytic effect size estimates were calculated using R 3.3.1 and Metafor 1.9-7.

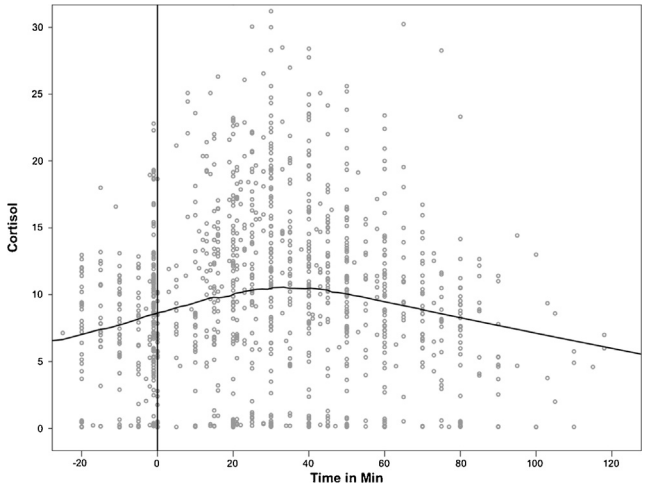

Fig. 1. Cortisol Sampling Protocol. 173 studies contributed 218 sub-samples of at least 2 or more cortisol samples. Cortisol collection times were standardized to the start of the TSST (time point 0).

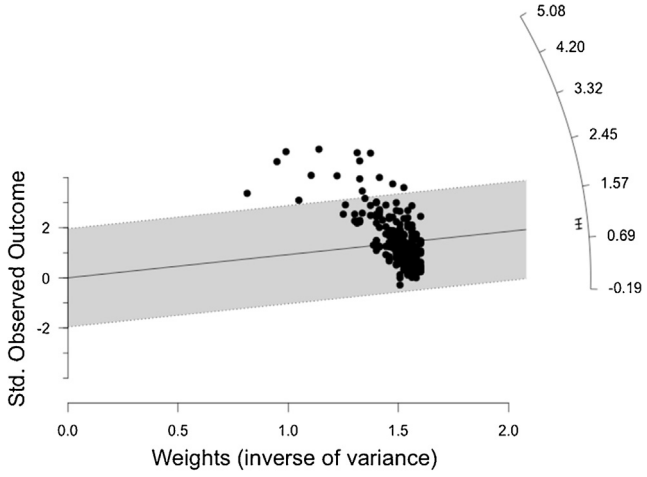

Fig. 2. Galbraith plot of overall effect. Radial plot demonstrating the overall effect size estimate (d' = .925) and 95% confidence interval [.84,1.01]. Individual effect sizes for each sub-sample (represented by dots) included can be found by drawing a line from the origin through the sub-sample to the radial axis.

3. Results

3.1. Cortisol sampling protocol

Sampling protocol variations were extracted to assess a prototypical cortisol stress response. From these examinations, cortisol was observed to peak 38 min after the start of a psychosocial stressor (Fig. 1). On average (weighted by sample size) studies sampled cortisol 4.1 (SD = 1.77) times after the cessation of the stressor. However there was broad variability in the number of post TSST samples ranging from 1 to 11, with the most commonly sampled times immediately after the TSST, 10–15 min, 20–25 min and 30–35 min. No sharp peaks in cortisol were observed during any time point, with highest cortisol levels occurring broadly between 35 and 45 min after the start of the TSST.

3.2. Study and TSST protocol

The average effect size across all assessments (Fig. 2) from all studies was *d'* = .925 (SEM = .043), which is significantly different from zero, $t(236) = 21.41$, $p < .001$; 95% CI = 0.84, 1.01.Total heterogeneity was .38, and was significant $X^2(236) = 1704.12$, $p < .001$. A

Table 1
Study characteristics.

	Mean (SD)	Range	Sub-samples reporting
Year published	2010 (5)	1993–2016	n = 237
Participants/study	36.1 (29.67)	6–219	n = 237
Gender	80 male 49 female 109 mixed		n = 237
Age (years)	28.46 (10.73)	18.38–65.78	n = 140
Age SD (years)	4.50 (3.58)	0.2–23.97	n = 132
Time of day			
Time in lab before TSST	44 min (41 min)	5 min–4 h	n = 212
Wait time after intravenous catheter placement	48.89 min (41.84 min)	20 min–4 h	n = 27
TSST Jury Panel			
Number of jury members	<2 jury: n = 8 2 jury: n = 154 3 jury: n = 33 4 jury: n = 2		n = 197
Reported estimated jury panel age (years)	26.05 (5.46)	20–50	n = 93
TSST			
Duration of preparation period	6.97 min (3.17 min)	2–20 min	n = 229
Duration of verbal period	<5 min: n = 7 5 min: n = 215 >5 min: n = 12		n = 234
Duration of math period	<5 min: n = 9 5 min: n = 212 >5 min: n = 9		n = 230

few sub-samples reported large effects (n = 9; $d > 3.0$), however, reanalysis of TSST protocol variations excluding those studies did not alter statistical results or conclusions, and subsequently were retained for analysis.

For study protocol factors, sufficient variations between studies existed to examine whether time of day (morning n = 17, lunchtime n = 30, afternoon n = 42, mixed times n = 129), pre-TSST lab time (0–15 min n = 49, 16–30 min n = 67, 31–60 min n = 65, >60 min n = 33), and pre-TSST activities (IV and questionnaires n = 24, questionnaires n = 109, and IV only n = 29) influenced cortisol responses. TSST protocol variations were broken into two analyses groups dependent on how much methodological factor variation between studies was present. Main analyses consisted of, speech preparation times (3 min n = 38, 5 min n = 77, 10 min n = 101), questionnaires administered during speech preparation times (yes n = 42, no n = 62), speech composition (job talk n = 184, other n = 49), and size of jury (two jurors n = 153, three jurors n = 32). The remaining methodological factors were examined in an exploratory nature due to reduced numbers of studies reporting these factors: speech task only TSST (n = 5), arithmetic subtract number (steps of 13 n = 66; steps of 17 n = 75), jury gender (all female n = 11, mixed gender n = 146), gender of jury member providing instructions (cross genders, e.g., male participants and female judge n = 15, unknown gender of jury member and participant n = 59), and juror feedback (negative n = 24, neutral n = 107).

3.2.1. Study protocol variations

Evidence suggests that time of day when the TSST was conducted does not significantly influence the overall cortisol response. Typically, lunchtime (12p.m.–2p.m.) is avoided due to potentially confounding effects from ingesting food. The average effect size for those studies where the TSST was conducted during lunch time (d' = .811, SE = .118, n = 30) did not differ significantly from other parts of the day (see Table 2). A general trend was evident with those studies conducted in the afternoon having a larger overall effect (d' = .962, SE = .090, n = 42). A number of studies did not report the time of day they conducted the TSST (n = 19) or conducted the TSST at different time during the day (generally a mix of lunchtime and afternoon timeframes; N = 129) with indications that conducting the TSST in the afternoon yielded larger cortisol responses.

Participants who spent significant amounts of time (>60 min) in the laboratory environment pre-TSST exhibited lower overall average effect sizes (d' = .799, SE = .128) than those who spent shorter amounts of time. More specifically, participants in the lab for 16–30 min displayed the highest meta-analytic effect size, and this was followed by a decreasing general trend in point estimates as participants spent longer amounts of time (0–15 min: d' = .928, SE = .091: 16–30 min: d' = .966, SE = .090: 31–60 min: d' = .889, SE = .071).

Pre-TSST activities included participants completing questionnaires and/or having an IV placed. Those studies (n = 29) that had an IV placed for blood sampling had the largest average overall effects (d' = .135, SE = .168). Meta-analytic point estimates for those studies where participants completed questionnaires only (d' = .823, SE = .054, n = 109) or in combination with IV placement (d' = .965, SE = .104, n = 24) were lower and did not fall within the confidence intervals for the IV only average effect suggesting lower cortisol responses.

See Fig. 3 for a summary of effects size differences.

3.2.2. TSST protocol variations

Studies varied in how long they had participants prepare for the speech portion of the TSST ranging from 3 min to 10 min (see Table 3). Few differences between 3 min, 5 min and 10 min were noted in either overall average effect sizes (3 min: d' = .891, 5 min: d' = .904, 10 min: d' = .947) or standard error estimates (3 min: SE = .083, 5 min: SE = .062, 10 min: SE = .082), as indicated by significant overlap of confidence intervals and meta-analytic point estimates. Similarly, the administration of questionnaires during this preparation time did not have a significant influence on overall average cortisol response (yes: d' = .846, SE = .106: no: d' = .849, SE = .051). Assessing differences due to arithmetic subtraction number chosen revealed that when 13 (d' = 1.012, SE = .011, n = 66) was used, overall average effects were higher than when 17 was used (d' = .907, SE = .051, n = 75) and fell outside the confidence interval constructed for 17's meta-analytic effect size.

Table 2
Summary of effect size estimates, standard errors (SE), and confidence intervals (CI; lower limit: LL, upper limit: UL) across study protocol variations.

Time of TSST	n	Estimate	SE	95% CI LL	95% CI UL
Morning	17	.784	.152	.486	1.083
Lunch	30	.811	.118	.581	1.042
Afternoon	42	.962	.090	.785	1.139
Mixed	129	.986	.065	.860	1.112
Unknown	19	1.103	.208	.694	1.511
Acclimation Duration					
0–15 min	49	.928	.091	.749	1.106
16–30 min	67	.966	.090	.789	1.143
31–60 min	65	.889	.075	.750	1.028
>60 min	33	.799	.128	.548	1.051
Pre-TSST Activity					
Intravenous catheter	29	1.345	.168	1.015	1.675
Questionnaire	109	.823	.054	.717	.929
Intravenous catheter + questionnaire	24	.965	.104	.762	1.168
No activity	49	1.031	.113	.810	1.252

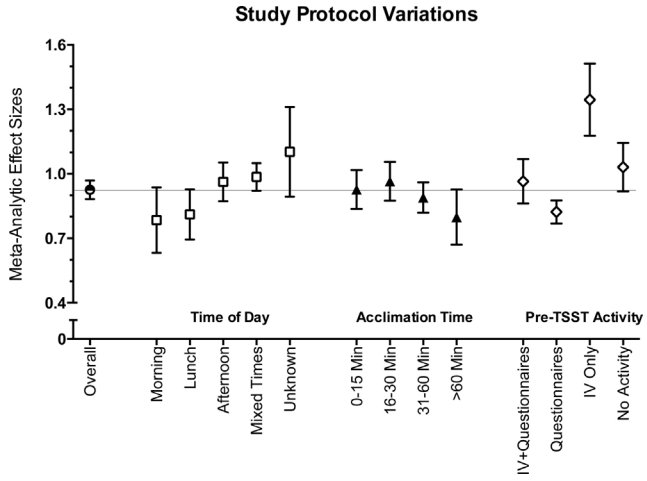

Fig. 3. Meta-analytic effect size estimates with 95% confidence intervals for study protocol variation (IV: intravenous catheter placement).

Table 3
Summary of effect size estimates, standard errors (SE), and confidence intervals (CI; lower limit: LL, upper limit: UL) across TSST Protocol Variations.

Prep Time	n	Estimate	SE	95% CI LL	95% CI UL
3 min	38	.891	.083	.728	1.055
5 min	77	.904	.061	.783	1.025
10 min	101	.947	.082	.787	1.107
Preparation Period Questionnaires					
Yes	42	.846	.078	.693	1.000
No	62	.849	.075	.703	.996
Serial subtraction					
Steps of 17	75	.907	.051	.807	1.006
Steps of 13	66	1.012	.106	.804	1.219
Jury #					
Two jurors	153	.891	.048	.796	.986
Three jurors	32	1.064	.148	.773	1.354
Jury Gender					
Female	11	.547	.105	.341	.753
Mixed	146	.975	.054	.870	1.080
Instructions					
Cross	15	.853	.100	.657	1.049
Uncontrolled	59	.796	.062	.675	.917
Missing/unknown	144	1.088	.071	.949	1.228
Feedback					
Negative	24	.713	.090	.536	.889
Neutral	107	.869	.055	.762	.975

The vast majority of studies followed the standard verbal TSST paradigm. Exploratory analysis revealed studies using only the speech portion of the TSST had similar overall average effects ($n = 5$, $d' = 1.066$) to other studies using job interviews and mental arithmetic, with a large 95% confidence interval constructed around the meta-analytic effect [.657, 1.473].

Comparing meta-analytic point estimates for the overall effect suggests that having more judges on the judging panel resulted in an increased cortisol response ($d' = 1.064$, SE = .148, n = 32). However, the overall average effect size estimates for those judging panel with two members ($d' = .891$, SE = .048) fell within the lower bounds of the three juror confidence intervals and the latter was associated with a large increase in confidence interval size [.773, 1.354] due to fewer studies.

Exploratory analyses of jury gender compositions suggested all female juries ($d' = .547$, SE = .105) to be less effective in terms of eliciting cortisol responses than mixed gender juries ($d' = .975$, SE = .054) with no overlap of confidence intervals. Preliminary analysis of whether gender of the juror providing the participant instructions (cross gender: $d' = .853$, uncontrolled gender: $d' = .796$) influenced overall average effects was inconclusive with large confidence intervals (cross gender: [.657, 1.05]; uncontrolled gender: [.675, .916]) around all meta-analytic effect sizes. A limited number of studies (n = 24) using negative feedback from the judging panel reported a lower overall effect ($d' = .713$, SE = .090) that fell outside the confidence interval range of the more frequently used neutral feedback ($d' = .869$, SE = .055).

See Fig. 4 for a summary of effects size differences.

4. Discussion

The current review aimed to assess determinants of the strength of cortisol responses to the most widely used acute psychosocial laboratory stress test, the TSST. Specifically, we focused on three groups of methodological variations: (1) sampling protocol variations, (2) general study protocol variations exerting indirect effects on cortisol response strength by affecting hormone dynamics, and (3) variations in the implementation of the TSST protocol itself. The results for each of the groups of protocol variations will be discussed in detail below.

4.1. Systematic review: cortisol sampling protocol

Examining the sampling protocols of 173 studies revealed not only great differences across studies in how many samples are collected post-TSST, but also a large variety in sampling time-points. The majority of studies covered the first half hour post-TSST with

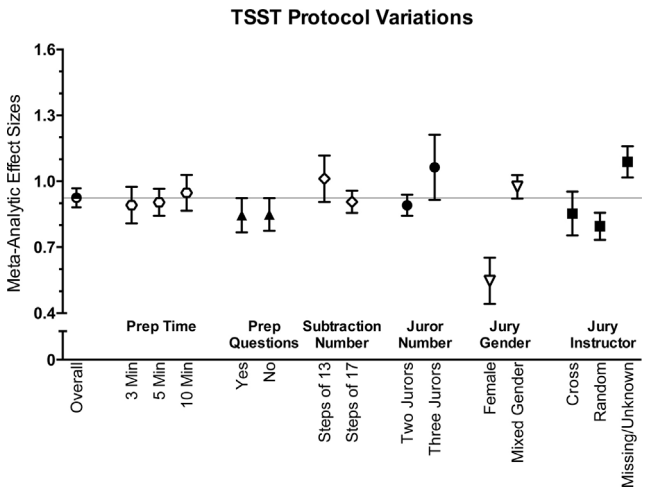

Fig. 4. Meta-analytic effect size estimates with 95% confidence intervals for TSST protocol variation.

multiple samples, allowing for high-resolution assessment of data points and extraction of an aggregated/collapsed cortisol response pattern. Most noteworthy, these analyses revealed no pronounced peak, but rather a gradual rise followed by a gradual fall in cortisol values and thus a prolonged period of increased cortisol availability. This observation is in line with the main characteristics of endocrine communication and as such reflects the underlying physiological processes. Furthermore, if circumstances may allow for only a very limited number of sample collection time-points, our findings support the recommendation to sample around 30–45 min after TSST *onset* to maximize the chance to capture the peak cortisol response.

4.2. Meta-analysis: study and TSST protocol variations

To examine study and TSST protocol variation effects on cortisol stress responses, we were able to extract sufficient information from a total of 186 TSST studies. From these studies, 237 effects sizes were computed and assessed meta-analytically. Overall, we observed a significant effect of the TSST on cortisol stress responses, such that exposing healthy adults to this specific psychosocial stress test led on average to an increase in cortisol levels over baseline levels in the magnitude of one standard deviation (.925). This estimate is in line with a sub-analysis presented in an earlier review that focused on stress tasks characterized by social-evaluative threat and uncontrollability, which included several TSST studies (Dickerson and Kemeny, 2004). As such, our finding supports the TSST as a valid laboratory stress protocol that reliably induces psychosocial stress and activates the HPA axis. However, the significant heterogeneity also clearly indicated that further exploration of factors explaining differences in cortisol stress responses between studies is warranted. This is particularly relevant in contexts in which it is important to be able to distinguish between response patterns that reflect individual differences in health risks from those that result from lab-dependent differences in TSST protocols.

4.2.1. Meta-analysis: physiologically relevant study protocol variations

In an effort to explain variability in cortisol responses to the TSST, the present meta-analysis focused first on factors that have the potential to affect cortisol stress response strength by influencing pre-TSST or baseline cortisol levels. I.e., we aimed at identifying factors that may systematically influence cortisol stress responses independent of the actual stress exposure experience.

a) Time of day

Cortisol levels show a pronounced circadian pattern characterized by a sharp increase in response to awakening followed by slow decreases over the course of the day and a nadir late at night/during the first half of sleep (Clow et al., 2004). Hence, conducting the TSST in the afternoon is often regarded as the best option to avoid large inter-individual variation in baseline cortisol levels. Furthermore, as meals are followed by cortisol increases (Kirschbaum et al., 1993) and pre-meal hypoglycemia is linked to a lack of cortisol stress responses (Davis et al., 1997), both lunch and dinner times are usually avoided as well. When comparing studies with TSST start-times scheduled during any of these time periods, we did not observe very pronounced differences in effects sizes. Closer inspection, however, revealed that cortisol stress response strengths in studies with TSSTs taking place in the morning/A.M. or during lunch time (12–2p.m.) are slightly lower and more variable than studies with TSSTs scheduled between 2p.m. and 5p.m. It can be speculated that with more studies in these categories, a clearer separation of confidence intervals would have supported the idea that TSSTs taking place during the A.M. and noon times are less reliable in inducing strong cortisol stress responses. Also, a substantial number of studies reported mixed TSST start-times, although taking into account the more detailed information provided by authors responding to the Qualtrics request revealed that mixed times mostly cover the afternoon hours, but may include 1p.m. or post-5p.m. Not surprisingly, the effect sizes of those studies are similar to the afternoon category and due to the larger number of studies falling into the mixed category, even present with smaller standard deviations. Taken together, our findings suggest that limiting TSST start-times to the afternoon may increase the likelihood of strong cortisol stress responses, while other start-times still present feasible alternatives.

b) Acclimation period duration

To avoid elevated pre-TSST baseline cortisol levels, most study protocols include a resting or acclimation period. The expectations is that participants may come to the laboratory already stressed due to events unrelated to the study or participation in a research study may itself be perceived as stressful (Lazarus, 1993). In either case, elevated baseline cortisol levels may mask TSST-induced increases in cortisol, resulting in supposedly blunted responses. Hence, allowing the participants to recover or acclimate is frequently regarded as a necessary measure to facilitate pre-study cortisol levels to return to baseline (Dickerson and Kemeny, 2004), and thereby to permit strong TSST responses and low inter-individual variability. Surprisingly, effect size estimates for studies with acclimation periods up to 15 min, 30 min, 60 min, or over 60 min did not substantially differ from each other. If at all, acclimation periods over 60 min were linked to reduced cortisol stress response strength. As such, our analyses do not support the notion that longer pre-TSST periods are necessarily better in terms of ensuring strong cortisol stress responses. However, one alternative explanation may be that specifically in studies with shorter acclimation periods, it was more likely that participants with high baseline cortisol levels were excluded from further data analyses by the individual study authors, thus cautioning against over-interpretation. Furthermore, it is noteworthy that the heterogeneity in studies within each category was among the highest seen across all analyses. This suggests that it may not so much be the sheer passage of time that is relevant, but that other factors occurring during this period may be more relevant or that acclimation durations may co-vary with multiple other factors. In this regard, it is important to consider the specific activities taking place during the acclimation period.

c) Pre-TSST activities

In many studies, participants are asked to answer questionnaires before the TSST and/or receive an intravenous catheter for sample collection. Both events can be upsetting and the latter potentially painful. We thus expected that compared to a true resting period, filling out questionnaires would result in lower cortisol stress responses and that this effect would be more pronounced for receiving a venous catheter, given that the latter is typically not removed until the end of the study protocol.

As expected, answering questionnaires pre-TSST resulted in a slightly lower effect size estimate compared to the overall estimate. However, contrary to our predictions, studies using intravenous catheters showed the highest cortisol stress responses out of all protocol variations assessed. While counterintuitive at first glance, one interpretation may be that participants' stress response systems may have been sensitized by the physical threat of having a catheter inserted into one's arm, while filling out questionnaires may be upsetting or in other ways interfere with acclimation aiming at reducing baseline pre-TSST cortisol levels. While highly speculative, follow-up analyses on baseline levels revealed no differences between the categories, thus supporting the theory of potential sensitization processes.

As mentioned above, pre-TSST activities often are confounded with acclimation period duration (Engert et al., 2013), such that studies using venous catheters allow for longer recovery post catheter placement and hence reduce the risk of elevated baseline cortisol levels. Future reviews should thus assess the combination of duration and type of activity to identify the most favorable protocol depending on study requirements.

4.2.2. Meta-analysis: psychologically relevant TSST protocol variations

While the protocol variations discussed above presented considerations relevant in terms of cortisol stress response confounds, changes to the TSST protocol may vary the level of psychological threat perceived by the participant and thus increase or diminish the chance to induce a robust and reliable cortisol stress response. We will first discuss the effects of variation in TSST task components, such as duration or combination of tasks, followed by an examination of the variations that affect the social interaction between the panel members and the participant.

a) TSST task components

The original TSST protocol included 10 min for the participant to prepare for the speech part of the stress test (Kirschbaum et al., 1993). Over the years, more and more studies reduced this period to 5 min and sometimes even to 3 min. As the speech preparation period was meant to contribute to the stressfulness of the experience by eliciting anticipatory stress (Engert et al., 2013), we predicted that shorter preparation times would result in smaller effect sizes. However, meta-analytical comparison did not reveal any differences between the three, suggesting that reducing the preparation period even down to 3 min will not negatively affect cortisol stress response strength. This is particularly noteworthy given the ethical obligation to minimize the burden for participants. It would be interesting in future studies to assess whether a preparation period is necessary at all, unfortunately however, we did not find any studies that excluded this component and could speak to this question.

Together with a reduction in preparation time, the incidence of studies asking participants to answer state questionnaires during that time period is increasing. In fact, these two variations may often influence each other. Despite this, having to answer a state questionnaire did not result in an effect size estimate that was any different from those studies in which no questionnaire was given. This finding opens up interesting possibilities, as for many research questions, it could be advantageous to assess participants' state not only before and after the stressful situation, but while in the situation itself as well.

Contrary to frequent variations in speech preparation time, too few studies varied the length of the verbal or arithmetic component or the composition of the two to allow for meaningful comparison of effect size estimates. The same was true for content of the speech part, i.e., the majority of studies followed the original instructions and yielded robust results suggesting that theoretical job interviews are an appropriate and effective stressor. Assessing the effects of variations in arithmetic task difficulty, the majority of the studies used either the original task of serial subtraction by 13 (Kirschbaum et al., 1993) or the later version asking participants to subtract in steps of 17 (Strahler et al., 2010). Interestingly, the former was associated with higher cortisol stress responses compared to the latter, although the overlap in confidence intervals makes it difficult to conclude that steps of 17 are truly less stressful. Nevertheless, it is noteworthy that the chosen approach to define cortisol stress responses was able to pick up this rather subtle and inter-individually varying difference in difficulty level.

b) Social interaction components

Several variations were introduced over the course of the last two decades that concern the social interaction component or TSST panel. This includes number of panel members, gender composition of the panel, gender match between the active panel member and participant, and last but not least, the panel members' behavior toward the participant.

Comparing cortisol stress response strengths between TSST studies that used a 3-member panel with those that used a 2-member paper suggest that the former may be slightly more effective. However, a significantly smaller number of studies followed the original protocol using a 3-member panel and hence overlapping confidence interval ranges make a clear recommendation difficult.

Despite occurring in only 11 studies, an all female panel resulted in one of the lowest effect size estimates across all variations assessed. It would be interesting to compare these estimates with studies using an all male TSST panel, however, this variation did not occur in any of the included studies. Similarly, insufficient information was available to investigate this effect further by considering *participant* gender.[1] Future studies should address this question, as it could be speculated that an all female panel would present less social identity threat and discrimination particularly for female participants, which would then lead to the weaker cortisol stress responses.

Directly related to this topic, we aimed to assess the relevance of receiving the instructions and feedback during the TSST by a panel member of the same versus the opposite gender for cortisol stress response strengths. Again, the vast majority of studies did not provide this information. Perhaps unsurprisingly, the magnitude of this group's cortisol stress response strength is thus very similar to the overall studies' effect size. On the contrary, studies that explicitly did not control the gender match between active panel member and participant showed lower cortisol stress responses. However, although these studies fell below the lower bounds of the not-specified studies' confidence intervals, the overlapping confidence intervals with the effect size estimate for the 15 studies explicitly

[1] Exploratory analyses revealed no cortisol stress response effect size differences ($F(2,8) = 0.46$, $p = .65$) between n = 5 samples assessing male participants only, n = 1 sample assessing female participants only, and n = 5 samples assessing both genders.

stating to have used only cross-gender interactions caution against delineating a strong recommendation.

Lastly, we assessed panel members' behavior toward the participant, specifically, whether feedback was given in a neutral manner or negatively. A total of 24 studies used negative feedback during the presentation portion of the TSST, which resulted in a clearly lower effect size estimate and thus weaker cortisol stress responses compared to the standard/neutral behavior. It could be argued that contrary to receiving non-valenced feedback, people have developed strategies to deal with negative feedback (Swann et al., 1992), which from an appraisal point-of-view, can be thought of as resources (Lazarus, 1993). Similarly, applying social self-preservation theory (Dickerson et al., 2004), negative feedback would give participants the opportunity to react and potentially defend him/herself. As such, both theoretical frameworks would predict decreases in perceived stressfulness of the situation and in line with our findings, diminished cortisol stress responses.

4.3. Limitations and outlook

Any recommendations based on the above findings have to be considered in light of several limitations. First, we used maximal cortisol increases to compute study effect sizes. However, the overall duration of the TSST varied very minimally across all assessed studies, thus reducing potential issues that may have resulted from dependent shifts in cortisol peak values. Furthermore, we were able to pick up subtle differences, such as in arithmetic task difficulty, further supporting our approach. Nevertheless, while we were particularly interested in the response component rather than the overall pattern or subsequent recovery, future studies targeting different questions may focus on area-under-the curve or slope of post-peak decline instead. Secondly, we did not assess potential research group effects. This decision was based on the fact that the largest research group was identical with the authors of the original protocol and the goal of the meta-analysis was to examine effects of deviations from the original protocol, i.e., assessing research group effects presented a circularity problem. Thirdly, significant study heterogeneity remained after examining each methodological variation. This suggests that an important next step will be to examine combinations of factors. This will then also allow for expanding current recommendations to suggestions for the most reliable and efficient TSST protocol in terms of cortisol increase strength in a given lab setting. Lastly, exposure to the TSST results in changes in multiple systems. The current review and meta-analysis focused on cortisol stress responses, as cortisol is one of the central effector of many of those changes (e.g., cortisol stress response modulation of inflammatory response, Sapolsky et al., 2000). However, as more studies become available that assess multiple systems simultaneously, we will be able to also investigate the effects of protocol variations on the stress response interplay between systems.

4.4. Summary

In summary, the current review revealed a cortisol TSST response pattern characterized by peak values occurring around 38 min post TSST start or 18 min after the end of the originally specified 20 min TSST. However, when aggregating cortisol values across 186 studies, the peak appeared to be less pronounced and instead consistent with endocrine dynamics resulting in gradual increases and recoveries. As such, in an effort to maximize effects and minimize costs, sampling protocols may not necessarily have to include short collection intervals immediately following the TSST, but instead focus on more frequent sample collections during the identified cortisol peak period.

With regard to study protocol variations that may confound baseline cortisol values and influence subsequent reactivity measures, meta-analytical examination revealed that acclimation periods pre-TSST below 30 or perhaps even 15 min may suffice, at least as long as no interfering activities, i.e., questionnaires, are taking place during that timeframe. It can be speculated that in contrast to leading to elevated baseline levels due to being stressful, an intravenous catheter may result in sensitization. However, this must be confirmed in future analyses assessing interactions between duration and activity, as intravenous catheters are often associated with longer acclimation periods.

Lastly, assessing the effects of TSST protocol variations on cortisol response strength, several observations are noteworthy. First, shortening speech preparation time did not change cortisol responses in any way, nor did including questionnaires during that period show an effect. As such, our findings suggest that speech preparation time is one TSST element that can be used to reduce the time burden for participants as well as laboratory logistics. Secondly, having panel members be all females and instructing panels members to show negative instead of neutral behavior towards the participants both were associated with considerably reduced cortisol stress response strengths. These observations emphasize the relevance of adhering to the original TSST protocol instructions of using a mixed gender panel and neutral panel feedback. Thirdly, while showing slight differences in effect size estimates, several variables of interest, such as content of the speech task, number of panel members, or gender match between active panel member and participant, were problematic to evaluate due to the small number of studies reporting those details. This cautions against brief and concise method sections if they come at the expense of sharing relevant protocol specifications.

Taken together, we presented a combination of a systematic review of sampling frequency and timing effects on cortisol stress response assessment and a meta-analytical examination of confounding study protocol variations as well as TSST protocol deviations effects on cortisol stress response strength. This approach was successful in pinpointing a time period in which to expect maximum cortisol levels. Furthermore, it revealed protocol components that help ensure reliable cortisol stress responses and identified elements that may be varied to increase efficiency of protocol implementation.

4.5. Recommendations

More specifically, our findings support the following recommendations:

- To capture peak cortisol stress responses on average occurring 38 min after TSST *onset*, frequent sampling between 30 and 45 min after TSST onset is recommended.
- Overall, the effectiveness of the TSST in terms of inducing cortisol responses is fairly robust to methodological variations. Particularly shortening the acclimation and the TSST preparation time times would aid in logistical concerns.
- IV placement, however, may potentially sensitize participants to future stressors and enhance cortisol responses to the TSST.
- Frequent lack of reporting relevant protocol details cautioned against interpretation of effect size differences between a number of protocol variations. Future studies should thus include more detailed descriptions of study protocols and particularly TSST implementation.
- Until stronger evidence suggests otherwise, adhering to the original TSST protocol specifications appears most promising in terms of maximizing cortisol stress responses (i.e., TSST conducted in the afternoon, with a three member-mixed gender jury panel giving neutral feedback, and utilizing 13 for subtraction tasks).

Acknowledgements

This work was supported by a Brain and Behavior Research Foundation NARSAD Young Investigator Award (J.M.W.).

The authors of this study wish to gratefully acknowledge the extensive support they received from the PNI/PNE community as well as thank all principle investigators who provided additional information via Qualtrics. The authors also wish to acknowledge all undergraduate research assistants who contributed to data coding and data entry.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.psyneuen.2017.02.030.

References

Al'Absi, M., Bongard, S., Buchanan, T., Pincomb, G.A., Licinio, J., Lovallo, W.R., 1997. Cardiovascular and neuroendocrine adjustment to public speaking and mental arithmetic stressors. Psychophysiology 34 (3), 266–275.

Allen, A.P., Kennedy, P.J., Cryan, J.F., Dinan, T.G., Clarke, G., 2014. Biological and psychological markers of stress in humans: focus on the Trier Social Stress Test. Neurosci. Biobehav. Rev. 38, 94–124.

Bellingrath, S., Kudielka, B.M., 2008. Effort-reward-imbalance and overcommitment are associated with hypothalamus–pituitary–adrenal (HPA) axis responses to acute psychosocial stress in healthy working schoolteachers. Psychoneuroendocrinology 33 (10), 1335–1343.

Buchanan, T.W., Laures-Gore, J.S., Duff, M.C., 2014. Acute stress reduces speech fluency. Biol. Psychol. 97, 60–66, http://dx.doi.org/10.1016/j.biopsycho.2014.02.005.

Campbell, J., Ehlert, U., 2012. Acute psychosocial stress: does the emotional stress response correspond with physiological responses? Psychoneuroendocrinology 37 (8), 1111–1134, http://dx.doi.org/10.1016/j.psyneuen.2011.12.010.

Clow, A., Thorn, L., Evans, P., Hucklebridge, F., 2004. The awakening cortisol response: methodological issues and significance. Stress 7 (1), 29–37.

Cornelisse, S., van Stegeren, A.H., Joëls, M., 2011. Implications of psychosocial stress on memory formation in a typical male versus female student sample. Psychoneuroendocrinology 36 (4), 569–578, http://dx.doi.org/10.1016/j.psyneuen.2010.09.002.

Cumming, G., 2013. Understanding the New Statistics: Effect Sizes, Confidence Intervals and Meta-analysis. Routledge.

Davis, S., Shavers, C., Davis, B., Costa, F., 1997. Prevention of an increase in plasma cortisol during hypoglycemia preserves subsequent counterregulatory responses. J. Clin. Invest. 100 (2), 429.

Dickerson, S.S., Kemeny, M.E., 2004. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. Psychol. Bull. 130 (3), 355.

Dickerson, S.S., Gruenewald, T.L., Kemeny, M.E., 2004. When the social self is threatened: shame, physiology, and health. J. Pers. 72 (6), 1191–1216.

Dickerson, S.S., Gable, S.L., Irwin, M.R., Aziz, N., Kemeny, M.E., 2009. Social-evaluative threat and proinflammatory cytokine regulation an experimental laboratory investigation. Psychol. Sci. 20 (10), 1237–1244.

Dykman, B.M., Abramson, L.Y., Alloy, L.B., Hartlage, S., 1989. Processing of ambiguous and unambiguous feedback by depressed and nondepressed college students: schematic biases and their implications for depressive realism. J. Pers. Soc. Psychol. 56 (3), 431.

Engert, V., Efanov, S.I., Duchesne, A., Vogel, S., Corbo, V., Pruessner, J.C., 2013. Differentiating anticipatory from reactive cortisol responses to psychosocial stress. Psychoneuroendocrinology 38 (8), 1328–1337.

Foley, P., Kirschbaum, C., 2010. Human hypothalamus–pituitary–adrenal axis responses to acute psychosocial stress in laboratory settings. Neurosci. Biobehav. Rev. 35 (1), 91–96.

Gerber, M., Kalak, N., Elliot, C., Holsboer-Trachsler, E., Pühse, U., Brand, S., 2013. Both hair cortisol levels and perceived stress predict increased symptoms of depression: an exploratory study in young adults. Neuropsychobiology 68 (2), 100–109, http://dx.doi.org/10.1159/000351735.

Het, S., Rohleder, N., Schoofs, D., Kirschbaum, C., Wolf, O., 2009. Neuroendocrine and psychometric evaluation of a placebo version of the 'Trier Social Stress Test'. Psychoneuroendocrinology 34 (7), 1075–1086.

Joëls, M., Pu, Z., Wiegert, O., Oitzl, M.S., Krugers, H.J., 2006. Learning under stress: how does it work? Trends Cogn. Sci. 10 (4), 152–158.

Johnson, E.O., Kamilaris, T.C., Chrousos, G.P., Gold, P.W., 1992. Mechanisms of stress: a dynamic overview of hormonal and behavioral homeostasis. Neurosci. Biobehav. Rev. 16 (2), 115–130.

Kemeny, M.E., 2003. The psychobiology of stress. Curr. Dir. Psychol. Sci. 12 (4), 124–129.

Kirschbaum, C., Pirke, K.-M., Hellhammer, D.H., 1993. The 'Trier Social Stress Test': a tool for investigating psychobiological stress responses in a laboratory setting. Neuropsychobiology 28 (1–2), 76–81, http://dx.doi.org/10.1159/000119004.

Kudielka, B.M., Hellhammer, D.H., Kirschbaum, C., Harmon-Jones, E., Winkielman, P., 2007. Ten years of research with the Trier Social Stress Test—revisited. Social Neurosci.: Integr. Biol. Psychol. Explan. Social Behav., 56–83.

Lazarus, R.S., 1993. Coping theory and research: past, present, and future. Psychosom. Med. 55 (3), 234–247.

Luethi, M., Meier, B., Sandi, C., 2009. Stress effects on working memory, explicit memory, and implicit memory for neutral and emotional stimuli in healthy men. Front. Behav. Neurosci. 2.

Mason, J.W., 1975. A historical view of the stress field. J. Human Stress 1 (2), 22–36.

Miller, G.E., Rohleder, N., Cole, S.W., 2009. Chronic interpersonal stress predicts activation of pro- and anti-inflammatory signaling pathways 6 months later. Psychosom. Med. 71 (1), 57–62, http://dx.doi.org/10.1097/PSY.0b013e318190d7de.

Sapolsky, R.M., Romero, L.M., Munck, A.U., 2000. How do glucocorticoids influence stress responses? Integrating permissive, suppressive, stimulatory, and preparative actions 1. Endocr. Rev. 21 (1), 55–89.

Strahler, J., Mueller, A., Rosenloecher, F., Kirschbaum, C., Rohleder, N., 2010. Salivary-amylase stress reactivity across different age groups. Psychophysiology 47 (3), 587–595, http://dx.doi.org/10.1111/j. 1469-8986.2009.00957.x.

Stroud, L.R., Foster, E., Papandonatos, G.D., Handwerger, K., Granger, D.A., Kivlighan, K.T., Niaura, R., 2009. Stress response and the adolescent transition: performance versus peer rejection stressors. Dev. Psychopathol. 21 (01), 47–68.

Swann, W.B., Wenzlaff, R.M., Krull, D.S., Pelham, B.W., 1992. Allure of negative feedback: self-verification strivings among depressed persons. J. Abnorm. Psychol. 101 (2), 293.

Thompson, S.G., Higgins, J.P.T., 2002. How should meta-regression analyses be undertaken and interpreted? Stat. Med. 21 (11), 1559–1573, http://dx.doi.org/10.1002/sim.1187.

Townsend, S.S., Major, B., Gangi, C.E., Mendes, W.B., 2011. From in the air to under the skin: cortisol responses to social identity threat. Pers. Soc. Psychol. Bull. 37 (2), 151–164.

Wiemers, U.S., Wolf, O.T., 2015. Cortisol broadens memory of a non-stressful social interaction. Psychopharmacology (Berl.) 232 (10), 1727–1733, http://dx.doi.org/10.1007/s00213-014-3808-7.

de Kloet, E.R., Oitzl, M.S., Joëls, M., 1999. Stress and cognition: are corticosteroids good or bad guys? Trends Neurosci. 22 (10), 422–426.

von Känel, R., Kudielka, B.M., Preckel, D., Hanebuth, D., Fischer, J.E., 2006. Delayed response and lack of habituation in plasma interleukin-6 to acute mental stress in men. Brain Behav. Immun. 20 (1), 40–48, http://dx.doi.org/10.1016/j.bbi.2005.03.013.