

# A systematic review of the Trier Social Stress Test methodology: Issues in promoting study comparison and replicable research

N.F. Narvaez Linares<sup>a,1</sup>, V. Charron<sup>a,1</sup>, A.J. Ouimet<sup>b</sup>, P.R. Labelle<sup>c</sup>, H. Plamondon<sup>a,\*</sup>

<sup>a</sup> Behavioural Neuroscience Group, School of Psychology, University of Ottawa, 136 Jean-Jacques Lussier, Ottawa, On, K1N 6N5, Canada

<sup>b</sup> Cognition & Anxiety Disorders Research Laboratory, School of Psychology, University of Ottawa, 136 Jean-Jacques Lussier, Ottawa, On, K1N 6N5, Canada

<sup>c</sup> University of Ottawa Library, 120 University Private, Ottawa, ON, K1N 6N5, Canada

## ARTICLE INFO

### Keywords:

Trier social stress test  
Stress paradigm  
Protocol  
Systematic review  
Standardization

## ABSTRACT

Since its development in 1993, the Trier Social Stress Test (TSST) has been used widely as a psychosocial stress paradigm to activate the sympathetic nervous system and hypothalamic-pituitary-adrenal axis (HPAA) stress systems, stimulating physiological functions (e.g. heart rate) and cortisol secretion. Several methodological variations introduced over the years have led the scientific community to question replication between studies. In this systematic review, we used the Preferred Reporting Items of Systematic Reviews and Meta-Analysis (PRISMA) to synthesize procedure-related data available about the TSST protocol to highlight commonalities and differences across studies. We noted significant discrepancies across studies in how researchers applied the TSST protocol. In particular, we highlight variations in testing procedures (e.g., number of judges, initial number in the arithmetic task, time of the collected saliva samples for cortisol) and discuss possible misinterpretation in comparing findings from studies failing to control for variables or using a modified version from the original protocol. Further, we recommend that researchers use a standardized background questionnaire when using the TSST to identify factors that may influence physiological measurements in tandem with a summary of this review as a protocol guide. More systematic implementation and detailed reporting of TSST methodology will promote study replication, optimize comparison of findings, and foster an informed understanding of factors affecting responses to social stressors in healthy people and those with pathological conditions.

## 1. Introduction

Prior to the development of the Trier Social Stress Test (TSST; Kirschbaum et al., 1993), researchers used diverse psychological stressors to assess the impact of stress on Hypothalamic-Pituitary-Adrenal axis (HPAA) activation (Berger et al., 1987). Given major concerns that these different stressors were associated with important inter-individual variability in physiological responses and that effects were often too small to be measured reliably, Kirschbaum et al. (1993) developed the TSST to provide researchers with a standardized psychophysiological paradigm for assessing the impact of psychosocial stress. Since then, researchers worldwide have used the TSST as a robust experimental protocol proven to produce reliable physiological outcomes (Dickerson and Kemeny, 2004). Since

its publication and up to 2007, the TSST was used in over 4000 studies, supporting its well-established status as a stress paradigm (Kudielka et al., 2007). TSST tasks (speech and arithmetic) are remarkably efficacious in studying stress effects on performance (Dickerson and Kemeny, 2004). Physiological measures, including heart rate (HR) and salivary cortisol (CORT), have helped researchers understand human stress responses, with findings suggesting that people experience up to a 70–80% increase in CORT secretion following TSST exposure (Kudielka et al., 2007).

Dysregulation of HPAA activation is a hallmark feature of various pathological conditions; chronic HPAA activation contributes to cardiovascular diseases (Brosschot et al., 2014), cognitive decline in aging populations (Scott et al., 2015), and exerts detrimental effects on physical and psychological well-being (Deak et al., 2016). Thus, it is vital that researchers use valid and reliable tools to better understand

\* Corresponding author. University of Ottawa, School of Psychology, Behavioral Neuroscience Group, 136 Jean-Jacques Lussier, Vanier building, room 2082, Ottawa, Ontario, K1N 6N5, Canada.

E-mail addresses: [nnarv038@uottawa.ca](mailto:nnarv038@uottawa.ca) (N.F. Narvaez Linares), [vchard074@uottawa.ca](mailto:vchard074@uottawa.ca) (V. Charron), [Allison.Ouimet@uottawa.ca](mailto:Allison.Ouimet@uottawa.ca) (A.J. Ouimet), [plabelle@uottawa.ca](mailto:plabelle@uottawa.ca) (P.R. Labelle), [Helene.Plamondon@uottawa.ca](mailto:Helene.Plamondon@uottawa.ca) (H. Plamondon).

<sup>1</sup> both researchers contributed equally to this manuscript.

<https://doi.org/10.1016/j.ynstr.2020.100235>

Received 20 April 2020; Received in revised form 18 May 2020; Accepted 19 May 2020

Available online 15 June 2020

2352-2895/© 2020 The Authors.

Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

### Abbreviations

|        |   |
|--------|---|
| CORT   | Cortisol  |
| CAR    | Cortisol awakening response                                       |
| HR     | Heart rate  |
| HPAA   | Hypothalamic–Pituitary–Adrenal Axis                               |
| Min    | Minutes   |
| PRISMA | Preferred Reporting Items of Systematic Reviews and Meta-Analysis |
| SD     | Standard deviation  |
| TSST   | Trier Social Stress Test  |

the functional dynamics influencing HPAA activation and behavioral responses in humans (Rubinow et al., 2012). Since its development, researchers have introduced several TSST variants to study different populations, maximize efficiency, and limit inter-individual variability. These variants include the children’s version (Buske-Kirschbaum et al., 1997), the virtual TSST (Jönsson et al., 2010), the TSST for groups (Von Dawans et al., 2011), and the friendly TSST (Wiemers et al., 2013).

#### 1.1. Objectives of the current review

Over 25 years have passed since Kirschbaum et al. (1993) developed the TSST. Although the paradigm appears at first glance unchanged, several researchers have pointed to methodological variations in the TSST protocol, raising concerns about validity and reproducibility of the measurements (Labuschagne et al., 2019; Allen et al., 2014, 2017; Goodman et al., 2017). Although there have been some attempts to examine the effects of some modifications or failure to control variables that are known to affect the TSST (e.g. morning TSST vs. afternoon TSST; Kudielka et al., 2004), to our knowledge, these attempts have been limited in scope.

Existing reviews (Labuschagne et al., 2019; Allen et al., 2014; Allen et al., 2017) addressing elements of the TSST paradigm have significantly raised awareness of the important factors to consider in studying the human stress response. These critical reviews have proposed selective assessments, which deviate from the validated protocol used by narrative systematic review. Goodman et al.’s (2017) meta-analysis focused primarily on the impact of protocol variations on CORT collection, not reviewing other variations of the TSST methodology. Additionally, authors did not use a registered standardized protocol, like the PRISMA-P statement or provide a systematic methodology for the review process (i.e., single vs double reviewer, risk of bias assessment). In this context, Labuschagne et al. (2019) acknowledged limitations of a critical review in that it does not provide a comprehensive assessment of the key elements of the TSST and related variabilities. At present, there is an important need for systematic reviews using “explicit, systematic methods to minimize bias in the identification, selection, synthesis, and summary of studies” (Moher et al., 2015, p. 3). Our review contributes to such a goal, providing a complementary but distinctive approach from that adopted by other review articles.

Considering the widespread and diverse use of the TSST, it is vital that we understand the ways in which methodological variations can affect outcomes. Indeed, such variations, if not considered, can account for divergent findings that researchers otherwise interpret as related to their study hypotheses. For example, hormonal fluctuations associated with menstrual cycles impact the HPAA (Kirschbaum et al., 1999). Maki et al. (2015) found that women in the luteal phase (high levels of estradiol and progesterone) compared to those in the follicular phase (low levels of estradiol and progesterone) performed better on a cognitive task and demonstrated significantly lower levels of cortisol following the TSST. Although the effects of menstrual cycle on HPAA have been documented for more than 20 years, researchers do not

consistently consider menstrual cycle phase in their studies using the TSST (e.g., post-traumatic stress disorder research; Summer et al., 2017; cardiac arrest research; Agarwal et al., 2018); researchers may misinterpret their findings if they do not control for menstrual cycle. Taylor et al. (2010) found that participants exposed to an unsupportive vs. supportive audience showed elevated CORT secretion compared to participants in the no audience condition; the difference between the audience conditions approached significance (Taylor et al., 2010). However, because the researchers did not assess for women’s menstrual cycles, it is unclear whether any of the observed differences were attributable to sex or hormonal differences. Finally, the original TSST protocol contained only a brief description of its methodology (Kirschbaum et al., 1993). As such, it is not surprising that researchers interested in studying stress have introduced variations in the protocol over time, given increasing knowledge and innovation in the field of stress and psychoneuroendocrinology. A formal update of the protocol, therefore, is warranted.

Thus, we aimed first to review various factors (e.g., exclusion criteria, sex/gender differences, menstrual cycle phase) that can impact stress induction and its associated outcomes. Second, we documented systematically the variations in how researchers have applied the TSST protocol since its development. Finally, we used our findings to generate guidelines for reporting TSST methodology, with the goal of fostering reliable comparisons between findings and maximizing replication efforts. We added a theoretical framework section before the methods and the results sections to provide a background rationale for the proposed methodological decisions for our systematic review.

## 2. Theoretical framework

### 2.1. Administration of the TSST and stress-sensitive factors

Before describing our findings in detail, we briefly review the various factors that are known to exert effects on the stress response and its associated effects on HPAA regulation (Hellhammer et al., 2009).

#### 2.1.1. Women participants and hormonal status

The influence of hormonal status on women’s stress responses is well documented (see Fang et al., 2014; Kajantie and Phillips, 2006, for reviews). For instance, women in the non-luteal phase show reduced CORT levels compared to women in the mid-luteal phase (Kirschbaum et al., 1999), and fluctuations of progesterone and estrogen levels can impact women’s emotional and cognitive responses to a stress-inducing task (Felmingahm et al., 2012). Among pre-menopausal and menopausal women, researchers have shown a variability on CORT measures depending on when it is measured (e.g., morning vs. afternoon; Kudielka et al., 2004). Similarly, Rotermann et al. (2015) found that approximately 15% of Canadian women and girls between the ages of 15 and 49 years reported using oral contraceptives in the previous month; oral contraceptive use is associated with reduced CORT secretion, and affects the diurnal neuroendocrine rhythm, particularly morning CORT levels (Roche et al., 2013). Notably, researchers have demonstrated elevated HPAA activation in men following a psychosocial stressor compared to women in the ovarian follicular phase (Stephens et al., 2016), highlighting the potential importance of considering this factor in data analysis and interpretation.

Further, pregnant women experience higher anxiety levels, depressive symptoms, and significantly elevated CORT levels compared to non-pregnant women (Mustonen et al., 2018). Relatedly, breastfeeding may be protective against stress responses. Johnston et al. (2017) found that women who breastfeed also demonstrate lower levels of CORT secretion, particularly in the moments following lactation.

## 2.1.2. Exclusion criteria

**2.1.2.1. Medication.** Pharmacological treatments come with therapeutic and side effects that can alter people's behavioral and physiological responses (e.g., cortisol secretion). Many drugs can influence (a) HPA activation, (b) associated biochemical systems (e.g., regulate sympathetic activation), or (c) participants' subjective experiences (Granger et al., 2009). In a study using the TSST for groups, Houtepen et al. (2015) demonstrated that participants diagnosed with bipolar disorder and taking antipsychotic medication showed a blunted cortisol response compared to their siblings who were also diagnosed with bipolar disorder, but not taking medication (Houtepen et al., 2015). Compiling drug prescriptions from the medical records of 142,377 American citizens, Zhong et al. (2013) found that 68.1% of individuals received a prescription for at least one type of medication, while 51.6% and 21.2% received prescriptions for a minimum of two and five medication types, respectively (Zhong et al., 2013).

**2.1.2.2. Mental health status.** Every year, 1 in 5 Canadians - Worldwide, 1 in 4 individuals (WHO, 2001) - is affected by a psychological disorder, and half the Canadian population will have experienced a psychological disorder before the age of 40 years (Mental Health Commission of Canada, 2013). Even though people with psychological disorders also tend to show reduced research participation rates, there is likely psychological health/disorder variability across samples (Loue and Sajatovic, 2008). Importantly, researchers have demonstrated that dysfunctional HPA activation is associated with various mental health conditions (Wingenfeld and Wolf, 2010). For instance, researchers reported that diagnoses of: (a) Major Depression with melancholic features (referred as melancholic depression in the article), panic disorder, obsessive-compulsive disorder, and schizophrenia could all exert a long-term impact on HPA activation (Jacobson, 2014); (b) social anxiety disorder, autism spectrum disorder (Jacobson, 2014), post-traumatic-stress disorder and atypical depression (Wichmann et al., 2017) can be associated with increased sensitivity to stress and/or lead to altered HPA activation. As such, the decision to include/exclude individuals who are or have been affected by a mental disorder needs careful consideration, not only for ethical reasons, but also when the association of the particular condition with stress responses is not the primary objective of the study (Allen et al., 2014).

**2.1.2.3. Tobacco use.** Assessments of the relationship between cortisol secretion and tobacco smoking have generated mixed findings. In a sample of 4231 people (73% men/27% women), Badrick et al. (2007) evaluated smoking status and salivary cortisol measures and they found that current smokers demonstrated increased salivary cortisol secretion (also observed in Cohen et al., 2019) compared to ex-smokers or people who had never smoked or quit smoking. Such findings support delayed impact of smoking on cortisol secretion and smoking a single cigarette is sufficient to induce cortisol secretion and HPA activation (i.e., increased HR; Rohleder and Kirschbaum, 2006). Despite the importance of smoking status, researchers define smoking status in various ways; there is no consensus on operational definitions of 'regular' or 'past' smokers (Ryan et al., 2012).

**2.1.2.4. Substance use.** Substance use disorders and psychoactive substance consumption are known to show differential effects on the neuroendocrine system. The central and peripheral nervous systems work in tandem to maintain homeostasis, a crosstalk that may be compromised by drug intake. For example, researchers have observed immediate increases in cortisol secretion (1–2 fold in magnitude) following 3,4-Methyl enedioxy methamphetamine (MDMA) consumption; people who reported regular MDMA consumption for 3 months also demonstrated a 400% increase in cortisol secretion compared to people who did not consume MDMA (Parrot et al., 2014). Other

psychostimulants, including amphetamines and cocaine, similarly increase plasma cortisol levels, and can affect mood and cardiovascular functions (Manetti et al., 2014). Findings are mixed among people who use cannabis, with some supporting both increased and decreased HPA activation, as well as effects on basal and awakening cortisol secretion profiles (Cservenka et al., 2018). Finally, one of the most common substances used is caffeine. The influence of caffeine on the HPA is well documented. For example, Patz et al. (2006) found that although low to moderate doses did not modulate HPA, they led to significant increases in corticosterone levels, which took 60 min to return to their initial levels. Higher doses impacted the HPA for up to 120 min. Burke et al. (2016) found that caffeine consumption impacted the circadian cycle and suggested that it could impact the secretion of other hormones. Additionally, there has been an increase among youth and college students in consumption of energy drinks, which not only contain higher doses of caffeine, but also contain other ingredients that are likely to have an impact on people's health (Malinauskas et al., 2007; Ibrahim and Ifitkhar, 2014; Shah et al., 2019).

**2.1.2.5. Body weight.** In North America, excluding Mexico, approximately two thirds of adults aged 20–64 years are overweight, and one in three people is considered obese (Flegal et al., 2012; Government of Canada, 2018). Worldwide about two in five people are overweight and one in ten are obese (WHO, 2016). At present, the impact of body weight on neuroendocrine system functioning remains uncertain (Incollingo Rodrigues et al., 2015); being overweight has been associated with hyper- (Odeniyo et al., 2015) and hypo- (Herhaus and Petrowski, 2018) responsiveness of HPA. On the other end of the continuum, Schorr et al. (2015) found that participants with a low body mass index (BMI; Kg/m<sup>2</sup>) demonstrated increased cortisol secretion compared with participants with a BMI in the normal weight range. More specifically, Monteleone et al. (2016) found an enhanced cortisol awakening response (CAR) in severely underweight participants with anorexia nervosa, but not in weight-restored participants; in other words, a dysregulation of CAR appears highly correlated with weight (Monteleone et al., 2016). Moreover, Pasquali et al. (2006) suggested that participants' BMI remains an important variable in understanding how diverse weight ranges may impact the stress response.

**2.1.2.6. Chronic diseases.** Chronic diseases include diverse medical conditions that persist across time and cause impairments in daily living. Such diseases represent the leading cause of death and disability worldwide (WHO, 2019). Diabetes, coronary heart disease, cancer and chronic obstructive pulmonary disease (Dis-Chaves et al., 2016; Matura et al., 2018) have all been associated with increased cortisol secretion. In North America, an estimated 25–33% of the adult population live with one or more chronic health conditions (Ward et al., 2012; Branchard et al., 2018), often associated with dysregulations of HPA activation (Allen et al., 2014). Furthermore, in a recent Danish study ( $N = 4,555, 439$ ), Hvidberg et al. (2020) reported that two-thirds of people aged 16 years and older (65.6%) had at least one chronic disease.

**2.1.2.7. Working night shifts.** Given that irregular shifts are a staple of a 24/7 global economy, 15–30% of American and European adults report working night shifts, and an additional 19% report working regularly for periods extending over 2 h between 10 p.m. and 5 a.m. (Boivin and Boudreau, 2014). Researchers have demonstrated that night shifts and irregular working schedules impact circadian cycles (Boivin and Boudreau, 2014) and exert profound effects on physical and psychological health associated with dysregulation of HPA and diurnal cortisol secretion profiles (Charles et al., 2016; Gonnissen et al., 2013). Therefore, more rigorous screening of work shifts is important to consider, especially if the sample of participants are undergraduate students. For example, in a longitudinal study, Lund et al. (2010) found that undergraduate students reported chronically limited sleep, which appeared to

lead to other problems (e.g., more frequent consumption of alcohol and drugs). Thus, it is important to take this factor into consideration.

### 2.1.3. Restriction of activities prior to participation

When assessing CORT secretion using blood or saliva samples, environmental factors can interact with the collected measures (Garde et al., 2009). Various activities performed prior to a saliva sample have been shown to enhance endocrine measures (Kudielka et al., 2009), including: brushing or flossing teeth, receiving false results due to blood contamination (Kudielka et al., 2012; Stalder et al., 2016), physical exercise (Rahman et al., 2010), food consumption, (Stalder et al., 2016), caffeinated beverages (Patz et al., 2006), smoking (Rohleder and Kirschbaum, 2006) and any substance (Zhou et al., 2010) or alcohol (Badrick et al., 2008) use. Additionally, heavy alcohol consumption can negatively impact cognitive functions and performance on everyday tasks (Gunn et al., 2018) and therefore affect TSST performance (Badrick et al., 2008).

## 2.2. TSST protocol application

### 2.2.1. Selected resting period

Researchers typically include a rest period before the TSST to account for multiple factors involved when participants take part in a laboratory session, including the participants' means of transportation being possibly associated with stress or increased sympathetic activation, general stress related to participating in a study, or anxiety associated with the performance task or with interacting with unknown people including the researcher (Kudielka et al., 2007). Additionally, Rahman et al. (2010) found that it takes the body up to 60 min to return to a homeostatic state after vigorous physical activity. Indeed, letting participants acclimatize to the research environment ensures that they have the lowest physiological activation possible before the TSST to assess the real impact on CORT secretion.

### 2.2.2. Period of the day

The circadian rhythm of CORT is the gold standard measure to establish an optimal starting time in neuroendocrine studies (Liu et al., 2017). CORT levels are lower in the afternoon than in the morning (Matsuda et al., 2012). Specifically, 30–45 min following awakening, people's CORT levels increase by 50 to 156% before declining throughout the day (Stalder et al., 2016). Recent research demonstrated that morning CORT measures tend to be more accurate and representative, and less affected by external factors (e.g., caffeine intake; Matsuda et al., 2012). Women's and older individuals' CORT levels tend to differ from normalized values (e.g., women demonstrate a greater and prolonged response than men); researchers suggest paying close attention to these differences (Stalder et al., 2016). Finally, because the body will naturally tend to re-establish homeostasis after a certain period, the duration of the TSST session may impact the TSST outcome measures.

### 2.2.3. Self-report anxiety/stress questionnaires

Researchers frequently use self-report questionnaires to assess participants' subjective experience of anxiety and its relation to their observed behavioral and physiological responses. There is some evidence that individual differences across a variety of factors can influence under- or over-reporting of subjective anxiety. For example, Karlson et al. (2011) found that women who reported higher job stress demonstrated higher CAR, whereas men who reported lower job stress demonstrated higher CAR. In other words, both sex and perceived work stress impacted upon HPA activation (Karlson et al., 2011).

Measuring subjective anxiety may be important not only for understanding its effect on psychophysiological functioning, but also for understanding the role of psychophysiological functioning on real-life distress or impairment. Indeed, there are multiple examples in the literature demonstrating a lack of relationship between subjective and objective measurement of anxiety (e.g., Campbell-Sills et al., 2006; De

Los Reyes et al., 2012; Puigcerver et al., 1989; Wilhelm and Roth, 2001). Some of these discrepancies are likely related to how people interpret their own physiological symptoms (i.e., as dangerous or not), and to the type of subjective anxiety that researchers measure (i.e., specific or general; De Los Reyes et al., 2012). Thus, it is possible that variations in subjective measurement result in different patterns of findings across studies using the TSST.

### 2.2.4. Panel of judges and video recording

Notwithstanding the gender of a participant, Allen et al. (2014) concluded in their critical review that a same gender across the panel of judges significantly influenced responses on a threat evaluating situation, especially when testing young men or women. Additionally, men and women demonstrate different HPA responses to stressors (Dickerson and Kemeny, 2004; Kudielka et al., 2007). Specifically, men tend to exhibit higher CORT levels than women, whose responses vary depending on menstrual cycle or hormonal contraceptive use. Duschesne et al. (2012) demonstrated that both men and women exhibited increased CORT secretion only when performing in front of judges of the opposite gender. However, this effect was present only in women in the follicular phase (Duschesne et al., 2012). Moreover, many researchers videorecord participants' performance, which is associated with amplified feelings of threat and elevated stress response (Biondi and Picardi, 1999). Thus, whether or not a TSST protocol includes videorecording may impact upon TSST outcomes.

### 2.2.5. Speech and arithmetic tasks

In a review article addressing the effects of public speaking on fear and anxiety and its impact on people's physiology and perception, Garcia-Leal et al. (2014) found that 30–50% of people reported a fear of public speaking with approximately 40% of those reporting anxiety about being negatively evaluated by others (Stein et al., 2010). Speech tasks are common across different research areas given their ability to create a social-evaluative threat (Dickerson and Kemeny, 2004). Buchanan et al. (2014) demonstrated that people who engaged in the TSST demonstrated reduced speech fluency and increased physiological reactivity compared to those in a non-stressful condition. Participants in the TSST condition showed higher word productivity (i.e. the ratio of productive words to total words) and paused more often during their speech than those in the non-stressful condition. This effect was pronounced in participants who evidenced higher cortisol and heart rate responses to the TSST, highlighting that speech tasks effectively induce stress and cause physiological changes in participants (Buchanan et al., 2014).

Similarly, close to 20% of the general population experiences some level of anxiety related to performing mathematical tasks (Dowker et al., 2016). Indeed, in a review performed by Caviola et al. (2017), they found that participants who performed an arithmetic task under timed conditions tended to show a weaker performance than participants under untimed conditions, a phenomenon related to the interference of different task-associated cognitive domains, including working memory (Caviola et al., 2017).

### 2.2.6. Recovery period

A recovery period following the TSST enables researchers to measure CORT variation until its return to baseline levels, when the effects of the TSST are attenuated. Physiological stress persists after stress exposure, but the duration of the effects is still unclear (Brosschot et al., 2014). Therefore, researchers have included a recovery period following the TSST (Kirschbaum et al., 1993) to quantify these differences subsequent to a stressor.

## 2.3. Physiological measures

### 2.3.1. Heart rate

HR measures represent a sensitive tool enabling researchers to

determine sympathetic nervous system activation in an experimental context. The [Task Force of the European Society of Cardiology \(1996\)](#) recommends taking HR measures every 5 min (cluster) for a short clinical study like the TSST. HR is a valid physiological measure to assess stress levels. [Hellhammer and Schubert \(2012\)](#) found that HR measures were significantly higher before and during the TSST and significantly lower during the recovery period, indicating a peak in physiological stress during the TSST.

### 2.3.2. Cortisol collection

Assessing salivary cortisol is a reliable, quick, and non-invasive way to determine changes in CORT levels ([Hellhammer et al., 2009](#)), making this method the most popular in human studies ([Liu et al., 2017](#)). Following a short-term stressor like the TSST, participants show increased salivary cortisol levels, reaching peak levels 15–20 min after the stressor ([Kudielka et al., 2009](#); [Kirschbaum et al., 1993](#)). Recommendations according to [Kirschbaum et al. \(1993\)](#) and a meta-analysis conducted by [Liu et al. \(2017\)](#) are that baseline cortisol measures be taken 30 min before the TSST, with time intervals between 10 and 25 min and the last measure occurring between 30 and 70 min following the start of the TSST. [Juster et al. \(2012\)](#) and [Dickerson and Kemeny \(2004\)](#) validated that cortisol reaches peak levels after the 20 min following the beginning of the TSST.

### 2.3.3. Blood collection

Blood collection can help characterize endogenous biochemical changes (e.g., neurotransmitter and/or hormones) to address specific objectives. To standardize and facilitate study replication, recommended time intervals for blood collection mirror those of cortisol. However, researchers who collect blood samples should minimize participant discomfort to avoid additional unwanted stress ([Birkett, 2011](#)) given its invasiveness.

## 3. Method

### 3.1. Protocol and registration

Prior to beginning our review, we registered our protocol in accordance with the PRISMA-P checklist ([Moher et al., 2015](#)) with PROSPERO ([CRD42017069908](#)). The protocol was registered on June 21, 2017 and was last updated on April 15, 2019. There were three updates to clarify the wording and update the status of the systematic review; no changes to the protocol were made.

### 3.2. Literature search strategy

We focused our review on studies wherein researchers used the original version of the TSST ([Kirschbaum et al., 1993](#)) with adult populations (18 years and older). A social sciences research librarian (P.R. L.) with expertise in knowledge syntheses assisted in drafting, developing, and implementing a search strategy that would retrieve relevant results from the following databases: APA PsycInfo (Ovid), Medline (Ovid), Web of Science, and Scopus. We developed a keyword search across multiple fields specific to the concepts of the TSST, psychosocial stressors, and speech tasks (See [Appendix D](#) for the complete search strategy). Database limits were not used at this stage of the review. The search was initially conducted in December 2017 (T<sub>1</sub>) and updated in July 2019 (T<sub>2</sub>). Articles collected at those time points were reviewed using the same criteria. We used [Zotero](#) (Roy Rosenzweig Center for History and New Media) and [Covidence](#) (Veritas Health Innovation) to manage references and complete the research for T<sub>1</sub> and T<sub>2</sub>, respectively (see [Fig. 1](#)).

### 3.3. Procedure

A group of undergraduate and graduate students were trained to

screen references. First, two independent reviewers screened the titles and abstracts to assess their relevance. In case of doubt or uncertainty, N.F.N.L. reviewed the title and abstract. If an article did not include an abstract, N.F.N.L. attempted to locate it online using Google Scholar, Pubmed, journal Web sites, and other databases. If unsuccessful, the study was categorized as not found and excluded. Studies with a brief or vague summary were triaged by N.F.N.L. to review and classify. During the second phase, all references that remained were screened by two independent reviewers according to the study eligibility criteria, section 3.4. The third phase consisted of a second full-text screening by two independent reviewers in conformity with study selection, section 3.5.

For clarification purposes, note that T<sub>1</sub> and T<sub>2</sub> refer to the dates when the literature searches were conducted. Only for T<sub>1</sub>, the principal authors (N.F.N.L. & V.C.) added an extra step to increase efficiency during data extraction by removing articles that did not provide adequate information about their use of the TSST, making data extraction and interpretation impossible. Specifically, while doing full-text screening for the first time (section 3.4), studies that reported using the TSST but provided limited or no procedural details (e.g., failed to report information concerning the speech or arithmetic tasks, only referred to the original article) were removed.

### 3.4. Exclusion criteria

During the title/abstract and full-text screening phases we used the following criteria to exclude ineligible studies. Studies were excluded if one of the following characteristics applied: (a) inclusion of participants 17 years old or younger, mixed sample of adults and children, or longitudinal studies with participants under 17 years or younger; (b) presented as a conference publication (given methodological information is summarized and not detailed) or erratum study (c) book chapter or graduate thesis (not peer-reviewed); (d) not published in English; (e) did not use the original TSST, for example: group TSST, virtual TSST, Toxic Shock Syndrome, animal studies, meta-analysis or systematic reviews, research on stress not using the TSST, research on psychosocial factors, program evaluations, Tubien Scotopic Threshold Test, and adaptations of the TSST (i.e., using a single task, or completely modifying all tasks from the original TSST version).

### 3.5. Study eligibility criteria

To be included, articles had to report on several specific methodological aspects. We assessed inclusion using sequential steps in order of importance from a to e: (a) testing time window; (b) number of judges (we also noted if there was a mention of judges giving positive or negative feedback at any point during the TSST); (c) nature of speech and arithmetic tasks; (d) collection of cortisol and at least one type of self-report measure (e.g., anxiety questionnaires); and (e) HR or blood collection. Specifically, articles failing to provide information about the testing time window were eliminated. Then, articles failing to indicate information about the judges were eliminated. Then, criteria related to reporting about the speech or arithmetic task (e.g., speech's type and duration, initial number for the arithmetic task), cortisol measures, anxiety questionnaires, HR or blood collection measurements were screened.

### 3.6. Data extraction

The following information was extracted from retained studies: (a) title of the article, authors' name, and year of publication; (b) number of participants for each sex<sup>2</sup> (men, women, or the combination of both),

<sup>2</sup> We use the term 'sex' to refer to biological sex at birth. We use the term 'gender' to refer to an individual's socially-constructed gender role and gender expression.

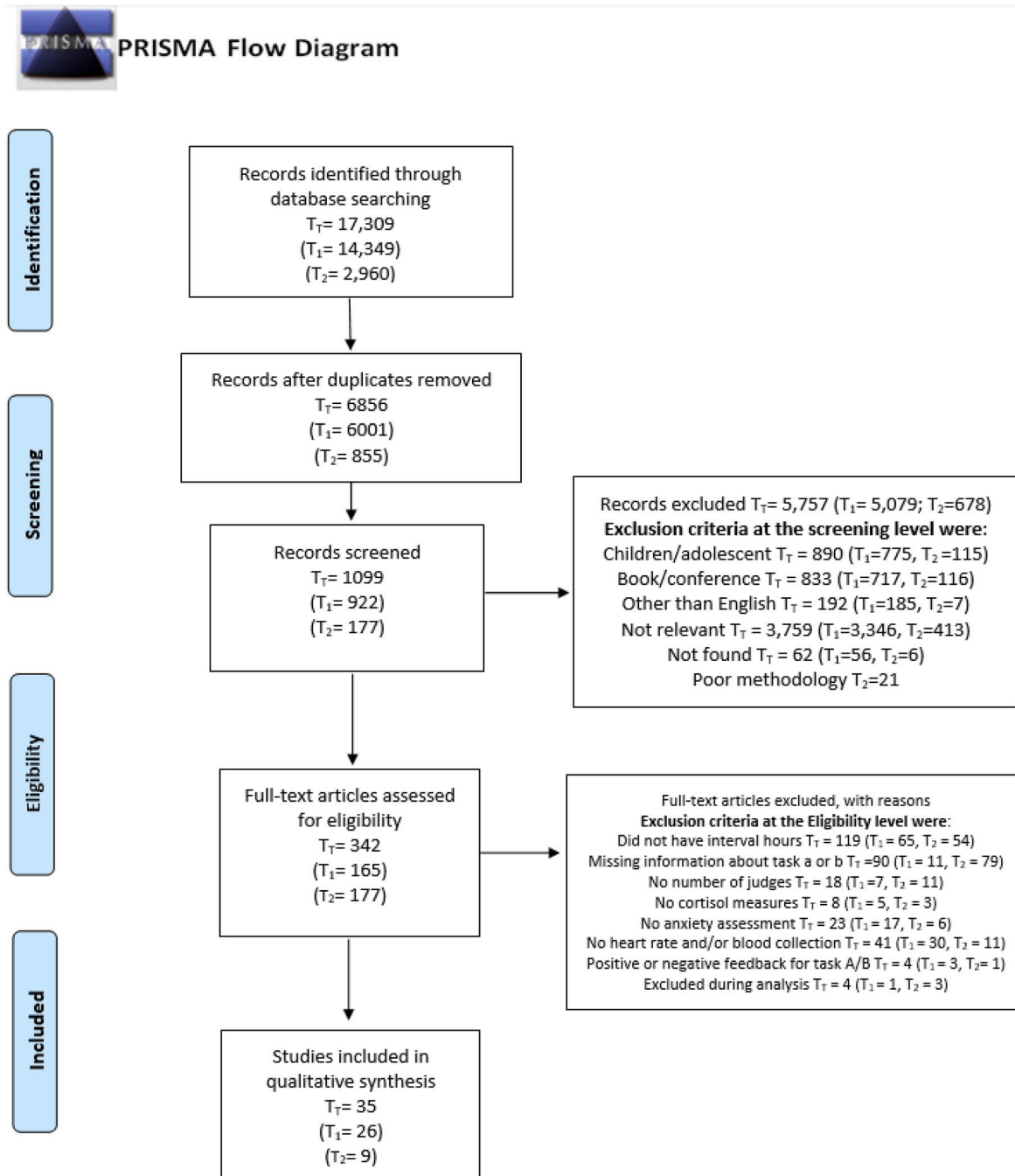


Fig. 1. Flow Diagram of the selection of the studies included in the systematic review (Moher et al., 2015).  $T_1$  and  $T_2$  refer to the dates when the literature searches were conducted.

mean age and standard deviation for each sex, menstrual cycle stage for women if provided and any exclusion or inclusion criteria related to menstrual cycle; (c) exclusion criteria; (d) self-report measures and associated administration time; (e) number and gender of judges; (f) presence of recording equipment; (g) modifications to the TSST protocol; (h) TSST administration time; (i) resting period duration from arrival to TSST initiation; (j) controlled or prohibited activities (e.g., drinking or smoking); and required abstinence period prior testing; (k) instructions for participants' speech preparation and delivery, and feedback received from judges; (l) characteristics of the arithmetic task, i.e., initial number used, number to be subtracted, and task duration; (m) resting period duration following TSST; (n) physiological measures (e.g., sampling frequencies and intervals).

### 3.7. Risk of bias

Given our goal of extracting methodological information rather than specific outcomes or effect sizes, we did not conduct a typical risk of bias assessment. In addition, we did two full text-screens to ensure accuracy.

## 4. Results

The database search (section 3.2) yielded a total ( $T_T$ ) of 17,309 studies ( $T_1 = 14,349$ ;  $T_2 = 2960$ ). After eliminating duplicates, 6856 ( $T_1 = 6001$ ;  $T_2 = 855$ ) studies remained in the database. Following title and abstract screening, 5757 ( $T_1 = 5079$ ;  $T_2 = 678$ ) studies were excluded. The remaining 1099 ( $T_1 = 922$ ;  $T_2 = 177$ ) studies proceeded to the full-

text examination portion of the procedure (section 3.5). As reported in the procedural section 3.3, following an initial review of studies, we added one supplementary exclusion criterion related to poor reporting of TSST details for T<sub>1</sub>. This step eliminated 757 articles, leaving 165 articles for T<sub>1</sub>. For T<sub>2</sub>, we applied these criteria during the study eligibility phase (section 3.4). Finally, we retained 39 articles (T<sub>1</sub>) following the study selection phase (section 3.5). The first authors excluded an additional four studies because the described methodology appeared inconsistent with the findings reported in the figures or in the result section. For example, the times at which cortisol measures were taken did not match what was reported in the methodology section and/or the figures. Therefore, 35 articles were ultimately included in this review. We provided a summary table of the study characteristics in appendix C and an excerpt in Table 1. In the reference list, articles included in the review are identified by an \*.

As explained in section 3.5, articles were excluded in sequential steps. However, we remained interested in the number of articles excluded based on the distinctive criteria. Overall, of the 165 articles examined in T<sub>1</sub>, insufficient details were found concerning the: testing time window (100 articles - 60.6%); speech or arithmetic tasks (13 articles - 9.7%); number of judges (23 articles - 13.9%); or self-report (13.9%), blood (93 articles - 56.4%) or HR (132 articles - 80%) measures. We excluded 12 articles (7.3%) because the judges provided positive or negative feedback during the speech or arithmetic task. Several articles fell into more than one of the exclusion criterion categories, thus demonstrating a general lack in systematic reporting of methodological details.

#### 4.1. Administration of the TSST and stress-sensitive factors

Please consult Appendix C for complete characteristics of all the screened studies.

##### 4.1.1. Women participants and hormonal status

Twenty-three studies included women participants. Among these studies, seven omitted mention of the menstrual cycle while 16 controlled for estrous cycle phases. Specifically, five included women in the luteal phase only, four included women in the follicular phase only, and three included women in either phase. Among the three studies wherein researchers did not control for either phase, only Maki et al. (2015) found a significant difference in stress responses between women in the follicular and luteal phase, average age in both groups 25.60 (5.39 SD) and 28.05 (5.83 SD), respectively. Twelve of the 23 studies (52%) controlled for oral contraceptive use and 10 controlled for pregnancy, lactation, or breastfeeding. Finally, two studies controlled for post-menopausal status while the other studies excluded women who were post-menopausal, ovulating, in the follicular phase, or had irregular estrous cycles.

##### 4.1.2. Exclusion criteria

In total, 26 different exclusion criteria were reported in the 35 articles. A total of 303 exclusion criteria were screened in the complete sample; studies used approximately nine exclusion criteria (mean = 8.66 ± 3.11 SD), some of the most susceptible to impact results being detailed here.

**4.1.2.1. Medication.** Thirty-two studies reported excluding participants taking prescribed drugs (i.e., psychoactive medication, all medication, or any medication consumed on a regular basis).

**4.1.2.2. Mental health status.** Twenty-seven studies reported excluding individuals with mental illness, psychiatric disorder and/or DSM-IV Axis 1 disorders.

**4.1.2.3. Tobacco use.** Twenty-two studies reported excluding

participants based on nicotine/tobacco consumption. Specifically, seven controlled for no smokers, five excluded people who reported smoking more than five cigarettes per day, three excluded “regular smokers”, two excluded “current smokers” and one excluded people with “low levels” of smoking. For the last three categories, the studies did not report the specific criteria used to define each state. Finally, four studies did not provide information about nicotine/tobacco consumption.

**4.1.2.4. Substance use.** Nineteen studies reported excluding participants based on (problematic) substance use or a substance use disorder.

**4.1.2.5. Body weight.** Twelve studies reported excluding participants based on their BMI; seven of these excluded participants who did not meet a minimum BMI threshold, while ten excluded participants for exceeding a specific BMI value. Generally, the BMI of the study participants ranged between a minimum of 18–20 kg/m<sup>2</sup> and a maximum of 26–35 kg/m<sup>2</sup>.

**4.1.2.6. Chronic diseases.** Ten studies reported excluding participants if they had a chronic disease, but provided no additional detail.

**4.1.2.7. Working night shifts.** Only four studies reported excluding participants if they worked night shifts.

##### 4.1.3. Restriction of activities prior to participation

In order for participants to be eligible for the TSST studies, abstinence from certain behaviors was required. Although restrictions varied, some were recurrent. Researchers asked participants to refrain from: eating (24 studies), drinking anything other than water (16 studies), engaging in physical exercise (16 studies), drinking coffee (14 studies), drinking alcohol (nine studies), using drugs (six studies), smoking (six studies), and brushing or flossing their teeth (two studies). Given the samples for each activity are either relatively small or the difference in time is large, it was difficult to assess an average time to refrain from each activity. Overall, the restriction period varied widely for the different behaviors extending between 60, 90, 120, 240, 270, 720, 1440, 2880, and 4320 min prior to participation, a consideration that will be addressed in relation to the specific activities in the discussion section.

#### 4.2. Differences in the TSST protocol application among the sampled studies

##### 4.2.1. Selected resting period

Nine studies reported having modified certain aspects of the TSST from the original protocol. Across reviewed studies, 12 different resting/habituation periods prior to TSST exposure were used, many substantially extending the initially proposed habituation time. In the original article, participants rested 30 min following a catheter placement or 10 min when no blood samples were collected (Kirschbaum et al., 1993). On average, participants waited 55 ± 7.42 min. However, no consensus emerged between studies; the resting periods were: 5 min (one study), 10 min (three studies), 20 min (two studies), 30 min (13 studies), 40 min (one study), 45 min (two studies), 60 min (six studies), 70 min (one study), 90 min (three studies), 180 min (one study), 210 min (one study), and 240 min (one study).

##### 4.2.2. Period of the day

Eleven studies reported initiating daily testing at the same time for every participant. Researchers often provided a time window for participants to be tested, rather than one specific time. For example, “Participants reported to the laboratory between 1200 and 1600” (Buchanan et al., 2014, section 2.4), “the TSST experimental sessions were run between the hours of 1100 and 1600” (Drake et al., 2017, TSST protocol section), or “TSSTs were administered in the afternoon between 2 p.m. and 5 p.m.” (Het et al., 2015, Procedure section).

**Table 1**  
Speech Task, Arithmetic Task and Judges and videotape - Example of information provided in [Appendix C](#).

| Study                        | Speech task              |                              |   | Arithmetic Task   |                                   |                                 | Judges and videotape                |                                |                                  |                             |
|------------------------------|--------------------------|------------------------------|---|-------------------|-----------------------------------|---------------------------------|-------------------------------------|--------------------------------|----------------------------------|-----------------------------|
|                              | Preparation period (min) | Time to deliver speech (min) | Type of speech  | Digit number used | Digit number used for subtraction | Time to complete the task (min) | Number of judges used for the panel | Number of judges that were men | Number of judges that were women | Mentioned it was videotaped |
| Abelson et al. (2014)        | 3                        | 5                            | Job interview   | 1022              | 13                                | 5                               | 2                                   | 1                              | 1                                | Yes                         |
| Bae et al. (2019)            | 5                        | 5                            | Job interview   | 2043              | 17                                | 5                               | 2                                   | –                              | –                                | Yes                         |
| Böbel et al. (2018)          | 3                        | 5                            | Job interview   | 3079              | 17                                | 5                               | 2                                   | –                              | –                                | Yes                         |
| Buchanan et al. (2014)       | 5                        | 5                            | Accusation defense                                      | 1022              | 13                                | 5                               | 1                                   | –                              | –                                | No                          |
| Drake et al. (2017)          | 10                       | 5                            | Job interview   | 1022              | 13                                | 5                               | 3                                   | Both males and females         | –                                | Yes                         |
| Erickson et al. (2017)       | 3                        | 5                            | Job interview   | 1022              | 13                                | 5                               | 2                                   | –                              | –                                | Yes                         |
| Elzinga & Roelofs (2005)     | 5                        | 5                            | Job interview   | 1587              | 13                                | 5                               | 3                                   | –                              | –                                | Yes                         |
| Fries et al. (2006)          | 3                        | 5                            | Job interview   | 2083              | 17                                | 5                               | 3                                   | 2                              | 1                                | Yes                         |
| Giles et al. (2014)          | 10                       | 5                            | Job interview   | 1223              | 17                                | 5                               | 3                                   | –                              | –                                | Yes                         |
| Gröpel et al. (2018)         | 5                        | 5                            | Job interview   | 2010              | 13                                | 5                               | 2                                   | –                              | –                                | Yes                         |
| Het et al. (2015)            | 5                        | 5                            | Job interview   | 2043              | 17                                | 5                               | 2                                   | –                              | –                                | No                          |
| Inagaki & Eisenberger (2016) | 5                        | 5                            | Administrative assistant position                       | 2083              | 13                                | 5                               | 2                                   | 1                              | 1                                | Yes                         |
| Jiang et al. (2017)          | 5                        | 5                            | Accusation defense                                      | 1022              | 13                                | 5                               | 3                                   | –                              | –                                | Yes                         |
| Kern et al. (2008)           | 3                        | 5                            | Job interview   | 2043              | 17                                | 5                               | 2                                   | –                              | –                                | Yes                         |
| Klatzkin et al. (2018a)      | 5                        | 5                            | Job as a campaign manager for a local politician        | 1022              | 13                                | 5                               | 3                                   | –                              | –                                | Yes                         |
| Klatzkin et al. (2018b)      | 5                        | 5                            | Job interview   | 2000              | 7                                 | 5                               | 3                                   | –                              | –                                | Yes                         |
| Klatzkin et al. (2019)       | 5                        | 5                            | Job interview   | 2000              | 7                                 | 5                               | 3                                   | –                              | –                                | Yes                         |
| Li et al. (2015)             | 5                        | 5                            | Job interview   | 2043              | 17                                | 5                               | 2                                   | 1                              | 1                                | Yes                         |
| Lupis et al. (2014)          | 5                        | 5                            | Job interview   | 2043              | 17                                | 5                               | 2                                   | –                              | –                                | Yes                         |
| Maki et al. (2015)           | 10                       | 5                            | Job interview   | 1687              | 13                                | 5                               | 3                                   | –                              | –                                | Yes                         |
| McInnis et al. (2014)        | 3                        | 5                            | Job interview   | 2043              | 17                                | 5                               | 2                                   | –                              | –                                | Yes                         |
| McInnis et al. (2015)        | 3                        | 5                            | Job interview   | 2043              | 17                                | 5                               | 2                                   | –                              | –                                | Yes                         |
| Oswald et al. (2004)         | 10                       | 5                            | Describe qualifications for hospital administrator      | 2322              | 13                                | 5                               | 3                                   | –                              | –                                | Yes                         |
| Polheber & Matchock (2014)   | 3                        | 5                            | Job interview   | 2023              | 17                                | 5                               | 3                                   | *mix-gendered*                 | –                                | Yes                         |
| Reinelt et al. (2019)        | 5                        | 5                            | Job interview   | 2043              | 17                                | 5                               | 2                                   | 1                              | 1                                | Yes                         |
| Shalev et al. (2011)         | 5                        | 5                            | Job interview   | 1687              | 13                                | 5                               | 2                                   | –                              | 2                                | Yes                         |
| Smith et al. (2016)          | 10                       | 5                            | Job interview   | 1022              | 13                                | 5                               | 3                                   | –                              | –                                | Yes                         |
| Souza et al. (2015)          | 10                       | 5                            | Good candidate for a future peacekeeping mission (army) | 910               | 7                                 | 5                               | 2                                   | 2                              | –                                | Yes                         |
| Thomas et al. (2011)         | 5                        | 5                            | Job interview   | 1022              | 13                                | 5                               | 3                                   | –                              | –                                | No                          |
| Veer et al. (2011)           | 10                       | 5                            | one's positive and negative characteristics             | 1033              | 13                                | 5                               | 3                                   | –                              | –                                | No                          |

(continued on next page)



Table 1 (continued)

| Study                 | Speech task              |                              |  | Arithmetic Task   |                                   |                                 | Judges and videotape                |                                |                                  |                             |
|-----------------------|--------------------------|------------------------------|--|-------------------|-----------------------------------|---------------------------------|-------------------------------------|--------------------------------|----------------------------------|-----------------------------|
|                       | Preparation period (min) | Time to deliver speech (min) | Type of speech                                     | Digit number used | Digit number used for subtraction | Time to complete the task (min) | Number of judges used for the panel | Number of judges that were men | Number of judges that were women | Mentioned it was videotaped |
| Wand et al. (2007)    | 10                       | 5                            | Describe qualifications for hospital administrator | 2322              | 13                                | 5                               | 2                                   | –                              | –                                | Yes                         |
| Wiegand et al. (2018) | 10                       | 5                            | Job interview                                      | 1022              | 13                                | 5                               | 2                                   | –                              | –                                | Yes                         |
| Xin et al. (2017)     | 5                        | 5                            | Accusation defense                                 | 1022              | 13                                | 5                               | 3                                   | 1                              | 2                                | Yes                         |
| Yao et al. (2016)     | 5                        | 5                            | Accusation defense                                 | 1022              | 13                                | 5                               | 3                                   | 1                              | 2                                | Yes                         |
| Zhang et al. (2019)   | 5                        | 5                            | Accusation defense                                 | 1022              | 13                                | 5                               | 3                                   | 1                              | 2                                | Yes                         |

On the one hand, providing time windows (rather than specific times) to participants allows researchers to accommodate different participant needs and increase study feasibility. On the other hand, time windows likely increase the variability of the stress response data, given that physiological measures are not taken at the exact same time of the day. Researchers reported sixteen times for participants to begin the study, varying from 08:30 to 18:30, with a window of  $3.16 \pm 1.77$  h in average. Forty-six ending times were reported, varying from 10:30 to 21:40. For example, if participants arrive between 14:00 and 17:00 (3-h interval) and the experiment lasts  $\sim 70$  min, the study end time will vary between 15:10 and 18:10. As shown in Table 2, providing such flexibility to participants increases the variability between studies given that both the initiation times for each study and the duration of each study (i.e., resting period, duration of TSST, and recovery period) vary. This variation in time may be an important reason explaining why systematic reviews/meta-analyses report conflicting results. We found that in half of the studies we reviewed, researchers asked participants to stay for an additional 60.29 ( $\pm 7.64$ ) min following the TSST, enabling them to evaluate the delayed impact of the test on physiological measures.

#### 4.2.3. Self-report anxiety/stress questionnaires

Researchers used 18 different questionnaires to assess participants' subjective state anxiety. On average, studies measured subjective state anxiety using more than a single questionnaire ( $2.23 \pm 1.24$ ). All studies used at least one measure, 20 used at least two measures, 15 used three measures, and eight used four measures. The State-Trait Anxiety Inventory (Spielberger et al., 1983), the Positive and Negative Affect Schedule (Watson et al., 1988), and Visual Analogue Scales (e.g., Heller, et al., 2016) were the most popular questionnaires; they were used in 18, 14 and 11 studies, respectively. Other scales included the Perceived Stress Scale (nine studies; Cohen et al., 1983) and the Beck Anxiety Inventory (nine studies; Steer and Beck, 1997); the remaining questionnaires were used in fewer than two studies and some of them seem to have been used as screening tools. Seventeen different questionnaire combinations were listed, although no consensus for a particular combination of questionnaires was observed. The majority of studies used questionnaires before and after TSST completion allowing researchers to determine changes in subjective state anxiety due to the stressors.

#### 4.2.4. Judges' number, gender and video recorded sessions

In most studies (31 out of 35), researchers told participants that the TSST session would be videorecorded for further analysis. Only one study used a single judge, whereas the other studies reported using two (17 studies) or three (17 studies) judges. As for the judges' gender, 10 studies used both men and women on the panel, one study used only women, one study used only men, and twenty-four studies failed to report the genders of the judges.

#### 4.2.5. Speech task

For the speech task, we noted three preparation times. Eight, eighteen, and nine studies allotted 3, 5, and 10 min of preparation time, respectively. All studies reported a speech duration period of 5 min. Seven types of speeches were reported. The majority (24 studies) used a specific job interview task (asking participants what makes them the best candidate for their dream job), five used an accusation defense task, two asked participants to describe personal qualifications in an administrative job at a hospital, one asked participants to comment on their qualifications for a future peacekeeping mission, one asked participants to present positive and negative personal characteristics, one asked participants to present their qualifications as a future administrative assistant, and one asked participants to describe their qualifications as a campaign manager for a local politician. We identified nine combinations of preparation times and speech types. The preparation times for studies using a job interview were three (8 studies), five (11 studies), and 10 min (5 studies). Five studies used an accusation defense speech and provided 5 min of preparation time and two studies provided 10 min preparation time for participants to describe qualifications as a hospital administrator. The remaining studies used either 5 min (i.e., administrative assistant or campaign manager for a local politician) or 10 min of preparation time (i.e., qualifications for future peacekeeping mission, personal positive and negative characteristics).

#### 4.2.6. Arithmetic task

Researchers used 13 different initiation and subtraction numbers in the arithmetic task across 14 different number combinations. The two number sets predominantly selected (initial and subtraction numbers) were 1022 and 13 (12 studies) and 2043 and 17 (8 studies). For the remaining studies, 1687/13, 2322/13, 2000/7 were used in two studies each, respectively; and 910/7, 1033/13, 1223/17, 1787/13, 2010/13, 2023/17, 2083/13 were used in one study each, respectively. All arithmetic tasks lasted 5 min.

#### 4.2.7. Recovery period following TSST

We observed a wide range of resting periods across studies. Sixteen studies used recovery periods that were shorter than 1 h (from 10 to 55 min); seven used over 1-h post testing recovery, and seven used recovery periods lasting over 90 min. On average, recovery periods lasted  $60.29 \pm 7.76$  min.

#### 4.3. Physiological measures

Collection times are preceded by minus '-' or plus '+' signs to indicate whether the physiological measures were collected prior or after TSST exposure, respectively, 0 min corresponding to TSST initiation. Given that studies used multiple collection intervals for the physiological measures, we decided to round up the stated collection times when

**Table 2**  
Procedural timeline- Example of information provided in Appendix C.

| Study                        | Initiation time for each |                             |                | Duration of each study  |                            |               |                             | Termination time for each study |                             |
|------------------------------|--------------------------|-----------------------------|----------------|-------------------------|----------------------------|---------------|-----------------------------|---------------------------------|-----------------------------|
|                              | Initiation time          | End of provided time window | Window (width) | Rest Period before TSST | Recovery period (duration) | TSST Duration | Experiment (Total duration) | Termination time                | End of provided time window |
|                              | Time 1 (24h)             | Time 2 (24h)                | In hours       | Min                     | Min                        | Min           | Min                         | Time 1 (24h)                    | Time 2 (24h)                |
| Abelson et al. (2014)        | 13                       | –                           | –              | 60                      | 75                         | 13            | 148                         | 15.47                           | –                           |
| Bae et al. (2019)            | 12                       | –                           | –              | 210                     | 130                        | 15            | 355                         | 17.92                           | –                           |
| Böbel et al. (2018)          | 13                       | –                           | –              | 60                      | 120                        | 15            | 195                         | 16.25                           | –                           |
| Buchanan et al. (2014)       | 12                       | 16                          | 4              | 30                      | 40                         | 15            | 85                          | 13.42                           | 17.42                       |
| Drake et al. (2017)          | 11                       | 16                          | 5              | 45                      | 40                         | 20            | 105                         | 12.75                           | 17.75                       |
| Erickson et al. (2017)       | 13.5                     | –                           | –              | 60                      | 60                         | 13            | 133                         | 15.72                           | –                           |
| Elzinga & Roelofs (2005)     | 9                        | –                           | –              | 45                      | 50                         | 15            | 110                         | 10.83                           | –                           |
| Fries et al. (2006)          | 15                       | 17.5                        | 2.5            | 180                     | 60                         | 13            | 253                         | 19.22                           | 21.72                       |
| Giles et al. (2014)          | 13                       | 15                          | 2              | 5                       | 20                         | 20            | 45                          | 13.75                           | 15.75                       |
| Gröpel et al. (2018)         | 13                       | 16                          | 3              | 30                      | 60                         | 15            | 105                         | 14.75                           | 17.75                       |
| Het et al. (2015)            | 14                       | 17                          | 3              | 30                      | 25                         | 15            | 70                          | 15.17                           | 18.17                       |
| Inagaki & Eisenberger (2016) | 13.5                     | 16.5                        | 3              | 90                      | 75                         | 15            | 180                         | 16.50                           | 19.5                        |
| Jiang et al. (2017)          | 13.5                     | 18.5                        | 5              | 30                      | 30                         | 15            | 75                          | 14.75                           | 19.75                       |
| Kern et al. (2008)           | 12                       | 16.5                        | 4.5            | 60                      | 90                         | 13            | 163                         | 14.72                           | 19.22                       |
| Klatzkin et al. (2018a)      | 14                       | 17                          | 3              | 10                      | 45                         | 15            | 70                          | 15.17                           | 18.17                       |
| Klatzkin et al. (2018b)      | 16                       | 17                          | 1              | 10                      | 80                         | 15            | 105                         | 17.75                           | 18.75                       |
| Klatzkin et al. (2019)       | 14                       | 16                          | 2              | 10                      | 30                         | 15            | 55                          | 14.92                           | 16.92                       |
| Li et al. (2015)             | 9                        | 12                          | 3              | 70                      | 40                         | 15            | 125                         | 11.08                           | 14.08                       |
| Lupis et al. (2014)          | 14                       | 17                          | 3              | 30                      | 45                         | 15            | 90                          | 15.50                           | 18.50                       |
| Maki et al. (2015)           | 13                       | 17                          | 4              | 60                      | 10                         | 20            | 90                          | 14.50                           | 18.50                       |
| McInnis et al. (2014)        | 13.5                     | 18.5                        | 5              | 30                      | 120                        | 13            | 163                         | 16.22                           | 21.22                       |
| McInnis et al. (2015)        | 13.5                     | 18.5                        | 5              | 30                      | 120                        | 13            | 163                         | 16.22                           | 21.22                       |
| Oswald et al. (2004)         | 12                       | –                           | –              | 90                      | 55                         | 20            | 165                         | 14.75                           | –                           |
| Polheber & Matchock (2014)   | 15                       | 16                          | 1              | 40                      | 30                         | 13            | 83                          | 16.38                           | 17.38                       |
| Reinelt et al. (2019)        | 11.75                    | –                           | –              | 240                     | 115                        | 15            | 370                         | 17.92                           | –                           |
| Shalev et al. (2011)         | 15                       | 18                          | 3              | 20                      | 60                         | 15            | 95                          | 16.58                           | 19.58                       |
| Smith et al. (2016)          | 15                       | 16                          | 1              | 30                      | 30                         | 20            | 80                          | 16.33                           | 17.33                       |
| Souza et al. (2015)          | 13                       | 17                          | 4              | 20                      | 20                         | 20            | 60                          | 14.00                           | 18                          |
| Thomas et al. (2011)         | 16                       | –                           | –              | 60                      | 90                         | 15            | 165                         | 18.75                           | –                           |
| Veer et al. (2011)           | 8.5                      | 10.5                        | 2              | 30                      | 70                         | 20            | 120                         | 10.50                           | 12.5                        |
| Wand et al. (2007)           | 12                       | –                           | –              | 90                      | 65                         | 20            | 175                         | 14.92                           | –                           |
| Wiegand et al. (2018)        | 15                       | –                           | –              | 30                      | 60                         | 20            | 110                         | 16.83                           | –                           |
| Xin et al. (2017)            | 13.5                     | –                           | –              | 30                      | 60                         | 15            | 105                         | 15.25                           | –                           |
| Yao et al. (2016)            | 14                       | 17                          | 3              | 30                      | 30                         | 15            | 75                          | 15.25                           | 18.25                       |
| Zhang et al. (2019)          | 14                       | 17                          | 3              | 30                      | 60                         | 15            | 105                         | 15.75                           | 18.75                       |

needed. For instance, 18 min and 31 min were rounded up to the most proximal number, 20 min and 30 min, respectively.

#### 4.3.1. Heart rate

Twenty-seven studies measured HR and thirty-four different time intervals were reported for recording initiation. On average, researchers

collected nine HR measures ( $9.42 \pm 5.14$ ). The time interval for HR collection varied greatly, HR sampling ranging from an initial collection at  $-70$  min and a final measure at  $+105$  min post initiation of the TSST. Five studies recorded HR throughout the study but did not specify time (i.e., baseline, pre-TSST, TSST, post-TSST).

#### 4.3.2. Cortisol collection

Researchers reported 38 different CORT collection times. On average, studies collected 6 CORT samples ( $6.29 \pm 3.06$ ) over the experimental procedure. However, we found inconsistencies in CORT time sampling procedures across the 35 studies. No consensus emerged on when the initial salivary sample was collected (i.e., initial measurement), the number of collected measures, or the interval between collections. As an illustration, two studies collected the 'initial' CORT level at  $-280$  min, one study at  $-180$  min, one study at  $-70$  min, one study at  $-50$  min, two studies at  $-45$  min, one study at  $-40$  min, six studies at  $-30$  min, four studies at  $-20$  min, two studies at  $-15$ , four studies at  $-10$  min, four studies at  $-5$  min, one study at  $-1$  min, and 6 studies at 0 min (immediately preceding TSST initiation - noted as 0 min sample).

#### 4.3.3. Blood collection

Twelve studies collected blood samples and thirty-three collection times were identified. On average, researchers collected seven samples ( $7.67 \pm 3.52$ ). Similar to salivary cortisol sampling, blood collection showed a wide range of intervals, with no consensus as to a validated sampling procedure. For example, only five studies collected a blood sample at the initiation of the TSST at 0 min, two collected at  $+5$  min, two collected at  $+10$  min, six collected at  $+15$  min, two collected at  $+20$  min, seven collected at  $+25$  min, two collected at  $+30$  min, five collected at  $+35$  min, two collected at  $+40$  min, seven collected at  $+45$  min, three collected at  $+50$  min, one collected at  $+55$ , five at  $+60$  min, two collected at  $+65$  min, two collected at  $+70$  min, three collected at  $+75$  min, and 15 collected blood samples between  $+80$  min and  $+133$  min. As for the method of collection, seven studies reported collecting blood samples using an intravenous catheter, three studies used an indwelling catheter, and two studies used a peripheral venous catheter. In all cases, the technique involved introducing a catheter to collect venous blood samples while minimizing the discomfort and stress effect associated with repeated venepuncture.

## 5. Discussion

We conducted a systematic review to examine the degree to which researchers report various methodological details when using the TSST. We aimed to document reporting practices, identify differences in protocols that may be limiting comparison across studies—and thus—replication efforts, and make recommendations to increase standardization in future research using the TSST. Overall, we found that most researchers: 1) did not provide sufficient details to enable replication, 2) ignored a variety of discrete variables that could influence their findings (e.g., alcohol use, number of judges), and 3) when testing women, often failed to consider menstrual cycle phases, peri- and menopausal periods, and oral contraceptive use. For example, the results in section 4.3.2., Cortisol collection, provide support that researchers have introduced several changes to the initial protocol over time, resulting in a lack of standardization of the actual testing procedure. Initial CORT levels are important because they establish the baseline to which future levels are compared, to determine whether the TSST produces an increase or decrease of HPAA. Therefore, if initial CORT levels are measured at different times and different rates, it is difficult—if not impossible—to compare HPAA responses across time. This comparison difficulty increases if we also take into consideration the variation in testing windows (section 4.2.2). Consequently, in the following sections, we provide a set of guidelines based on the literature and on our findings with the ultimate goal of facilitating strong replication practices by reducing the impact of known confounding variables in HPAA responses to stress.

### 5.1. Researchers should control stress-sensitive factors when using the TSST

#### 5.1.1. Women participants and hormonal status

Guideline 1. Knowing that hormonal fluctuations in women can strongly influence responses to a stressor, researchers should assess and report women's menstrual cycle phase or menopausal state. Moreover, they should include these variables in their statistical analyses (e.g., by creating separate groups or specifying a covariate; Gaffey et al., 2014). Ignoring or failing to report these variables can reduce sound study replication. If groups are too small to control for hormonal status in data analyses, researchers should—at a minimum—report frequencies for this variable so that the information is available for future knowledge synthesis efforts. Researchers could alternatively determine inclusion or exclusion criteria based on menstrual cycle, contraceptive use, pregnancy, or breastfeeding status.

#### 5.1.2. Exclusion criteria

Guideline 2. Researchers using the TSST should assess and report: (a) medication use (type/duration of treatment), (b) mental or psychiatric disorders, (c) history of substance use, nicotine/tobacco and alcohol intake, (d) Body Mass Index range, (e) chronic diseases, (f) poor health/physically unhealthy, (g) cardiovascular diseases, and (h) working night shifts. Reporting decisions about whether to include participants, exclude participants, or implement statistical controls based on these factors will enable clearer interpretation of the effects of any independent variables on HPAA, and will also enhance the validity and replication of studies (Lilienfeld, 2017). Moreover, researchers should clearly indicate how they are measuring each of these factors. For example, how did they assess for current psychological disorders? What chronic diseases were considered?

In 1946, the WHO defined health in their constitution as “the state of complete physical, mental, and social well-being and not merely the absence of disease and infirmity” (WHO, 1946, p. 1). As such, health is often assessed via participant self-report (e.g. any medical conditions, medication, mental health symptoms, etc.). For research with human participants, it is up to the experimenters to decide *a priori* how they will define the health status of participants and, based on their definition, determine eligibility criteria for the study. Again, depending on the specific research goals, inclusion and exclusion criteria may vary. Regardless of how researchers define health for a given study, it is crucial to clearly report the definition in the methodology section, which was not the case for most of the studies reported in this systematic review. In the conclusion (section 6.0), we offer some ideas for how researchers can account for different confounding variables when interpreting their findings, using a guidance document we created. According to participants' responses, researchers can operationalize their own definition of poor health/physically unhealthy and exclude participants from participation or analysis accordingly.

#### 5.1.3. Restriction of activities prior to participation

Guideline 3. Brushing or flossing teeth, smoking, using substances, drinking alcohol or caffeinated beverages, engaging in physical exercise, and eating are all behaviors that, to some extent, have an impact on HPAA. However, we found that researchers do not apply consistent standards for accounting for these behaviors. Therefore, we recommend that researchers systematically ask participants to report engagement in these behaviors prior to participating in the TSST (Stalder et al., 2016). Moreover, researchers should ask all participants to abstain from engaging in these behaviors at least 60 min prior to arriving at the laboratory. For alcohol and substance consumption, this period should be extended to 24 h. Thus, including the rest period prior to beginning the TSST, 90–120 min should elapse before researchers collect physiological measures to assess the impact of the TSST. That time frame will provide sufficient time for CORT secretion to stabilize prior to engaging in the TSST.

Additionally, caffeine withdrawal occurs when abstinence from caffeine leads to primary onset of symptoms 12–24 h after the last consumption, with the peak between 20 and 51 h, and a duration of 2–9 days (Juliano et al., 2004). Therefore, researchers should carefully plan their TSST with consideration of how to restrict caffeine consumption while avoiding a withdrawal state for participants.

## 5.2. Differences in the TSST protocol

### 5.2.1. TSST protocol alterations and selected resting period

Guideline 4. Considering our guideline to restrict the above-mentioned activities for 60 min prior to testing, we recommend that researchers implement at least 30 min resting time for participants upon arriving at the lab and prior to initiating the TSST to allow the individual's physiological responses to stabilize. When participants initially arrive, their physiological activity may vary because of different personal, situational, and environmental factors (e.g., walking to the lab, anxiety about being in a strange environment or participating in research). During this period, researchers can greet participants, complete the consent process, and/or administer preliminary questionnaires—especially those that measure the various factors for which researchers should control (e.g., oral contraceptive use, medical history). We believe the acclimatization period is very important and we recommend a minimum of 30 min be implemented for every study.

### 5.2.2. Period of the day

Guideline 5. Researchers should choose a specific time of day at which participants can begin the study. We suggest avoiding wide time intervals as starting points, given that collected measures might not be comparable, and people's circadian rhythms can influence CORT values. The variability in time windows that we observed may explain some inconsistencies found in the current research. Additionally, researchers should review the literature related to CORT circadian rhythms relevant to their target population (i.e., age and sex) prior to setting a starting point. Finally, researchers should carefully analyze and interpret measures taken beyond 90 min post-TSST to confirm that these values are associated with the experimental conditions and not to extraneous factors.

Lastly, to our knowledge, there are no studies on the impact of waking time and cortisol levels during the day on TSST outcomes. However, in one study, Williams et al. (2005) measured CAR in 32 men and women over 6 days across three conditions: 2 early shift days, 2 day-shift days, and 2 control (leisure) days. They found that waking time had no effect on CAR even when controlling for stress and sleep disturbances. Future studies are warranted to investigate the impact of awakening time on diurnal cortisol levels. At present, researchers may ask participants to wake up at a specific time or within a defined morning time window to minimize influence of this possible confounding variable. However, at this time, no scientific research is available to sustain this claim.

### 5.2.3. Self-report anxiety/stress questionnaires

Guideline 6. Researchers should use a minimum of two different self-report questionnaires to assess participants' subjective anxiety or stress state to ensure accurate estimates of anxiety experiences. Researchers should administer the questionnaires before and after the TSST to measure the impact of the TSST on participants' subjective state anxiety (i.e., change in anxiety due to TSST). Given the large volume and variety of self-report state anxiety measures, researchers should carefully consider their research goals to select the measure(s) that best measures their target variables. For assessing in-the-moment state anxiety, we recommend the State-Trait Anxiety Inventory (STAI), the Positive or Negative Affect Schedule (PANAS), or the Visual Analogue Scale (VAS) because there is evidence that they are sensitive to change, some are available for free, and available in several languages (e.g., STAI in French [Gauthier and Bouchard, 1993]; STAI in Spanish [Virella et al., 1994]).

Furthermore, given evidence that subjective and physiological anxiety measurements often diverge, researchers can better understand the impact of the TSST on different components of anxiety experiences across time by using these questionnaires at multiple timepoints and comparing subjective and physiological responses.

### 5.2.4. Judges' number, gender, and video recorded sessions

Guideline 7. The original TSST included a panel of 3 judges. However, researchers have deviated considerably from this suggestion, using a smaller panel of 2 judges, primarily due to study feasibility. To our knowledge, there are no research findings suggesting significant effects of this procedural difference. However, given the large number of studies that have used two judges and the real need to facilitate study feasibility, adopting a standard of two judges would promote reliable effects that could easily be compared. We suggest using one man and one woman, rather than two same-gendered judges given findings that a mixed panel more effectively induces optimal HPA activation (Allen et al., 2014; Duschesne et al., 2012). Because videorecording is effective in enhancing perceived stress and HPA activation (Kudielka et al., 2007), we recommend telling participants that their performance will be recorded and judges will evaluate their non-verbal communication.

### 5.2.5. Speech and arithmetic tasks

Guideline 8. As shown in Table 1, we found considerable variability in the TSST speech task administration. In a meta-analysis, Goodman et al. (2017) found few differences between preparation times, but did not assess speech length or type. Therefore, in the interest of increasing consistency between studies, researchers should implement the following standard speech task process: 5 min preparation using a piece of paper and pencil to organize their thoughts, 5 min speech (without their notes) about why the participant considers themselves the best candidate for the job of their dreams.

Guideline 9. We observed a wide variety of number combinations. To ensure researchers induce similar stress levels across participants and across studies, we recommend that researchers adopt a standard initial number and subtraction variable. We suggest maintaining the initial arithmetic combination described by Kirschbaum et al. (1993), asking participants to repeatedly subtract 13 from an initial number of 1022. Some systematic reviews/meta-analysis have tried to determine whether variation in the arithmetic task impacts CORT measures, but have not yielded significant findings (see Goodman et al. (2017) for an example), and one could not ascertain whether observed differences were accounted by the selection of the initial number, the subtracting number or the combination/interaction of both. Furthermore, given the variability observed across several variables (e.g., window time, menstrual cycle, exclusion criteria), current knowledge synthesis efforts may not accurately estimate differences based on a single variable because of these multiple confounds.

### 5.2.6. Recovery period

Guideline 10. We must remain cautious about the possible lasting impact of the TSST on HPA activation in certain populations, especially when there is so much variation across variables (e.g., menstrual cycle, age, time windows). Nonetheless, researchers should analyze and interpret with caution any physiological measures taken more than 90 min following TSST completion. To our knowledge, no researchers have demonstrated HPA activation related to TSST exposure above this time interval. In other words, researchers investigating HPA dysfunction, which is not typically observed at delayed intervals should do so through re-exposure to a second TSST session or to another acute stressor. Finally, researchers should include a post-test recovery period lasting at least 60 min to enable them to collect additional physiological measures and examine optimal recovery intervals in diverse populations.

### 5.3. Physiological measures

#### 5.3.1. Heart rate

Guideline 11. HR measures are a practical, non-invasive method to quantify physiological activation of HPA and a way to supplement salivary and/or blood measures (Liew et al., 2016). Based on the reviewed research and considering that HR can be taken continuously throughout the session without disturbing the TSST administration, we suggest that *researchers collect HR measures at time intervals of -30 min, -5 min, 0 min, +5 min, +10 min, +15 min, +20 min, +30 min, +40 min, +50 min and +60 min.* Collecting HR measurements at sufficient time intervals will help better appreciate rapidly occurring changes in participants' physiological arousal.

#### 5.3.2. Cortisol collection

Guideline 12. *We recommend that researchers take 6 CORT samples at -30 min, 0 min, +15 min, +25 min, +35 min and +45 min.* The proposed time intervals have proven effective in the reviewed studies and appear optimal with protocols involving repeated blood or saliva sampling. As researchers know, theory and practice do not always map onto one another. Thus, instead of taking a sample at +10 min, which would occur between the end of the speech task and the beginning of the arithmetic task, we recommend taking it at +15 min. By shifting the interval collection by 5 min, we may not be able to separate contribution of the speech and arithmetic task to cortisol secretion, but this strategy will enable the TSST to be completed without interruption. Indeed, some participants experience difficulty giving a saliva sample (i.e., participants taking medication or older adults; see Dickerson and Kemeny, 2004; Juster et al., 2012). Considering the results in this review and the time profile of endocrine secretions following various moderate psychogenic stressors, including the TSST, the proposed collection enables to cover an appropriate secretion time range and is aligned with results from previous findings (section 2.3.2).

#### 5.3.3. Blood collection

Guideline 13. To our knowledge, no specific recommendations have been proposed for blood collection aimed at CORT measures during the TSST. Therefore, based on CORT secretion profile and numerous studies using TSST induced cortisol detection in saliva samples, we recommend *blood sample collection be performed using the same intervals as for salivary CORT detection.* Given that blood collection remains an invasive procedure, researchers should ensure it is completed by a team member with the appropriate qualifications (e.g., a registered nurse). Researchers should consider whether this additional burden on research feasibility is on par with the desired outcome.

Moreover, they should choose a blood sampling method (e.g., repeated venepuncture, indwelling line) taking into consideration the additional activities in which participants will be involved during the study. For example, if participants undergo a neuropsychological assessment, the presence of a needle inserted on the top of the hand could interfere with full arm dexterity and exert unanticipated effects on any neuropsychological task requiring a manipulation of objects. Therefore, researchers should carefully select the proper method to take blood samples based on the study paradigm. In this context, consulting the WHO guide is a good start (WHO, 2010).

Although we focused primarily on the main physiological measures assessed during the TSST (i.e., CORT), it is worth mentioning that saliva and/or blood samples allow researchers to measure several other endocrine responses. For example, researchers could assess the influence of the main sex hormones in women (i.e., estradiol and progesterone) on study outcomes while other researchers study the impact of the TSST exposure on immune parameters (e.g., cytokines) using collected saliva and/or blood samples (for a review see Allen et al., 2014). These analyses need to be carefully planned in advance for two main reasons: 1) analysis of different endocrine parameters can be very costly, and 2) the amount of saliva and/or blood can differ depending on the number

and/or nature of the endocrine/immune responses assessed. Researchers interested in collecting and analyzing endogenous biomarkers should develop a detailed plan before beginning data collection.

## 6. Conclusion

The TSST has been a true asset in investigating the impact of social stress exposure on various functional outcomes. Nonetheless, in reviewing these individual studies, we noted no systematic processes among researchers across a number of variables and factors. First, there is considerable variability across studies in exclusion criteria, which makes it difficult to replicate or compare studies. Second, we found substantial variability in time intervals implemented in all phases of the TSST, with a notable lack of similarity across articles, and in whether or not researchers considered hormonal status among women. These two major factors may play a primary role in the wide variety of findings in this research area. We found no standardization across studies of optimal times for collection of various physiological measures, complicating future researchers' ability to select time intervals when designing a study. We acknowledge the cost associated with collecting multiple, repeated physiological measures. However, selecting consistent intervals, even if it is not feasible to use all measures at a high frequency, whether due to costs or participants not being suited for repeated sampling, still represents a minimum standard, and an improvement over the current situation. We strongly believe that establishing key time collection intervals can set the stage for improved knowledge synthesis and transfer.

Throughout the years, researchers have implemented many changes to the original protocol. Several studies reported divergent findings, which may be partly attributable to methodological differences. In this review, we noted that only 20% of the articles eligible for this review (i.e., because they reported a sufficient amount of methodological detail) were published before 2014. Thus, the scientific field may be becoming more aware of this methodological variability, and prioritizing high reporting standards to facilitate research replication (Lilienfeld, 2017).

Given that our findings are similar to those reported in past reviews (Labuschagne et al., 2019; Allen et al., 2014, 2017; Goodman et al., 2017), our work provides additional support to past criticisms of TSST variations using best practices for knowledge syntheses. Using a systematic approach, our review validates the recommendations from previous work, and through rigorous assessment of the controlled factors across studies, enabled us to provide informed guidelines to be used in future research.

Since the inception of our review, we realized that researchers devoted to studying stress have made significant gains to increase our understanding of the consequences and influence of stress in our daily lives. Nevertheless, even one of the most rigorous paradigms like the TSST is not immune to methodological drift following new discoveries. As such, we used this systematic review to develop a TSST researcher's guide—a shortlist of elements researchers should consider prior to using the TSST. Our ultimate goal is to facilitate standardization of the research protocol via a detailed methodology, including a proposed timeline for the collected physiological measures (appendix A) and a background questionnaire (appendix B) to guide researchers in considering several factors that could influence findings. Specifically, we aim to raise awareness of the different variables that can influence TSST outcomes. However, specific research objectives and resource limitations will undoubtedly require adaptations in applying the proposed recommendations. Rather than a strict set of implementation rules, we conceptualize Appendix B as a flexible guide to aid in controlling potentially influential or confounding variables by taking these factors into account when designing the study, collecting a large amount of data, analyzing data, and interpreting their findings. To our knowledge, there are no comparable documents available in the literature. We hope that researchers will use these resources to facilitate data collection, research replication, data analysis, and eventually improve study

comparisons and knowledge synthesis (Stark, 2018).

### CRedit authorship contribution statement

**N.F. Narvaez Linares:** Conceptualization, Methodology, Formal analysis, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization, Project administration. **V. Charron:** Conceptualization, Investigation, Data curation, Writing - original draft, Writing - review & editing, Visualization. **A.J. Ouimet:** Conceptualization, Writing - review & editing, Visualization. **P.R. Labelle:** Methodology, Investigation, Writing - review & editing. **H. Plamondon:** Conceptualization, Writing - original draft, Visualization, Supervision.

### Acknowledgements

We would like to thank Paulette Ewalefo, Matilda Berr, Valérie Ranger and Aaron Boughen for their assistance in data coding and database data entry. We declare no conflict of interest. NFNL and VC are supported by scholarships from the Québec Research Funds in Nature and Technologies (FRQNT). This work was supported by a discovery grant to HP from the National Science and Engineering Research Council (NSERC) of Canada (RG203596-13).

### Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jynstr.2020.100235>.

### References

- \* Abelson, J.L., Erokkson, T.M., Mayer, S.E., Crocker, J., Briggs, H., Lopez-Duran, N.L., Liberson, I., 2014. Brief cognitive intervention can modulate neuroendocrine stress responses to the Trier Social Stress Test: buffering effects of a compassionate goal orientation. *Psychoneuroendocrinology* 44, 60–70. <https://doi.org/10.1016/j.psyneuen.2014.02.016>.
- Allen, A.P., Kennedy, P.J., Cryan, J.F., Timothy, G.D., Gerard, C., 2014. Biological and psychological markers of stress in humans: focus on the trier social stress test. *Neuroscience & Biobehavioral Reviews* 38, 94–124. <https://doi.org/10.1016/j.neubiorev.2013.11.005>.
- Agarwal, S., Prescittutti, A., Verma, J., Pavol, M.A., Anbarasan, D., Brodie, D., Stern, Y., 2018. Women have worse cognitive, functional, and psychiatric outcomes at hospital discharge after cardiac arrest. *Resuscitation* 125, 12–15. <https://doi.org/10.1016/j.resuscitation.2018.01.036>.
- Allen, A.P., Kennedy, P.J., Dockray, S., Cryan, J.F., Dinan, T.G., Clarke, G., 2017. The trier social stress test: principles and practice. *Neurobiology of Stress* 6, 113–126. <https://doi.org/10.1016/j.jynstr.2016.11.001>.
- Badrick, E., Bobak, M., Britton, A., Kirschbaum, C., Marmot, M., Kumari, M., 2008. The relationship between alcohol consumption and cortisol secretion in an aging cohort. *J. Clin. Endocrinol. Metabol.* 93 (3), 750–757. <https://doi.org/10.1210/ajc.2007-0737>.
- Badrick, E., Kirschbaum, C., Kumari, M., 2007. The relationship between smoking status and cortisol secretion. *J. Clin. Endocrinol. Metabol.* 92 (3), 819–824. <https://doi.org/10.1210/jc.2006-2155>.
- \* Bae, Y.J., Reinelt, J., Netto, J., Uhlig, M., Willenberg, A., Ceglarek, U., Kratzsch, J., 2019. Salivary cortisone, as a biomarker for psychosocial stress, is associated with state anxiety and heart rate. *Psychoneuroendocrinology* 101, 35–41. <https://doi.org/10.1016/j.psyneuen.2018.10.015>.
- Berger, M., Bossert, S., Krieg, J.C., Dirllich, G., Ettmeir, W., Schreiber, W., von Zerssen, D., 1987. Interindividual differences in the susceptibility of the cortisol system: an important factor for the degree of hypercortisolism in stress situations? *Biol. Psychiatr.* 22 (11), 1327–1339. [https://doi.org/10.1016/0006-3223\(87\)90067-9](https://doi.org/10.1016/0006-3223(87)90067-9).
- Biondi, M., Picardi, A., 1999. Psychological stress and neuroendocrine function in humans: the last two decades of research. *Psychother. Psychosom.* 68 (3), 114–150. <https://doi.org/10.1159/000012323>.
- Birkett, M., 2011. The trier social stress test protocol for inducing psychological stress. *JoVE* 56, e3238. <https://doi.org/10.3791/3238>.
- \* Böbel, T.S., Hackl, S.B., Langgartner, D., Jarczok, M.N., Rohleder, N., Rook, G.A., Reber, S.O., 2018. Less immune activation following social stress in rural vs. urban participants raised with regular or no animal contact, respectively. *Proc. Natl. Acad. Sci. U.S.A.* 115 (20), 5259–5264. <https://doi.org/10.1073/pnas.1719866115>.
- Boivin, D.B., Boudreau, P., 2014. Impacts of shift work on sleep and circadian rhythms. *Pathol. Biol.* 62 (5), 292–301. <https://doi.org/10.1016/j.patbio.2014.08.001>.
- Branchard, B., Deb-Rinker, P., Dubois, A., Lapointe, P., O'Donnel, S., Pelletier, L., Williams, G., 2018. At-a-glance – how Healthy are Canadians? A brief update. *Health Promotion and Chronic Disease Prevention in Canada* 38 (10). <https://doi.org/10.24095/hpcdp.38.10.05>.
- Brosschot, J.F., Geurts, S.A., Kruizinga, I., Radstaak, M., Verkuil, B., Quirin, M., Kompier, M.A., 2014. Does unconscious stress play a role in prolonged cardiovascular stress recovery? *Stress Health* 30 (3), 179–187. <https://doi.org/10.1002/smi.2590>.
- Buchanan, T.W., Laures-Gore, J.S., Duff, M.C., 2014. Acute stress reduces speech fluency. *Biol. Psychol.* 97, 60–66. <https://doi.org/10.1016/j.biopsycho.2014.02.005>.
- Burke, T.M., Markwald, R.R., McHill, A.W., Chinoy, E.D., Snider, J.A., Bessman, S.C., Jung, C.M., O'Neill, J.S., Wright Jr., K.P., 2016. Effects of caffeine on the human circadian clock *in vivo* and *in vitro*. *Science Translational Medicine* 7 (305), 305ra146. <https://doi.org/10.1126/scitranslmed.aac5125>.
- Buske-Kirschbaum, A., Jobst, S., Wustmans, A., Kirschbaum, C., Rauh, W., Hellhammer, D., 1997. Attenuated free cortisol response to psychosocial stress in children with atopic dermatitis. *Psychosomatic Medicine* 59 (4), 419–426. <https://doi.org/10.1097/00006842-199707000-00012>.
- Campbell-Sills, L., Barlow, D.H., Brown, T.A., Hofmann, S.G., 2006. Effects of suppression and acceptance on emotional responses of individuals with anxiety and mood disorders. *Behav. Res. Ther.* 44 (9), 1251–1263. <https://doi.org/10.1016/j.brat.2005.10.001>.
- Caviola, S., Carey, E., Mammarella, I.C., Szucs, D., 2017. Stress, time pressure, strategy selection and math anxiety in mathematics: a review of the literature. *Front. Psychol.* 8 (1488), 1–13. <https://doi.org/10.3389/fpsyg.2017.01488>.
- Charles, L.E., Fekedulegn, D., Burchfiel, C.M., Hartley, T.A., Andrew, M.E., Violanti, J. M., Miller, D.B., 2016. Shiftwork and diurnal salivary cortisol patterns among police officers. *J. Occup. Environ. Med.* 58 (6), 542–549. <https://doi.org/10.1097/JOM.0000000000000729>.
- Cohen, S., Kamarck, T., Mermelstein, R., 1983. A global measure of perceived stress. *J. Health Soc. Behav.* 24 (4), 385–396. <https://doi.org/10.2307/2136404>.
- Cohen, A., Colodner, R., Masalha, R., Haimov, I., 2019. The relationship between tobacco smoking, cortisol secretion, and sleep continuity. *Subst. Use Misuse* 54 (10), 1705–1714. <https://doi.org/10.1080/10826084.2019.1608250>.
- Covidence systematic review software, Veritas Health Innovation, Melbourne, Australia. Available at [www.covidence.org](http://www.covidence.org).
- Cservenka, A., Lahanas, S., Dotson-Bossert, J., 2018. Marijuana use and hypothalamic-pituitary-adrenal Axis functioning in humans. *Front. Psychiatr.* 9, 472. <https://doi.org/10.3389/fpsyg.2018.00472>.
- Deak, T., Quinn, M., Cidlowski, J.A., Victoria, N.C., Murphy, A.Z., Sheridan, J.F., 2016. Neuroimmune mechanisms of stress: sex differences, developmental plasticity, and implications for pharmacotherapy of stress-related disease. *Stress* 18 (4), 367–380. <https://doi.org/10.3109/10253890.2015.1053451>.
- De Los Reyes, A., Aldao, A., Thomas, S.A., Daruwala, S., Swan, A.J., Van Wie, M., Lachner, W.V., 2012. Adolescent self-reports of social anxiety: can they disagree with objective psychophysiological measures and still be valid? *J. Psychopathol. Behav. Assess.* 34, 308–322. <https://doi.org/10.1007/s10862-012-9289-2>.
- Dickerson, S.S., Kemeny, M.E., 2004. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Add this article to my list* 130 (3), 355–391. <https://doi.org/10.1037/0033-2909.130.3.355>.
- Dis-Chaves, Y., Gil-Lozano, M., Toba, L., Fandino, J., Ogando, H., Gonzalez-Matias, L.C., Mallo, F., 2016. Stressing diabetes? The hidden links between insulinotropic peptides and the HPA axis. *J. Endocrinol.* 230 (2), R77–R94. <https://doi.org/10.1530/JOE-16-0118>.
- Dowker, A., Sarkar, A., Looi, C.Y., 2016. Mathematics anxiety: what have we learned in 60 Years? *Front. Psychol.* 7, 1–16. <https://doi.org/10.3389/fpsyg.2016.00508>.
- \* Drake, C.L., Cheng, P., Almeida, D.M., Roth, T., 2017. Familial risk for insomnia is associated with abnormal cortisol response to stress. *Sleep* 40 (1), zsx143. <https://doi.org/10.1093/sleep/zsx143>.
- Duschesne, A., Tessera, E., Dedovic, K., Engert, V., Pruessner, J.C., 2012. Effects of panel sex composition on the physiological stress responses to psychosocial stress in healthy young men and women. *Biol. Psychol.* 89 (1), 99–106. <https://doi.org/10.1016/j.biopsycho.2011.09.009>.
- \* Elzinga, B.M., Roelofs, K., 2005. Cortisol-induced impairments of working memory require acute sympathetic activation. *Behav. Neurosci.* 119 (1), 98–103. <https://doi.org/10.1037/0735-7044.119.1.98>.
- \* Erickson, T.M., Mayer, S.E., Lopez-Duran, N.L., Scarsella, G.M., McGuire, A.P., Croker, J., Abelson, 2017. Mediators of compassionate goal intervention effects on human neuroendocrine responses to the Trier Social Stress Test. *Stress* 20 (6), 533–540. <https://doi.org/10.1080/10253890.2017.1368489>.
- Fang, C.Y., Egleston, B.L., Manzur, A.M., Townsend, R.R., Stanczyk, F.Z., Spiegel, D., Dorgan, J.F., 2014. Psychological reactivity to laboratory stress is associated with hormonal responses in postmenopausal women. *J. Int. Med. Res.* 42 (2), 444–456. <https://doi.org/10.1177/0300060513504696>.
- Felmingahm, K.L., Fong, W.C., Bryant, R.A., 2012. The impact of progesterone on memory consolidation of threatening images in women. *Psychoneuroendocrinology* 37 (11), 1896–1900. <https://doi.org/10.1016/j.psyneuen.2012.03.026>.
- Flegal, K.M., Carroll, M.D., Kit, B.K., Ogden, C.L., 2012. Prevalence of obesity and trends in the distribution of body mass index among US adults, 1999–2010. *J. Am. Med. Assoc.* 307 (5), 491–497. <https://doi.org/10.1001/jama.2012.39>.
- \* Fries, E., Hellhammer, D.H., Hellhammer, J., 2006. Attenuation of the hypothalamic-pituitary-adrenal axis responsivity to the trier social stress test by the benzodiazepine alprazolam. *Psychoneuroendocrinology* 31 (10), 1278–1288. <https://doi.org/10.1016/j.psyneuen.2006.09.009>.
- Gaffey, A.E., Wirth, M.M., Hoks, R.M., Jahn, A.L., Abercrombie, H.C., 2014. Circulating cortisol levels after exogenous cortisol administration are higher in women using hormonal contraceptives: data from two preliminary studies. *Stress* 17 (4), 314–320. <https://doi.org/10.3109/10253890.2014.919474>.

- Garcia-Leal, C., Graef, F.G., Del-Ben, C.M., 2014. Experimental public speaking: contributions to the understanding of the serotonergic modulation of fear. *Neurosci. Biobehav. Rev.* 46 (3), 407–417. <https://doi.org/10.1016/j.neubiorev.2014.09.011>.
- Garde, A.H., Persson, R., Hansen, A.M., Osterberg, K., Ørbaek, P., Eek, F., Karlson, B., 2009. Effects of lifestyle factors on concentrations of salivary cortisol in healthy individuals. *Scand. J. Clin. Lab. Investig.* 69 (2), 242–250. <https://doi.org/10.1080/00365510802483708>.
- Gauthier, J., Bouchard, S., 1993. Adaptation canadienne-française de la forme révisée du State-Trait Anxiety Inventory de Spielberger [A French-Canadian adaptation of the revised version of Spielberger's State-Trait Anxiety Inventory]. *Can. J. Behav. Sci./Rev. Can. Sci. Comport.* 25 (4), 559–578. <https://doi.org/10.1037/h0078881>.
- \* Giles, G.E., Mahoney, C.R., Brunyè, T.T., Taylor, H.A., Kanarek, R.B., 2014. Stress effects on mood, HPA Axis, and autonomic response: comparison of three psychosocial stress paradigms. *PLoS One* 9 (12), e113618. <https://doi.org/10.1371/journal.pone.0113618>.
- Gonnissen, H.K., Mazuy, C., Rutters, F., Martens, E.A., Adam, T.C., Westerterp-Plantenga, M.S., 2013. Sleep architecture when sleeping at an unusual circadian time and associations with insulin sensitivity. *PLoS One* 8 (8), e72877. <https://doi.org/10.1371/journal.pone.0072877>.
- Goodman, W.K., Janson, J., Wolf, J.M., 2017. Meta-analytical assessment of the effects of protocol variations on cortisol responses to the Trier Social Stress Test. *Psychoneuroendocrinology* 80, 26–35. <https://doi.org/10.1016/j.psyneuen.2017.02.030>.
- Government of Canada, 2018. *Publications – Healthy Living*. Retrieved 02 18, 2019, from Tackling Obesity in Canada: Obesity and Excess Weight Rates in Canadian Adults, 01 25. <https://www.canada.ca/en/public-health/services/publications/healthy-living/obesity-excess-weight-rates-canadian-adults.html>.
- Granger, D.A., Hibel, L.C., Fortunato, C.K., Kapelushki, C.H., 2009. Medication effects on salivary cortisol: Tactics and strategy to minimize impact in behavioral and developmental science. *Psychoneuroendocrinology* 34 (10), 1437–1448. <https://doi.org/10.1016/j.psyneuen.2009.06.017>.
- \* Gröpel, P., Urner, M., Pruessner, J.C., Quirin, M., 2018. Endurance- and resistance-trained men exhibit lower cardiovascular responses to psychosocial stress than untrained men. *Front. Psychol.* 9, 852. <https://doi.org/10.3389/fpsyg.2018.00852>.
- Gunn, C., Mackus, M., Griffin, C., Munafo, M.R., Adams, S., 2018. A systematic review of the next-day effects of heavy alcohol consumption on cognitive performance. *Addiction* 113 (12), 2182–2193. <https://doi.org/10.1111/add.14404>.
- Hellhammer, D.H., Wust, S., Kudielka, B.M., 2009. Salivary cortisol as a biomarker in stress research. *Psychoneuroendocrinology* 34 (2), 163–171. <https://doi.org/10.1016/j.psyneuen.2008.10.026>.
- Hellhammer, J., Schubert, M., 2012. The physiological response to Trier Social Stress Test relates to subjective measures of stress during but not before or after the test. *Psychoneuroendocrinology* 37 (1), 119–124. <https://doi.org/10.1016/j.psyneuen.2011.05.012>.
- Heller, G.Z., Manuguerra, M., Chow, R., 2016. How to analyze the visual Analogue scale: myths, truths and clinical relevance. *Scandinavian Journal of Pain* 13 (1), 67–75. <https://doi.org/10.1016/j.jpain.2016.06.012>.
- Herhaus, B., Petrowski, K., 2018. Cortisol stress reactivity to the trier social stress test in obese adults. *Obesity Facts* 11 (6), 491–500. <https://doi.org/10.1159/000493533>.
- \* Het, S., Vocks, S., Wolf, J.M., Hammelstein, P., Herpertz, S., Wolf, O.T., 2015. Blunted neuroendocrine stress reactivity in young women with eating disorders. *J. Psychosom. Res.* 78 (3), 60–267. <https://doi.org/10.1016/j.jpsychores.2014.11.001>.
- Houtepen, L.C., Boks, M.P., Kahn, R.S., Joëls, M., Vinkers, C.H., 2015. Antipsychotic use is associated with a blunted cortisol stress response: a study in euthymic bipolar disorder patients and their unaffected siblings. *Eur. Neuropsychopharmacol.* 25 (1), 77–84. <https://doi.org/10.1016/j.euroneuro.2014.10.005>.
- Hvidberg, M.F., Johnsen, S.P., Davidsen, M., Ehlers, L., 2020. A nationwide study of prevalence rates and characteristics of 199 chronic conditions in Denmark. *Pharmacoeconomics - Open* 4, 361–380. <https://doi.org/10.1007/s41669-019-0167-7>.
- Ibrahim, N.K., Ifikhar, R., 2014. Energy drinks: Getting wings but at what health cost? *Pakistan Journal Medical Sciences* 30 (6), 1415–1419. <https://doi.org/10.12669/pjms.306.5396>.
- \* Inagaki, T.K., Eisenberger, N.I., 2016. Giving support to others reduces sympathetic nervous system-related responses to stress. *Psychophysiology* 53 (4), 427–435. <https://doi.org/10.1111/psyp.12578>.
- Incollingo Rodrigues, A.C., Epel, E.S., White, M.L., Standen, E.C., Seckl, J.R., Tomiyama, A.J., 2015. Hypothalamic-pituitary-adrenal axis dysregulation and cortisol activity in obesity: a systematic review. *Psychoneuroendocrinology* 62, 301–318. <https://doi.org/10.1016/j.psyneuen.2015.08.014>.
- Jacobson, L., 2014. Hypothalamic-pituitary-Adrenocortical Axis: neuropsychiatric aspects. *Comprehensive Physiology* 4 (2), 715–738. <https://doi.org/10.1002/cphy.c130036>.
- \* Jiang, C., Buchanan, T.W., Yao, Z., Zhang, K., Wu, J., Zhang, L., 2017. Acute psychological stress disrupts attentional bias to threat-related stimuli. *Sci. Rep.* 7 (1), 14607. <https://doi.org/10.1038/s41598-017-14138-w>.
- Johnston, C., Campebell-Yeo, M., Disher, T., Benoit, B., Fernandes, A., Streiner, D., Zee, R., 2017. Skin-to-skin care for procedural pain in neonates. *Cochrane Systematic Review* 2, CD008435. <https://doi.org/10.1002/14651858.CD008435.pub3>, 2017.
- Jönsson, P., Wallergård, M., Österberg, K., Hansen, Å.M., Johansson, G., Karlson, B., 2010. Cardiovascular and cortisol reactivity and habituation to a virtual reality version of the Trier Social Stress Test: a pilot study. *Psychoneuroendocrinology* 35 (9), 1397–1403. <https://doi.org/10.1016/j.psyneuen.2010.04.003>.
- Juliano, L.M., Griffiths, R.R., 2004. A critical review of caffeine withdrawal: Empirical validation of symptoms and signs, incidence, severity, and associated features. *Psychopharmacology* 176 (1), 1–29. <https://doi.org/10.1007/s00213-004-2000-x>.
- Juster, R.-P., Perna, A., Marin, M.-F., Sindi, S., Lupien, S.J., 2012. Timing is everything: anticipatory stress dynamics among cortisol and blood pressure reactivity and recovery in healthy adults. *Stress* 15 (6), 569–577. <https://doi.org/10.3109/10253890.2012.661494>.
- Kajantie, E., Phillips, D.I.W., 2006. The effects of sex and hormonal status on the physiological response to acute psychosocial stress. *Psychoneuroendocrinology* 31 (2), 151–178. <https://doi.org/10.1016/j.psyneuen.2005.07.002>.
- Karlson, B., Eek, F., Hansen, A.M., Garde, A.H., Ørbaek, P., 2011. Cortisol variability and self-reports in the measurement of work-related stress. *Stress Health* 27 (7), e11–24. <https://doi.org/10.1002/smi.1330>.
- \* Kern, S., Oakes, T.E., Stone, C.K., McAuliff, E.M., Kirschbaum, C., Davidson, R.J., 2008. Glucose metabolic changes in the prefrontal cortex are associated with HPA axis response to a psychosocial stressor. *Psychoneuroendocrinology* 33 (4), 517–529. <https://doi.org/10.1016/j.psyneuen.2008.01.010>.
- Kirschbaum, C., Kudielka, B.M., Gaab, J., Schommer, N.C., Hellhammer, D.H., 1999. Impact of gender, menstrual cycle phase, and oral contraceptives on the activity of the hypothalamus-pituitary-adrenal axis. *Psychosom. Med.* 62 (2), 154–162.
- 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. <https://doi.org/10.1159/000119004>.
- \* Klatzkin, R.R., Baldassaro, A., Hayden, E., 2018a. The impact of chronic stress on the predictors of acute stress-induced eating in women. *Appetite* 123, 345–351. <https://doi.org/10.1016/j.appet.2018.01.007>.
- \* Klatzkin, R.R., Baldassaro, A., Rashid, S., 2019. Physiological responses to acute stress and the drive to eat: the impact of perceived life stress. *Appetite* 133, 393–399. <https://doi.org/10.1016/j.appet.2018.11.019>.
- \* Klatzkin, R.R., Gaffney, S., Cyrus, K., Bigus, E., Brownley, K.A., 2018b. Stress-induced eating in women with binge-eating disorder and obesity. *Biol. Psychol.* 131, 96–106. <https://doi.org/10.1016/j.biopsycho.2016>.
- Kudielka, B.M., Hellhammer, D.H., Kirschbaum, C., 2007. Ten years of research with the trier social stress test (TSST) - revisited. In: Harmon-Jones, E., Winkelman, P. (Eds.), *Social Neuroscience: Integrating Biological and Psychological Explanations of Behaviour*. Guilford Press, New-York, pp. 56–83.
- Kudielka, B.M., Hellhammer, D.H., Wust, S., 2009. Why do we respond so differently? Reviewing determinants of human salivary cortisol responses to challenge. *Psychoneuroendocrinology* 34 (1), 2–18. <https://doi.org/10.1016/j.psyneuen.2008.10.004>.
- Kudielka, B.M., Schommer, N.C., Hellhammer, D.H., Kirschbaum, C., 2004. Acute HPA axis responses, heart rate, and mood changes to psychosocial stress (TSST) in humans at different times of day. *Psychoneuroendocrinology* 29 (8), 983–992. <https://doi.org/10.1016/j.psyneuen.2003.08.009>.
- Kudielka, B., Gierens, A., Hellhammer, D.H., Wust, S., Schlotz, W., 2012. Salivary cortisol in ambulatory assessment-some dos, some don'ts, and some open questions. *Psychosom. Med.* 74 (4), 418–431. <https://doi.org/10.1097/PSY.0b013e31825434c7>.
- Labuschagne, I., Grace, C., Rendell, P., Terret, G., Heinrichs, M., 2019. An introductory guide to conducting the trier social stress test. *Neurosci. Biobehav. Rev.* 107, 686–695. <https://doi.org/10.1016/j.neubiorev.2019.09.032>.
- \* Li, S., Weerda, R., Milde, C., Wolf, O.T., Thiel, C.M., 2015. ADRA2B genotype differentially modulates stress-induced neural activity in the amygdala and hippocampus during emotional memory retrieval. *Psychopharmacology* 232 (4), 755–764. <https://doi.org/10.1007/s00213-014-3710-3>.
- Lilienfeld, S.O., 2017. Psychology's replication crisis and the grant culture: righting the ship. *Perspect. Psychol. Sci.* 12 (4), 660–664. <https://doi.org/10.1177/1745691616687745>.
- Loue, S., Sajatovic, M., 2008. Research with severely mentally ill Latinas: successful recruitment and retention strategies. *J. Immigr. Minority Health* 10 (2), 145–153. <https://doi.org/10.1007/s10903-007-9063-9>.
- Liu, J.J.W., Ein, N., Peck, K., Huang, V., Pruessner, J.C., Vickers, K., 2017. Sex differences in salivary cortisol reactivity to the Trier Social Stress Test (TSST): a meta-analysis. *Psychoneuroendocrinology* 82, 26–37. <https://doi.org/10.1016/j.psyneuen.2017.04.007>.
- Liew, W.S., Seera, L., Loo, C.K., Lim, E., Kubota, N., 2016. Classifying stress from heart rate variability using salivary biomarkers as reference. *IEEE Transactions on Neural Networks and Learning Systems* 27 (10), 2035–2046. <https://doi.org/10.1109/TNNLS.2015.2468721>.
- \* Lupis, S.B., Lerman, M., Wolf, J.M., 2014. Anger responses to psychosocial stress predict heart rate and cortisol stress responses in men but not women. *Psychoneuroendocrinology* 49, 84–95. <https://doi.org/10.1016/j.psyneuen.2014.07.004>.
- Lund, H.G., Reider, B.D., Whiting, A.B., Prichard, J.R., 2010. Sleep patterns and predictors of disturbed sleep in a large population of college students. *J. Adolesc. Health* 46 (2), 124–132. <https://doi.org/10.1016/j.jadohealth.2009.06.016>.
- \* Maki, P.M., Mordecai, K.L., Rubin, L.H., Sundermann, S.A., Eatough, E., Drogos, L., 2015. Menstrual cycle effects on cortisol responsivity and emotional retrieval following a psychosocial stressor. *Horm. Behav.* 74, 201–208. <https://doi.org/10.1016/j.yhbeh.2015.06.023>.
- Malinauskas, B.M., Aebly, V.G., Overton, R.F., Carpenter-Aeby, T., Barber-Heidal, K., 2007. A survey of energy drink consumption patterns among college students. *Nutr. J.* 6 (35) <https://doi.org/10.1186/1475-2891-6-35>.
- Manetti, L., Cavagnini, F., Martino, E., Ambrogio, A., 2014. Effects of cocaine on the hypothalamic-pituitary-adrenal axis. *J. Endocrinol. Invest.* 37 (8), 701–708. <https://doi.org/10.1007/s40618-014-0091-8>.

- Matsuda, S., Yamaguchi, T., Okada, K., Gotouda, A., Mikami, S., 2012. Day-to-day variations in salivary cortisol measurements. *Journal of Prosthodontic Research* 56 (1), 37–41. <https://doi.org/10.1016/j.jpor.2011.04.004>.
- Matura, L.A., Malone, A., Jaime-Lara, R., 2018. A systematic review of biological mechanisms of fatigue in chronic illness. *Biological Research For Nursing* 20 (4), 140–421. <https://doi.org/10.1177/1099800418764326>.
- \* McClinnis, C.M., Thoma, M.V., Gianferante, D., Hanlin, L., Chen, X., Brenies, J., Rohleder, N., 2014. Measures of adiposity predict interleukin-6 responses to repeated psychosocial stress. *Bren, Behavior, and Immunity* 42, 33–40. <https://doi.org/10.1016/j.bbi.2014.07.018>.
- McInnis, C.M., Wang, D., Gianferante, D., Hanlin, L., Chen, X., Thoma, M.V., Rohleder, N., 2015. Response and habituation of pro- and anti-inflammatory gene expression to repeated acute stress. *Brain Behav. Immun.* 46, 237–248. <https://doi.org/10.1016/j.bbi.2015.02.006>.
- Mental Health Commission of Canada, 2013. Canadian Mental Health Association. Retrieved 11 16, 2018, from Fast Facts about Mental Illness: <https://cmha.ca/abo-ut-cmha/fast-facts-about-mental-illness>.
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Prisma-P Group, ., 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Syst. Rev.* 4 (1), 1. <https://doi.org/10.1186/2046-4053-4-1>.
- Monteleone, A.M., Monteleone, P., Serino, I., Amodio, R., Monaco, F., Maj, M., 2016. Underweight subjects with anorexia nervosa have an enhanced salivary cortisol response not seen in weight restored subjects with anorexia nervosa. *Psychoneuroendocrinology* 70, 118–121. <https://doi.org/10.1016/j.psyneuen.2016.05.004>.
- Mustonen, P., Karlsson, L., Scheinin, N.M., Kortelasma, S., Coimbra, B., Rodrigues, A.J., Karlsson, H., 2018. Hair cortisol concentration (HCC) as a measure for prenatal psychological distress — a systematic review. *Psychoneuroendocrinology* 92, 21–28. <https://doi.org/10.1016/j.psyneuen.2018.03.019>.
- Odeniyi, I.A., Fasanmade, O.A., Ogbera, A.O., Ohwovoriole, A.E., 2015. Body mass index and its effect on serum cortisol level. *Niger. J. Clin. Pract.* 18 (2), 194–197. <https://doi.org/10.4103/1119-3077.151040>.
- \* Oswald, L.M., Mathena, J.R., Wand, G.S., 2004. Comparison of HPA axis hormonal responses to naloxone vs psychologically-induced stress. *Psychoneuroendocrinology* 29 (3), 371–388. [https://doi.org/10.1016/S0306-4530\(03\)00048-9](https://doi.org/10.1016/S0306-4530(03)00048-9).
- Parrot, A.C., Montgomery, C., Wetherell, M.A., Downey, L.A., Stough, C., Scholey, A.B., 2014. MDMA, cortisol, and heightened stress in recreational ecstasy users. *Behav. Pharmacol.* 25 (5–6), 458–472. <https://doi.org/10.1097/FBP.0000000000000060>.
- Pasquali, R., Vicennati, V., Cacciari, M., Pagotto, U., 2006. The hypothalamic-pituitary-adrenal axis activity in obesity and the metabolic syndrome, 1083 (1), 111–128. <https://doi.org/10.1196/annals.1367.009>.
- Patz, M.D., Day, H.E., Burrow, A., Campeau, S., 2006. Modulation of the hypothalamic-pituitary-adrenocortical axis by caffeine. *Psychoneuroendocrinology* 31 (4), 493–500. <https://doi.org/10.1016/j.psyneuen.2005.11.008>.
- \* Polheber, J.P., Matchock, R.L., 2014. The presence of a dog attenuates cortisol and heart rate in the Trier Social Stress Test compared to human friends. *J. Behav. Med.* 37 (5), 860–867. <https://doi.org/10.1007/s10865-013-9546-1>.
- Puigcerver, A., Martinez-Selva, J.M., Garcia-Sanchez, F.A., Gomez Amor, J., 1989. Individual differences in psychophysiological and subjective correlates of speech anxiety. *J. Psychophysiol.* 3 (1), 75–81.
- Rahmani, Z.A., Abdullah, N., Singh, R., Sosroseno, W., 2010. Effect of acute exercise on the levels of salivary cortisol, tumor necrosis factor-alpha and nitric oxide. *J. Oral Sci.* 52 (1), 133–136.
- \* Reinelt, J., Uhlig, M., Muller, K., Lauckner, M.E., Kumral, D., Schaare, H.L., Gaebler, M., 2019. Acute psychosocial stress alters thalamic network centrality. *Neuroimage* 199, 680–690. <https://doi.org/10.1016/j.neuroimage.2019.06.005>.
- Roche, D.J., King, A.C., Cohonn, A.J., Lovallo, W.R., 2013. Hormonal contraceptive use diminishes salivary cortisol response to psychosocial stress and naltrexone in healthy women. *Pharmacol. Biochem. Behav.* 109, 84–90. <https://doi.org/10.1016/j.pbb.2013.05.007>.
- Rohleder, N., Kirschbaum, C., 2006. The hypothalamic-pituitary-adrenal (HPA) axis in habitual smokers. *Int. J. Psychophysiol.* 59 (3), 236–243. <https://doi.org/10.1016/j.ijpsycho.2005.10.012>.
- Rotermann, M., Dunn, S., Black, A., 2015. Oral contraceptive use among women aged 15 to 49: results from the Canadian Health Measures Survey. *Health Rep.* 26 (10), 21–28.
- Rubinow, D.R., Crowder, V.L., Schmidt, P.J., Rubinow, K.B., 2012. *Psychoneuroendocrinology*. In: Fink, G., Levine, J.E., Plaff, D.W., ScienceDirect (Eds.), *Handbook of Neuroendocrinology*. Academic Press, UK, pp. 779–790. <https://doi.org/10.1016/C2009-0-04284-6>.
- Ryan, H., Troscclair, A., Gfroerer, J., 2012. Adult current smoking: differences in definitions and prevalence estimates—NHIS and NSDUH, 2008. *Journal of Environmental and Public Health*. <https://doi.org/10.1155/2012/918368>. 2012: 918368, 1–11.
- Schorr, M., Lawson, E.A., Ditchel, L.E., Klibanski, A., Miller, K.K., 2015. Cortisol measures across the weight spectrum. *J. Clin. Endocrinol. Metabol.* 100 (9), 3313–3321. <https://doi.org/10.1210/JC.2015-2078>.
- Scott, S.B., Graham-England, J.E., Engeland, C.G., Smyth, J.M., Almeida, D.M., Katz, M. J., Sliwinski, M.J., 2015. The effects of stress on cognitive aging, physiology and emotion (ESCAPE) Project. *BioMed Central Psychiatry* 15, 146. <https://doi.org/10.1186/s12888-015-0497-7>.
- Shah, S.A., Szeto, A.H., Farewell, R., Shek, A., Fan, D., Quach, K.N., Kaul, S., 2019. Impact of high volume energy drink consumption on electrocardiographic and blood pressure parameters: a randomized trial. *Journal of the American Heart Association* 8 (11). <https://doi.org/10.1161/JAHA.118.011318>, 2019;8:e011318.
- \* Shalev, I., Israel, S., Uzevovsky, F., Gritsenko, I., Kaitz, M., Ebstein, R.P., 2011. Vasopressin needs an audience: neuropeptide elicited stress responses are contingent upon perceived social evaluative threats. *Horm. Behav.* 60 (1), 121–127. <https://doi.org/10.1016/j.yhbeh.2011.04.005>.
- \* Smith, A.L., Hantsoo, L., Malyhina, A.P., File, D.W., Valentino, R., Wein, A.J., Epperson, C.N., 2016. Basal and stress-activated hypothalamic pituitary adrenal axis function in postmenopausal women with overactive bladder. *International Urogynecology Journal* 27 (9), 1383–1391. <https://doi.org/10.1007/s00192-016-2988-6>.
- \* Souza, G.G., Medonça-de-Souza, A.C., Duarte, A.F., Fischer, N.L., Souza, W.F., Coutinho, E.S., Volchan, E., 2015. Blunted cardiac reactivity to psychological stress associated with higher trait anxiety: a study in peacekeepers. *BMC Neurosci.* 16 (81), 1–8. <https://doi.org/10.1186/s12868-015-0216-9>.
- Spielberger, C.D., Gorsuch, R.L., Lushene, R., Vagg, P.R., Jacobs, G.A., 1983. *Manual for the State-Trait Anxiety Inventory*. Consulting Psychologists Press, Palo Alto, CA.
- Stalder, T., Kirschbaum, C., Kudielka, B.M., Adam, E.K., Pruessner, J.C., Wust, S., Clow, A., 2016. Assessment of the cortisol awakening response: Expert consensus guidelines. *Psychoneuroendocrinology* 63, 414–432. <https://doi.org/10.1016/j.psyneuen.2015.10.010>.
- Stark, P.B., 2018. Before reproducibility must come reproducibility. *Nature* 557, 613. <https://doi.org/10.1038/d41586-018-05256-0>.
- Stein, D.J., Ruscio, A.M., Lee, S., Jordi Alonso, M.P., Andrade, L.H., Benjet, C., Kessler, R. C., 2010. Subtyping social anxiety disorder in developed and developing countries. *Depress. Anxiety* 27 (4), 390–403. <https://doi.org/10.1002/da.20639>.
- Stephens, M.A., Mahon, P.B., McCaul, M.E., Wand, G.S., 2016. Hypothalamic-pituitary-adrenal axis response to acute psychosocial stress: effects of biological sex and circulating sex hormones. *Psychoneuroendocrinology* 66, 47–55. <https://doi.org/10.1016/j.psyneuen.2015.12.021>.
- Steer, R.A., Beck, A.T., 1997. Beck anxiety inventory. In: Zalaquett, C.P., Wood, R.J. (Eds.), *Evaluating Stress: A Book of Resources*. Scarecrow Education, pp. 23–40.
- Taylor, S.E., Seeman, T.E., Eisenberger, N., Kozanian, T.A., Moore, A.N., Moons, W.F., 2010. Effects of a supportive or an unsupportive audience on biological and psychological responses to stress. *J. Pers. Soc. Psychol.* 98 (1), 47–56. <https://doi.org/10.1037/a0016563>.
- Summer, J.A., Hagan, K., Grodstein, F., Roberts, A.L., Harel, B., Koenen, K.C., 2017. Posttraumatic stress disorder symptoms and cognitive function in a large cohort of middle-aged women. *Depress. Anxiety* 34 (4), 356–366. <https://doi.org/10.1002/da.22600>.
- \* Task Force of the European Society , o. A, 1996. Heart rate variability: standards of measurement, physiological interpretation, and clinical use. *Circulation* 93 (5), 1043–1065.
- Thomas, S.E., Randall, P.K., Brady, K., See, R.E., Drobos, D.J., 2011. An acute psychosocial stressor does not potentiate alcohol cue reactivity in non-treatment-seeking alcoholics. *Alcoholism Clinical & Experimental Research* 35 (3), 464–473. <https://doi.org/10.1111/j.1530-0277.2010.01363.x>.
- \* Veer, I.M., Oei, N.Y., Spinhoven, P., van Buchem, M.A., Elzinga, B.M., Rombouts, S.A., 2011. Beyond acute social stress: increased functional connectivity between amygdala and cortical midline structures. *Neuroimage* 57 (4), 1534–1541. <https://doi.org/10.1016/j.neuroimage.2011.05.074>.
- Virella, B., Arbona, C., Novy, D.M., 1994. Psychometric properties and factor structure of the Spanish version of the state-trait anxiety inventory. *Journal of Personality Assessment* 63 (3), 401–412. [https://doi.org/10.1207/s15327752jpa6303\\_1](https://doi.org/10.1207/s15327752jpa6303_1).
- Von Dawans, B., Kirschbaum, C., Heinrichs, M., 2011. The Trier Social Stress Test for Groups (TSST-G): a new research tool for controlled simultaneous social stress exposure in a group format. *Psychoneuroendocrinology* 36 (4), 514–522. <https://doi.org/10.1016/j.psyneuen.2010.08.004>.
- \* Wand, G.S., Oswald, L.M., McCaul, M.E., Wong, D.F., Johnson, E., Zhou, Y., Kumar, A., 2007. Association of amphetamine-induced striatal dopamine release and cortisol responses to psychological stress. *Neuropsychopharmacology* 32 (11), 2310–2320. <https://doi.org/10.1038/sj.npp.1301373>.
- Ward, B.W., Schiller, J.S., Goodman, R.A., 2012. Multiple chronic conditions among US adults: a 2012 update. *Prev. Chronic Dis.* 11, E62. <https://doi.org/10.5888/pcd11.130389>.
- Watson, D., Clark, L.A., Tellegen, A., 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. *J. Pers. Soc. Psychol.* 54 (6), 1063–1070. <https://doi.org/10.1037/0022-3514.54.6.1063>.
- Wichmann, S., Kirschbaum, C., Böhme, C., Petrowski, K., 2017. Cortisol stress response in post-traumatic stress disorder, panic disorder, and major depressive disorder patients. *Psychoneuroendocrinology* 83, 135–141. <https://doi.org/10.1016/j.psyneuen.2017.06.005>.
- \* Wiegand, C., Heusser, P., Klinger, C., Cysarz, D., Bussing, A., Ostermann, T., Savelsbergh, A., 2018. Stress-associated changes in salivary microRNAs can be detected in response to the Trier Social Stress Test: an exploratory study. *Sci. Rep.* 8 (1), 7112. <https://doi.org/10.1038/s41598-018-25554-x>.
- Wiemers, U.S., Schoofs, D., Wolf, O.T., 2013. A friendly version of the Trier Social Stress Test does not activate the HPA axis in healthy men and women. *Stress* 16 (2), 254–260. <https://doi.org/10.3109/10253890.2012.714427>.
- Wilhelm, F.H., Roth, W.T., 2001. The somatic symptom paradox in DSM-IV anxiety disorders: suggestions for a clinical focus in psychophysiology. *Biol. Psychol.* 57 (1–3), 105–140. [https://doi.org/10.1016/S0301-0511\(01\)00091-6](https://doi.org/10.1016/S0301-0511(01)00091-6).
- Williams, E., Magid, K., Steptoe, A., 2005. The impact of time of waking and concurrent subjective stress on the cortisol response to awakening. *Psychoneuroendocrinology* 30 (2), 139–148. <https://doi.org/10.1016/j.psyneuen.2004.06.006>.
- Wingenfeld, K., Wolf, O.T., 2010. HPA Axis Alterations in mental disorders: impact on memory and its relevance for therapeutic interventions. *CNS Neurosci. Ther.* 17 (6), 714–722. <https://doi.org/10.1111/j.1755-5949.2010.00207.x>.



- Who, 1946. World health organization - constitution, 07 22, Retrieved from World Health Organization. [https://www.who.int/governance/eb/who\\_constitution\\_en.pdf](https://www.who.int/governance/eb/who_constitution_en.pdf).
- Who, 2001. The World Health Report : 2001 : Mental Health : New Understanding, New Hope. World Health Organization, Geneva.
- Who, 2010. WHO Guidelines on Drawing Blood: Best Practices in Phlebotomy. Retrieved from World Health Organization. [http://www.euro.who.int/\\_data/assets/pdf\\_file/0005/268790/WHO-guidelines-on-drawing-blood-best-practices-in-phlebotomy-Eng.pdf?ua=1](http://www.euro.who.int/_data/assets/pdf_file/0005/268790/WHO-guidelines-on-drawing-blood-best-practices-in-phlebotomy-Eng.pdf?ua=1).
- Who, 2016. Obesity and Overweight. Retrieved 03 03, 2020, from World Health Organization: <https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight>.
- Who, 2019. World health organization, 02 18, Retrieved 02 18, 2019, from Chronic disease and health promotion: [https://www.who.int/chp/about/integrated\\_cd/en/](https://www.who.int/chp/about/integrated_cd/en/).
- \* Xin, Y., Wu, J., Yao, Z., Guan, Q., Aleman, A., Luo, Y., 2017. The relationship between personality and the response to acute psychological stress. *Sci. Rep.* 7 (1), 16906. <https://doi.org/10.1038/s41598-017-17053-2>.
- \* Yao, Z., Zhang, L., Jiang, C., Zhang, K., Wu, J., 2016. Stronger cortisol response to acute psychosocial stress is correlated with larger decrease in temporal sensitivity. *Peer J* 26 (4), e2061. <https://doi.org/10.7717/peerj.2061>.
- \* Zhang, H., Yao, Z., Lin, L., Sun, X., Shi, X., Zhang, L., 2019. Early life stress predicts cortisol response to psychosocial stress in healthy young adults. *Psych J.* 8 (3), 353–362. <https://doi.org/10.1002/pchj.278>.
- Zhong, W., Maradit-Kremers, H., St Sauver, J.L., Yawn, B.P., Ebbert, J.O., Roger, V.L., Rocca, W.A., 2013. Age and sex patterns of drug prescribing in a defined American population. *Mayo Clin. Proc.* 88 (7), 697–707. <https://doi.org/10.1016/j.mayocp.2013.04.021>.
- Zhou, Y., Proudnikov, D., Yuferov, V., Kreek, M.J., 2010. Drug-induced and genetic alterations in stress-responsive systems: implications for specific addictive diseases. *Brain Res.* 1314, 35–252. <https://doi.org/10.1016/j.brainres.2009.11.015>.
- Zotero [Computer software], Roy Rosenzweig Center for History and New Media. Available at [www.zotero.org/download](http://www.zotero.org/download).