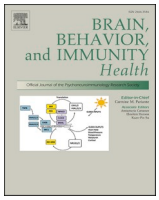
Brain, Behavior, & Immunity - Health 43 (2025) 100936

Contents lists available at ScienceDirect

Brain, Behavior, & Immunity - Health

journal homepage: www.editorialmanager.com/bbih/default.aspx



Measuring salivary cortisol in biobehavioral research: A systematic review and methodological considerations

Fanghong Dong [a], Justine S. Sefcik [b], Elizabeth Euiler [b], Nancy A. Hodgson [c,*]

[a] School of Medicine, Washington University, St. Louis, MO, 63110, USA
[b] College of Nursing and Health Professions, Drexel University, Philadelphia, PA, 19104, USA
[c] School of Nursing, University of Pennsylvania, Philadelphia, PA, 19104, USA

ARTICLE INFO

Keywords:
Salivary cortisol
Biobehavioral research
Community setting

ABSTRACT

The assessment of salivary cortisol in community settings has gained popularity in biobehavioral research due to its noninvasive sampling, ease of handling and storage, and suitability for repeated sampling in short intervals. Ensuring consistent methodological practices for salivary cortisol is essential. This systematic review critically examines salivary cortisol collection procedures, data cleaning, and analysis to better understand its role in biobehavioral research within community populations. Fifty-eight articles met the inclusion criteria. Results indicated significant variability in study designs and cortisol measurement procedures, particularly regarding the biobehavioral role of cortisol, sampling periods, covariate considerations, cortisol analysis parameters, and data analysis plans. The review highlights commonly used and promising study designs while identifying methodological issues in cortisol measurement and analysis that should be addressed to improve comparability in future research.

1. Introduction

Biobehavioral research is the investigation of interactions among behavioral, psychological, socioenvironmental and biological factors that contribute to our understanding of stress and health (Benedict, 2013). Stress in this context is defined as the pervasive phenomenon of everyday life which activates the hypothalamic pituitary adrenal (HPA) axis and has been shown to cause long-lasting negative health effects, both psychologically and physically (O'Connor et al., 2021; Yaribeygi et al., 2017). Cortisol is one biomarker of the HPA axis function that helps us understand a person's response to both daily activity and chronic stress (Hellhammer et al., 2009). Salivary cortisol is widely used for different research purposes because of its ease of collection and relative stability. Commonly used diurnal indices of cortisol activity include cortisol awakening response (surge in cortisol that occurs 30–45 min after waking), diurnal cortisol slope (degree of change in cortisol levels from morning to evening), and area under the daytime cortisol curve (area under all cortisol data points measured across the day) (Adam and Kumari, 2009). These are to understand the association between stress and various health outcomes common in biobehavioral research.

Saliva data collection is increasingly popular for understanding cortisol levels as it is relatively easy to obtain and handle without requiring specialized personnel to obtain cortisol blood, urine, or hair samples. This technique is also particularly helpful with vulnerable populations such as children, older adults, or individuals who may have difficulty or resistiveness to donating blood samples, especially in the naturalistic, community setting. Additionally, research participants (and their caregivers when needed) can collect saliva cortisol samples in their home and store them in a home freezer until returning the specimens to research staff (Hodgson and Granger, 2013). This community-based process effectively reduces study cost, is more convenient than lab-based collection, and allows for measurement in the naturalistic setting of everyday life.

As the assessment of salivary cortisol becomes more common in biobehavioral research, consistency of methodological practices in collection and data analysis is paramount to ensure measurement reliability and to aid in cross-study comparisons. While there are several guidelines entailing recommendations for a saliva cortisol collection protocol (Adam and Kumari, 2009; Stalder et al., 2016), these are written for a controlled, lab setting, and do not have considerations for the nuances that may occur in naturalistic settings such as the home. In the community setting, obtaining saliva samples may not be easy, as successful collection depends on the implementation of strict protocols

* Corresponding author.
E-mail address: hodgsonn@nursing.upenn.edu (N.A. Hodgson).

https://doi.org/10.1016/j.bbih.2024.100936
Received 2 June 2024; Received in revised form 21 November 2024; Accepted 21 December 2024
Available online 29 December 2024

by participants and/or their caregivers. Outside of the lab there is a lack of moment-to-moment oversight of a researcher to ensure participants' compliance with the procedures. Therefore, it is critical to have an understanding of common procedures that incorporate unique considerations for the community setting.

Previous literature reviews on salivary cortisol have focused on sampling protocols for clinical purposes (Bellagambi et al., 2020), without attention to the procedures used for measurement and analysis of cortisol in naturalistic settings. Prior reviews focused primarily on specific types of saliva data collection (e.g., spitting, swab-based sampling, drool), procedures for saliva data collection, commercially available sampling device, handling, transporting, and archiving samples (Bellagambi et al., 2020; Padilla et al., 2020). One notable gap in these reviews was the lack of attention on procedures for cleaning and analyzing raw cortisol data that have been collected in the community.

Understanding the best practices for cleaning and analyzing raw salivary cortisol data is particularly important for researchers working with community-based participants as protocol adherence deficiencies from participants collecting their own saliva (e.g., late collection of saliva, skipping collection times) are common. Cortisol analysis requires precise timing of data collection, and when research participants do not follow protocols and adhere to scheduled collection times it may influence interpretations of a person's diurnal cortisol profile. Knowing best practices for cleaning and analyzing raw salivary cortisol data can inform researchers approach to cortisol analysis and contribute additional information for future guideline revisions.

Furthermore, there are no known reviews of the literature that synthesize what protocols biobehavioral researchers are using to direct their participants to collect salivary cortisol in the community. Elucidating this information will inform those working on study design that includes the self-collection of saliva samples from participants. To address these gaps in the literature, the aim of this systematic review was to summarize the sampling protocols, analysis parameters, data cleaning, and statistical approaches for salivary cortisol collected by research participants in the community.

Specifically, we were interested in answering the following questions.

1) What were the salivary cortisol sampling protocols reported?
2) What were the salivary cortisol parameters reported?
3) What data cleaning and statistical approaches were reported in the association between salivary cortisol and a biobehavioral component?

2. Methods

We followed the Cochrane Handbook for Systematic Reviews for conducting the review (Higgins et al., 2019) and use Preferred Reporting Items for Systematic Reviews and Meta-Analyses Protocols (PRISMA-P) statement for reporting the results of this study (Moher et al., 2015; Shamseer et al., 2015). This systematic review protocol was registered with the PROSPERO (register No. CRD42021237402). We followed five stages in this systematic review: (1) literature search, (2) article selection, (3) data extraction, (4) quality assessment, and (5) data synthesis.

2.1. Literature search strategy

We developed the following Population Intervention Comparison Outcome (PICO) framework to guide the search strategy. Studies that examined the association between salivary cortisol (either exposure or outcome) and the biobehavioral measure (intervention, exposure, or outcome) in the target population aged 10 and above (population). Studies did not need to include a control or comparison group for inclusion in this systematic review.

This PICO has been converted to a search strategy as shown in Table 1S. The main search terms are related to salivary cortisol and biobehavioral research. We searched the following five databases between January 1, 2016 and June 30, 2021: (1) PubMed; (2) Embase; (3) Scopus, (4) CINHAL, and (5) PsycINFO. To identify potentially relevant grey literature, we searched Google Scholar and Google search engines. The search strategy for the five databases was developed in consultation with a medical librarian. The complete search strategy is included in Supplement Table 2S.

2.2. Article selection

The detailed inclusion and exclusion criteria for this review are shown in Table 3S. All records identified from the database or search engines was recorded in a software management program EndNote X9 (Clarivate Analytics). The EndNote library was also used to remove any duplicates. The library was uploaded into Covidence, an online software to help manage systematic reviews. Two independent reviewers (FD and JS) screened the title and abstract of all identified studies against the eligibility criteria. The full text of the identified studies was then reviewed and assessed for eligibility. Disagreements were resolved by discussion or by consultation with a third reviewer (NH). Once the final list of studies was determined, the references for each included article was searched to identify additional studies that should be considered for inclusion.

A PRISMA flow diagram was created to document the selection process and reasons for article exclusions to ensure repeatability of the search results. This included (1) Identification: records identified through database searching, additional records identified through other sources, and records after duplicates removed; (2) Screening (by title and abstract): including the number of records screened and records excluded; (3) Eligibility: full-text articles assessed for eligibility and full-text articles excluded, with reasons; and (4) Included: studies included in qualitative synthesis.

2.3. Data extraction

Study characteristics *were* extracted by one author (FD) and completely audited by another author (EE). Differences were reconciled through meetings. Data were extracted using a data extraction sheet including the following information: (1) publication details: author, date of publication, and country of study population; (2) study design: aims of study, type of study (cross-sectional, longitudinal, experimental/randomized clinical trial(RCT), quasi-experimental), role of salivary cortisol in the study; (3) study participants, including number of participants, population characteristics including age, gender, race/ethnicity, socioeconomic status and body mass index; (4) saliva collection device and method, and salivary cortisol collection protocol, saliva collection time, protocol adherence, and sample transportation before doing the lab analysis; (5) cortisol parameter measured; (6) salivary cortisol data cleaning procedure: raw data preparation (including data completeness, quality and consistency of both saliva sample and collection time if applicable), defining impossible value, missing values, and outliers; (7) main statistical analyses of the associations between the salivary cortisol and measures of biobehavioral components.

2.4. Quality assessment

The Crowe Critical Appraisal Tool (CCAT), version 1.4 (Crowe and Sheppard, 2011) was used to assess the quality of all included studies. Total scores on the CCAT ranges from 0 to 40, with a higher score indicating higher overall quality of the study. Two reviewers independently completed the tools and met to reach consensus on scores.

2.5. Data synthesis

We summarized the findings and provided a synthesis in Table 1 and in narrative form. These results summarized and described the salivary

Table 1
Characteristics of included studies.

Study	Design	Behavioral components	Cortisol	Sample characteristics	Saliva collection device	Salivary cortisol collection protocol	Saliva collection time recording	For longitudinal study, list follow-up sampling time points	Temperature of frozen during transportation to lab	Saliva assay in lab analysis	Unit of raw cortisol level
1. Abshire et al., (2018). United States	cross-sectional	Quality of life and functional status, fatigue	predictor	N = 44; 73% male; mean age 57.7 ± 13 years; 45.5% white; SES: NR; BMI: NR	NR	three samples per day (at waking, 30 min after waking, and before going to bed) on 2 days when they expected to have a "normal" routine.	self-recorded collection time (log)		at −20 °C	EIA	ug/dl
2. Anderson et al. (2021). United States	cross-sectional	physical activity	outcome	N = 85; 72.9% Female; mean age 19.06; 44.7% White; SES: NR; BMI: 26.1 ± 6.2	passive drool	two samples per day (after waking (S1) and 30 min after waking (S2)) for four consecutive weekdays and nights (Monday to Thursday, or Tuesday to Friday), beginning at 5 pm on the first day of the study (either a Monday or Tuesday), and ending between 10 a.m. and 5 pm on the final day of the study (corresponding to either a Thursday or Friday).	medication event monitoring system		at −20 °C	DELFIA	nmol/L
3. Armer et al. (2018). United States	longitudinal	life stress	outcome	N = 337; all Female; mean age 59.7 ± 11.68; 96.4% Caucasian (93.5% non-Hispanic); SES: NR; BMI: NR	NR	three samples per day (upon awakening, between 4pm and 6:30pm, and at bedtime) on three days before surgery	self-recorded collection time	before surgery, the 6-month follow-up appointment (was typically completed 1 month post-chemotherapy completion), and the 1-year follow-up (was completed at the routine 12-month clinic visit)	at −80 °C	CLIA	nmol/L
4. Ayala-Grosso et al. (2021). Venezuela	cross-sectional	behavioral attitudes indexes	predictor	N = 135; 30% Female; mean age 46.52 ± 4.24; Population from Valle la Pascua; SES: 55% university level education; BMI: NS, 54% overweight	cotton-based collection	four samples per day (at time of awakening, 2 h later, at noon and at 6 p.m. before dinner time) for one day	NS		4 °C; then at −20 °C	DELFIA	nmol/L

(*continued on next page*)

Table 1 (*continued*)

Study	Design	Behavioral components	Cortisol	Sample characteristics	Saliva collection device	Salivary cortisol collection protocol	Saliva collection time recording	For longitudinal study, list follow-up sampling time points	Temperature of frozen during transportation to lab	Saliva assay in lab analysis	Unit of raw cortisol level
5. Basson et al. (2019). Canada	cross-sectional	sexual function	outcome	N = 275; all female; mean age 33.01 ± 11.68 (control), 31.81 ± 12.05 (experimental); 60.6%/58.7% Euro-Caucasian, 18.2%/25.4% East Asian; SES: NR; BMI: NR	passive drool	four samples per day (at awakening, 30 min and 60 min after waking, and immediately before bedtime) on three separate, typical weekdays	NR		at −15 °C	DELFIA	nmol/L
6. Benz et al., (2019). Germany	cross-sectional	Self-reports from female participants on use of OC and menstrual cycle phase, depression and anxiety as covariates	outcome	N = 51; 41 women, 10 men; mean age = 21.32 ± 3.28 (women), 24.90 ± 3.00 (men); race: NR; SES: NR; BMI: range 17.5–29.9	cotton-based collection (swab, Salivette)	ten samples per day (during 270 min after awakening, at intervals of 30 min to get one sample) for two weekdays, two observations per participant	medication event monitoring system		dry place, then at −20 °C	DELFIA	nmol/L
7. Bernsdorf and Schwabe. (2018). Germany	cross-sectional	sleep- and stress-related factors	outcome	N = 48 (24 children, 24 adults); 50% female; children mean age = 7.58 ± 0.26, adults mean age = 41.33 ± 0.79; German; SES: NR; BMI: children 15.28 ± 0.33, adults 25.19 ± 0.82	cotton-based collection (swab, Salivette)	four samples per day (the first immediately after awakening, while still lying in bed, as well as 15, 30, and 45 min after awakening) on four days (2 weekdays and on 2 weekend days.)	self-recorded collection time		at −18 °C	CLIA	nmol/L
8. Bitsika et al., 2017. Australia	cross-sectional	child based behaviors	outcome	N = 149; 135 female; race: NR; SES: NR; BMI: NR	cotton-based collection (swab, Salivette)	two samples per day (30–45 min after they awoke in the morning as well as between the hours of 2.00 pm and 4.00 p.m.) for one day	NR		under 20 °C, then at −80 °C	ELISA	nmol/L
9. Boss et al. (2016). United States	cross-sectional	religious coping	outcome	N = 88; 66% females; mean age 75.4 ± 9.0; 94% Caucasians; SES: 44% high school education (M = 12.3 years); BMI: 29.6 ± 6.22	NR	one afternoon saliva sample between 1:00 pm and 5:00 pm	NR		iced bag, then at −80 °C	EIA	ug/dl
10. Chandola et al. (2018). UK	longitudinal	sleep (hours)	outcome	N = 1143; gender NS; race: NR; SES: NR; BMI: NS as covariate	cotton-based collection (swab, Salivette)	Six samples per day (At waking, after waking 30 min, 2.5 h, 8 h, 12 h, and bedtime)	self-recorded collection time (log)	at phases 7 and 9	elsewhere	CLIA	nmol/L

(*continued on next page*)

Table 1 (*continued*)

Study	Design	Behavioral components	Cortisol	Sample characteristics	Saliva collection device	Salivary cortisol collection protocol	Saliva collection time recording	For longitudinal study, list follow-up sampling time points	Temperature of frozen during transportation to lab	Saliva assay in lab analysis	Unit of raw cortisol level
						on a normal weekday					
11. Charles et al. (2020). United States	cross-sectional	cognitive function, and allostatic load.	predictor	N = 1735 (final N = 1001); 892 Female; mean age 55.99 ± 12.3; 93% European-American; SES: 48% well-educated; BMI: NR	cotton-based collection (swab, Salivette)	four samples per day (immediately upon waking, 30 min after waking, before lunch, and before bed) on four consecutive days on days 2–5 of the NSDE 8-day study:	self-recorded collection time (both log and nightly telephone interview; 25% "smart boxes" that contained a computer chip that recorded)		at −60 °C	CLIA	log units
12. Chiang et al. (2016). United States	cross-sectional	sleep	outcome	N = 316; 180 Female; mean age 16.40 ± 0.74; 29.1% European, 41.8% Latino, 23.1% Asian; SES: middle-class, median income $50,000; BMI: 23.16 ± 5.01	cotton-based collection (swab, Salivette)	During the first three days, participants provided saliva samples at 5 time points throughout the day: at waking, 15 min post-wake, 30 min post-wake, before dinner, and before bed.	self-recorded collection time (stamping booklet)		fridge, then at −80 °C	CLIA	nmol/L
13. Chin et al., 2017. United States	cross-sectional	marital status	outcome	N = 572; 48% Female; mean age 33.7 ± 10.2; 63% white, 32% African-American, 5% other; SES: NR; BMI: NS as covariate	cotton-based collection	Among two viral-challenge studie, seven samples per day (1, 2, 4, 6, 8, 12, and 14 h post-waking) on each of the pre-quarantine days, and eight samples during the first 24 h of quarantine (0, 1, 2, 4, 5, 7, 9, and 14 h post-waking). The third viral-challenge studies, seven samples from pre-quarantine days (assessed at 1, 2, 4, 7, 9, 11, and 14 h post-waking) and eight samples from the baseline day of quarantine (0, 1, 4.25, 6.25, 7.25, 9.25, 12.75, and 16.75 h post-waking)	Both detailed written instructions and either a pre-programmed wristwatch or handheld computer were provided to signal participants at each collection time. In addition, the signaling device also provided a unique alphanumeric code for each collection. Participants were instructed to write the code as well as the exact time and date of collection on each tube right after it was sealed		NS (their own fridge)	DELFIA	log units

(*continued on next page*)

Table 1 (*continued*)

Study	Design	Behavioral components	Cortisol	Sample characteristics	Saliva collection device	Salivary cortisol collection protocol	Saliva collection time recording	For longitudinal study, list follow-up sampling time points	Temperature of frozen during transportation to lab	Saliva assay in lab analysis	Unit of raw cortisol level
14. Corominas-Roso et al., 2017. Spain	cross-sectional	Attention Deficit Hyperactivity Disorder (ADHD) subtype	just descriptive and correlation	N = 108 ADHD + 27 controls; 44 female (ADHD), 13 female (control); mean age 35.5 ± 10.23 (inattentive), 35.6 ± 9.20 (combined), 32 ± 8.6 (control); race: NR; SES: NR; BMI: 24.99 ± 5.62 (inattentive), 24.02 ± 5.22 (combined), 24.86 ± 4.82 (control)	sampling reported elsewhere	four morning samples at 0, 30, 45 and 60 min after awakening on one day	self-recorded collection time		at −80 °C	ELISA	NS
15. Cuneo et al. (2017). United States	cross-sectional	fatigue	outcome	N = 30; all Female; mean age 63.2 ± 13; all Caucasian, Non-Hispanic; SES: NR; BMI: NR	cotton-based collection (swab, Salivette)	three samples per day(upon awakening, at 5pm, and at bedtime) for three consecutive days	NR		NR	CLIA	nmol/L; log unit
16. D'Cunha et al., 2019. Australia	quasi-experimental	intervention, behavioral observations and exit questionnaire	outcome	N = 22; 16 Female; mean age 84.6 ± 7.27; race: NR; SES: NR; BMI: 26.1 ± 5.09	passive drool	four times per day (upon waking, after 30 min, 60 min after breakfast, and 45 min after dinner) for one day	NS	baseline (the day before the first visit), post-intervention at six weeks (the day after the final visit), and follow-up at twelve weeks (six weeks post-intervention).	dry ice, then at −20 °C	NR	nmol/L
17. Darabos et al., 2019. United States	cross-sectional	constructive and unconstructive processing, as measured from a cancer related expressive writing task	outcome	N = 17; all male; mean age 25.41 ± 3.24; 47.1% White, 23.5% Hispanic; SES: NR; BMI: NR	cotton-based collection (swab, Salivette)	four samples per day(upon waking (morning), 30 min after awakening, 8 h after awakening, and at bedtime) on 3 consecutive weekdays	NR		at −20 °C	EIA	ng/dl
18. Doolin et al., 2017. Ireland	cross-sectional	to compare HPA axis activity between depressed patients (MDD) and healthy controls, with a more specific measure of salivary cortisol and cortisone concentrations using the liquid chromatography-	outcome	N = 97 (57 MDD, 40 control); 37 Female (MDD); mean age 28.26 ± 8.41 (MDD), 27.48 ± 5.61 (control); most Europeans; SES: NR; BMI: 24.96 ± 6.17 (MDD), 22.81 ± 3.25 (control)	cotton-based collection (swab, Salivette)	five samples per day (at post-wakening time points (0, +30, +60, +720 and + 750 min) for one day	NR		at −80 °C	LC-MS	nmol/L

(*continued on next page*)

Table 1 (*continued*)

Study	Design	Behavioral components	Cortisol	Sample characteristics	Saliva collection device	Salivary cortisol collection protocol	Saliva collection time recording	For longitudinal study, list follow-up sampling time points	Temperature of frozen during transportation to lab	Saliva assay in lab analysis	Unit of raw cortisol level
		mass spectrometry (LC-MS) technique									
19. Engert et al. (2018). Germany	cross-sectional	health and sleep	outcome	N = 328; 195 women; mean age 40.65 ± 9.25; race: NR; SES: NR; BMI: NR	cotton-based collection (swab, Salivette)	seven samples per day (free awakening (while still in bed) and at 30, 60, 240, 360, 480 and 600 min after awakening) on two consecutive days; 60 min not use in this study	self-recorded collection time (preprogrammed mobile device to remind)		fridge, then at −30 °C	DELFIA	nmol/L
20. Fuentecilla et al. (2019). United States	cross-sectional	support provision	outcome	N = 151; 54% Female; mean age 55.65 ± 4.58; race: NR; SES: 65% work full time; BMI: NR	cotton-based collection (swab, Salivette)	four samples a day (upon waking, 30 min after waking, at noon, before bed) for one day	self-recorded collection time		at −80 °C	CLIA	nmol/L
21. Garcia, A.F. et al., 2017. United States	cross-sectional	health, acculturative stress	mediator	N = 89; 46 Female (51.7%); median age 20; adult Mexican Americans; SES: 56.6% income <40K; BMI: 24.82 ± 2.79	cotton-based collection (swab, Salivette)	Four samples per day (at awakening, 30, 45, and 60 min thereafter) on two consecutive weekdays	self-recorded collection time (log); medication event monitoring system		fridge, then at −80 °C	EIA	ug/dl
22. Garcia, M. et al., 2021. United States	cross-sectional	loneliness, disability	outcome; aim 1 is correlation	N = 62; all Female; age 18–54; 89% Caucasian, 2.9% Black, 5.2% multi-racial; SES: NR; BMI: NR	cotton-based collection (swab, Salivette)	waking, 30 min after waking, 45 min after waking, noon, 4 p.m., 8 p.m. in two consecutive weekdays	self-recorded collection time		NS, their own fridge	ELISA	NR
23. Goldstein et al., 2017. United States	cross-sectional	maternal histories of anxiety and depression; parenting bullying	predictor	N = 476; all Female; mean age 14.4 ± 0.62; 81.3% white and non-Hispanic; SES: most parents completed 4-year college; BMI: NS as covariate	cotton-based collection (MEMSCap™ bottle)	three samples per day (immediately upon waking, 30 min after waking, and approximately 8:00 p.m.) on three consecutive weekdays.	self-recorded collection time; medication event monitoring system		at −80 °C	DELFIA	nmol/L
24. Herane-Vives et al., 2018. UK and Chile	cross-sectional	Depression (Atypical major depressive episodes, A-MDE)	outcome	N = 111 (44 non-A-MDE, 27 A-MDE, 40 controls); 28/20/29 females; mean age 34.5/31.9/33.2; race: NR; SES: NR; BMI: 25.4/26.7/24.3	cotton-based collection (swab, Salivette)	six saliva samples ((i) immediately after awakening, (ii) 30 min after awakening, (iii) 60 min after awakening, (iv) at noon, (v) at 4 p.m. and (vi) at 8 p.m.) on a single day	self-recorded collection time (log)		fridge, then at −20 °C	CLIA	nmol/L

(*continued on next page*)

Table 1 (*continued*)

Study	Design	Behavioral components	Cortisol	Sample characteristics	Saliva collection device	Salivary cortisol collection protocol	Saliva collection time recording	For longitudinal study, list follow-up sampling time points	Temperature of frozen during transportation to lab	Saliva assay in lab analysis	Unit of raw cortisol level
						between Tuesday and Friday					
25. Ho, Lo et al., 2020. Hongkong, China	RCT	intervention	outcome	N = 51 parent-child dyads; 92.3% Female parent; mean age 40/39.2/38.3; Chinese; SES: average monthly income 13.7K HKD; BMI: NR	cotton-based collection (swab, Salivette)	four sample (after wakeup around 07:30, before lunchtime around 12:00, late-afternoon 17:30, and before sleep21:30) on a week day	self-recorded collection time (reminder notes)	The baseline sample were collected within one week before the first session of the intervention. The post-intervention sample were collected within three days after the last session of the intervention	NS	ELISA	nmol/L
26. Ho, Fong, Yau et al., 2020. Hongkong, China	longitudinal	daily functioning; functional performance	mediator	N = 189 (final N = 157); 82% female; mean age 78.9 ± 8.1; Chinese population; SES: 66.5% ≤ 10 years education; BMI: NR	cotton-based collection	five times on a weekday (wake-up (Sample 1), 1 h after wake-up (Sample 2), noon (Sample 3), late afternoon at 5 p.m. (Sample 4), and evening at 9 p.m. (Sample 5)	self-recorded collection time (reminder notes)	only at Time 1	NS (keep frozen)	ELISA	nmol/L
27. Ho, Fong, Chan et al., 2020. Hongkong, China	RCT	intervention	outcome	N = 166; 81.9% Female; mean age 79 ± 8.0; Chinese population; SES: NR; BMI: NR	cotton-based collection	five times on a weekday (wake-up (Sample 1), 1 h after wake-up (Sample 2), noon (Sample 3), late afternoon at 5 p.m. (Sample 4), and evening at 9 p.m. (Sample 5)	self-recorded collection time	All participants were assessed at four-time points over 12 months. Baseline data were collected 1 week before the start of the intervention (Time 1). Postintervention assessment (Time 2) was administered at the end of the intervention, that is, 3 months after baseline. Two follow-up assessments were conducted at 6 months (Time 3) and 12 months (Time 4) after baseline.	NS	ELISA	nmol/L
28. Holmqvist-Jämsén et al., 2017. Finland	cross-sectional	vocal symptoms (health)	predictor	N = 170; 121 Female; race: NR; SES: NR; BMI: NR	cotton-based collection (swab, Salivette)	one sample in the morning immediately after waking up,	self-recorded collection time		under 20 °C, then at −80 °C	RIA	nmol/L

(*continued on next page*)

Table 1 (*continued*)

Study	Design	Behavioral components	Cortisol	Sample characteristics	Saliva collection device	Salivary cortisol collection protocol	Saliva collection time recording	For longitudinal study, list follow-up sampling time points	Temperature of frozen during transportation to lab	Saliva assay in lab analysis	Unit of raw cortisol level
						preferably before 9 a.m.					
29. Hooper, 2019. United states	quasi-experimental	intervention	outcome	N = 115; 44% Female; mean age 48 ± 10.38; 72 African American, 43 White; SES: 85% ≥high school, 55% income <$10,000; BMI: NR	cotton-based collection (swab, not specify, Salimetrics)	four samples per day (upon waking, 30 min after waking, 4:00 p.m., and at 6:30 p.m.) for one day	NR	at baseline, the EOT, and at the one-month follow-up.	iced bag or fridge, then at −80 °C	RIA	ug/dl
30. Huang et al., 2020. Taiwan, China	cross-sectional	sleep	predictor	N = 108 (75 HCC, 33 controls); 81.3%/66.7% Female; mean age 61.25 ± 12.56/ 55.55 ± 11.55; Chinese; SES: NR; BMI: NR	cotton-based collection (swab, Salivette)	five time points (on waking, 30 min after waking, 12 pm, 5 pm, and bedtime) on 3 consecutive days	self-recorded collection time (daily phone)		fridge, then at −70 °C	ELISA	ug/dl
31. Huynh et al., 2016. United states	cross-sectional	discrimination, sleep (wake time)	outcome	N = 292; 58% Female; mean age 16.39 ± 0.74; 42% Latin American, 29% European, 23% Asian; SES: mean household income $71,374; BMI: NS as covariate	cotton-based collection (swab, Salivette)	five samples at designated times (wake, 15 min after wake, 30 min after wake, before dinner, and at bed time) for three consecutive day; Adolescents provided three days of cortisol samples on different days of the week. Only weekday samples were included in the analyses.	self-recorded collection time (stamping booklet)		frozen, then at −20 °C	ELISA	log units
32. Jakuszkowiak-Wojten et al., 2016. Poland	cross-sectional case control	diseases	outcome and descriptive	N = 28 (14 PD, 14 control); gender NR; median age 32.3/ 32.2; race: NR; SES: NR; BMI: NS	cotton-based collection (swab, Salivette)	three samples (immediately after awakening and 15 and 30 min later) in one morning	self-recorded collection time (stamping booklet)		at −80 °C	ELISA	nmol/L
33. Johnso et al., 2020. Canada	RCT	fatigue	mediator	N = 77; 85.7% female; mean age 58.1 ± 10; 72 white; SES: NR; BMI: 27.5 (range 18–45)	cotton-based collection (swab, Salivette)	four time a day (waking, noon, 5 p.m., bedtime) on three consecutive days at baseline; as close to the end of the week as possible	self-recorded collection time (also stamping tube with time)	Four weeks (final week of light use)	fridge or freezer, then at −80 °C	CLIA	nmol/L

(*continued on next page*)

Table 1 (*continued*)

Study	Design	Behavioral components	Cortisol	Sample characteristics	Saliva collection device	Salivary cortisol collection protocol	Saliva collection time recording	For longitudinal study, list follow-up sampling time points	Temperature of frozen during transportation to lab	Saliva assay in lab analysis	Unit of raw cortisol level
34. Keefe et al., 2019. United states	RCT	intervention, anxiety?	outcome	N = 45; 29 Female; mean age 45.60 ± 16.40; 62% Caucasian; SES: NR; BMI: NR	cotton-based collection (swab, Salivette)	four samples per day (at 8a. m.,12pm, 4pm, and 8pm) on three concurrent days	NS	prior to the initiation of treatment (Baseline), and subsequently for concurrent three days prior to their final assessment for the open-label phase of treatment (Week 8)	at −20 °C	CLIA	nmol/L
35. Kristiansen et al., 2020. Sweden	cross-sectional	perceived stress, sleep	outcome	N = 167 (63 adults); 43 Female (adults); mean age 36.7 ± 11.1 (control), 44.3 ± 12.1 (diabetes); Swedish; SES: NR; BMI: 25.8 ± 4.0 (control), 26.0 ± 3.3 (diabetes)	cotton-based collection (swab, SalivaBio®)	three samples an evening sample, collected within 1 h before going to bed; a morning sample, collected directly at awakening; and a second morning sample, collected 30 min after the first morning sample) for one day	self-recorded collection time (diary) and recording device		fridge, then at −20 °C	ELISA	nmol/L
36. Labad et al. (2018). Spain	cross-sectional	clinical symptoms	outcome	N = 89 (21 ARMS, 34 FEP, 34 control); 6/10/10 Female; mean age 22.1 ± 5.1/23.9 ± 5.0/ 24.3 ± 4.3; race: NR; SES: NR; BMI: 22.7 ± 3.5/24.1 ± 3.8/23.2 ± 3.7	cotton-based collection (swab, Salivette)	6 saliva samples per day (at the following sampling times: awakening (T1), 30-post-awak-ening (T2), 60-post-awakening (T3), 10:00 h (T4), 23:00 h (T5). Participants were told to intake DEX at 23:00 h just after T5 sample collection, and the next day at 10:00 h, another salivary sample was obtained for assessing post-DEX cortisol levels (T6).)	NR		at −20 °C	ELISA	nmol/L
37. Landau et al. (2021). Australia	RCT	depression	either outcome or predictor	N = 122; 73 Female; mean age 12.71 ± 1.01; race: NR; SES: NR; BMI: NR	passive drool	two samples (in the morning upon waking and in the evening) for two consecutive weekdays	NS, elsewhere, but report time difference	at baseline (T1),and a two-year follow-up (T3)	elsewhere (at −80 °C)	ELISA	ug/dl

(*continued on next page*)

Table 1 (*continued*)

Study	Design	Behavioral components	Cortisol	Sample characteristics	Saliva collection device	Salivary cortisol collection protocol	Saliva collection time recording	For longitudinal study, list follow-up sampling time points	Temperature of frozen during transportation to lab	Saliva assay in lab analysis	Unit of raw cortisol level
38. Laures-Gore et al. (2018). United States	cross-sectional	language production/ perceived stress	outcome	N = 33 (19 aphasia, 14 control); 7/7 Female; mean age 55.47 ± 11.86/ 55.53 ± 11.9; 17/8 Caucasian, 2/6 African American; SES: income reported; BMI: NR	cotton-based collection (swab, Salivette)	seven times per day(upon awakening, 30 and 60 min later, AQ8 and then at 1100 h, 1500 h, 1800 h, and bedtime)for one day	medication event monitoring system		at −20 °C	EIA	nmol/L
39. Liu et al., 2017. United States	longitudinal	health change	predictor	N = 141; 133 Female; mean age 60.65 ± 10.84; race: NR; SES: NR; BMI: NR	cotton-based collection (swab, Salivette)	five samples per day (i.e., before getting out of bed, 30-min after getting out of bed, before lunch, before dinner, and before bed) for 8 consecutive days.	self-recorded collection time (daily phone)	at baseline, 6 and 12 months	fridge, then at −80 °C	EIA	nmol/L
40. Mitchell et al. (2020). United States	cross-sectional	health indicators	outcome	N = 88; 63 Female; mean age 49.1 ± 14.57; 55.7% Hispanic, 28.4% Non-Hispanic white, 15.9% non-Hispanic black; SES: household income reported; BMI: 28.7 ± 6.9	cotton-based collection (swab, Salivette)	three times per day (at waking, late afternoon, and bedtim) for two consecutive days	NR		NR	CLIA	nmol/L
41. Morgan et al. (2017). United States	cross-sectional	sleep	outcome	N = 672; 364 Female (53.7%); mean age 71.5; 83.2% White non-Hispanic, 3.9% White Hispanic, 6.8% African American; SES: NR; BMI: NR	cotton-based collection (swab, Salivette)	three samples (at the beginning of the interview, partway through the interview, and at the completion of the interview) for one day	NS, Each sample had a time stamp.		fridge, then at −80 °C	CLIA	nmol/L
42. Otto et al. (2018). United States	cross-sectional	trait emotion regulation strategy	outcome	N = 46; 23 Female; mean age 54.04 ± 10.24; 63% Caucasian; SES: NR; BMI: NS as covariates	cotton-based collection (swab, Salivette)	four saliva samples per day (immediately upon waking, 30 min after waking, before lunch, and before bed) on four consecutive days	self-recorded collection time (both log and nightly telephone interview)		elsewhere	EIA	nmol/L
43. Pace et al., 2020. United States	cross-sectional	HRQOL	just descriptive and correlation	N = 22 dyads; mean age 52.41 ± 11.25 (survivor)/45.32 ± 14.77 (caregiver); all Latina; SES: income ranges reported; BMI:	cotton-based collection (swab, Salivette)	three samples (immediately on awakening in the morning, between 4:30 p.m. and 6:00 p.m., and bedtime)on over 2	self-recorded collection time (both log and nightly telephone interview/text reminder)		NS, freezer, dry ice	EIA	ug/dl

(*continued on next page*)

Table 1 (*continued*)

Study	Design	Behavioral components	Cortisol	Sample characteristics	Saliva collection device	Salivary cortisol collection protocol	Saliva collection time recording	For longitudinal study, list follow-up sampling time points	Temperature of frozen during transportation to lab	Saliva assay in lab analysis	Unit of raw cortisol level
				32.38 ± 7.00 (survivor)/30.07 ± 7.00(caregiver)		consecutive days					
44. Ramos-Quiroga et al. (2016). Spain	experimental, not RCT	emotion lability, ADHD disease	outcome	N = 136 (109 ADHD, 27 control); 45/13 Female; mean age 35.56 ± 9.55; all Caucasian; SES: NR; BMI: NR	NR	four saliva samples (at 0, 30, 45 and 60 min after awakening) for one day; on weekdays at home while patients were performing standard morning activities.	NS		at −80 °C	ELISA	nmol/L
45. Rosnick et al., 2016. United States	RCT	intervention	outcome	N = 42; 81% (cognitive behavioral therapy, CBT/76% (no CBT) female; mean age 71.19 ± 8.68 (CBT)/68.71 ± 7.97 (no CBT); 86%/81% White; SES: NR; BMI: NR	sampling reported elsewhere, (Salimetrics, LLC, State College, PA)	three daily saliva samples (immediately upon awakening, 30 min after waking, and at bedtime) on two consecutive days	self-recorded collection time (both diary and phone reminder)	both at the beginning and end of the 16-week CBT vs. no-CBT augmentation phase	elsewhere	ELISA	ug/dl
46. Sampedro-Piquero et al., 2020. Spain	cross-sectional	craving	predictor	N = 27 (14 substance use disorder, 13 control); all male; mean age 36.2 ± 2.3/40.6 ± 3.2; white Caucasian; SES: 15.7 ± 0.5/ 17.3 ± 0.8 education years; BMI: NR	cotton-based collection (swab, Salivette)	three samples per day(between 08.00 and 09.00 before breakfast and at least one hour after waking to avoid interfering with the cortisol awakening response, at 16.00 to 17.00 and before going to sleep (23.00–24.00))for one day	NR		fridge, then at −20 °C	ELISA	ug/dl
47. Schreier & Chen. 2017. United States	cross-sectional	life stress	outcome	N = 261; 53.3% female; mean age 14.3 ± 1.07; 49% European, 36%	cotton-based collection	four saliva samples (1, 4, 9, and 11 h following wake-	self-recorded collection time (provided stamper (DYMO Datemark))		at −30 °C	ELISA	nmol/L

(*continued on next page*)

Table 1 (*continued*)

Study	Design	Behavioral components	Cortisol	Sample characteristics	Saliva collection device	Salivary cortisol collection protocol	Saliva collection time recording	For longitudinal study, list follow-up sampling time points	Temperature of frozen during transportation to lab	Saliva assay in lab analysis	Unit of raw cortisol level
				Asian; SES: family income range <$5000 to >$200,000; BMI: 21.37 ± 3.70		up) for six consecutive days					
48. Schuler et al., 2017. United States	longitudinal	depression, stressful events	predictor	N = 527; all female; mean age 14.39 ± 0.62; 81.6% non-Hispanic white; SES: NR; BMI: 21.79 ± 4.14	cotton-based collection (swab, Salivette)	At baseline, three saliva samples (at waking, 30 min after waking, and 8 p.m.) on 3 consecutive days.	self-recorded collection time (diary); medication event monitoring system	cortisol only at baseline, depression were assessed both baseline and 18 months follow up, stressful life events assess at both 9 and 18 months follow up	freezer, then at −80 °C	DELFIA	nmol/L
49. Seidenfaden et al., 2017. Denmark	cross-sectional case control	adversity, stress	outcome	N = 76 (37 patients, 39 controls); 20 (patient)/19 (control) female; mean age 32.3 ± 10.7patient/31.7 ± 9.7 control; race: NR; SES: NR; BMI: NR	cotton-based collection (swab, Salivette)	Seven samples per day immediately upon awakening, at 15, 30, 45 and 60 min after awakening, at 6 pm and at 11 pm) for one day	NR		at −80 °C	ECLIA	nmol/L
50. Sin et al., 2017. United States	cross-sectional	stressors	outcome	N = 1657; 57% female; mean age 56.44 ± 12.11; race: NR; SES: 40% bachelor's or higher; BMI: NR	cotton-based collection (swab, Salivette)	4 times per day on 4 interview days (Day 2–5) (upon waking, 30-min post-waking, before lunch, and before bed)	self-recorded collection time (both log and nightly telephone interview)		at −60 °C	CLIA	nmol/L
51. Starr et al., 2017. United States	cross-sectional	depression, episodic stress	outcome and moderator	N = 241; 54% female; mean age 15.90 ± 1.09; 73.9% White, 12.2% Black, 4.1% Asian; SES: median family income $80,000–89,999; BMI: NR	cotton-based collection (swab, Salivette)	four samples a day (immediately after waking ("before you get out of bed, right after you open your eyes"), 30 min after waking, 60 min after waking, and 12 h after waking) for 2 consecutive days.	medication event monitoring system		at −20 °C	DELFIA	NS

(*continued on next page*)

Table 1 (*continued*)

Study	Design	Behavioral components	Cortisol	Sample characteristics	Saliva collection device	Salivary cortisol collection protocol	Saliva collection time recording	For longitudinal study, list follow-up sampling time points	Temperature of frozen during transportation to lab	Saliva assay in lab analysis	Unit of raw cortisol level
52. Strahler and Nater, 2018. Germany	cross-sectional	eating and drinking	outcome and mediator	N = 77; 38 female; mean age 23.9 ± 4.5; European (German); SES: all upper secondary education; BMI: 22.0 ± 2.8	passive drool	Sample collection days were timed between Tuesday and Thursday six sampling occasions each day (awakening, 30 min after awakening, 11 a.m., 2 pm, 6 pm, and 9 pm) on four consecutive days (always Tuesday to Friday to exclude influences of weekday vs. week- end on parameters of interest, see Skoluda, Linnemann and Nater, 2016)	NS, a pre-programed iPod Touch		fridge or freezer, then at −20 °C	ELISA	NR
53. Tada, 2018. Japan	experimental not RCT, longitudinal	exercise intervention	outcome	N = 61; 42 female; mean age 70.9 ± 5.9; race: NR; SES: NR; BMI: NR	cotton-based collection	one morning sample (at the beginning of the com- prehensive health promotion program, at 10 a.m. prior to start of exercise)	NR	baseline, and 6-month	NR	EIA	ug/dl
54. Uriza et al., 2021. United States	RCT	health behavior intervention	outcome	N = 48; 69% female; mean age ~55.7 ± 5.8; majority Non-Hispanic white; SES: 45–79% income ≥$80,000; BMI: range 28.1 ± 4.4 to 31.5 ± 5.3	cotton-based collection (swab, Salivette)	four times a day (waking, 30 min after waking, 4 pm, and bedtime) on two consecutive weekdays	self-recorded collection time (both log and reminder)	at baseline and at 4 months post-intervention	NS, home freezer and transport using a freezer bag	DELFIA	nmol/L

(*continued on next page*)

Table 1 (*continued*)

Study	Design	Behavioral components	Cortisol	Sample characteristics	Saliva collection device	Salivary cortisol collection protocol	Saliva collection time recording	For longitudinal study, list follow-up sampling time points	Temperature of frozen during transportation to lab	Saliva assay in lab analysis	Unit of raw cortisol level
55. Walls et al. (2020). United States	cross-sectional	smoking, eating, medication factors	outcome	N = 188; 56% female; mean age 46.3; African Indian; SES: mean income $9,862, 89% ≥high school; BMI: NR	cotton-based collection (swab, Salivette)	Four samples (upon waking, 1 h after waking, 2 h after waking, and at 8 pm) for one day	medication event monitoring system		under 20 °C, then at −80 °C	CLIA	nmol/L
56. Wong and Shobo, 2017. United States	cross-sectional	daily stressor	outcome	N = 253; 54.90% female; mean age 66.8 ± 4.96; race: NR; SES: 50.80% high school/some college; BMI: NR	NR	three sample (on awakening, 30 min post awakening, before lunch, and before bed) on three consecutive days. This study focused on the awakening cortisol level and 30 min post awakening cortisol level.	self-recorded collection time		NS (asked to store all samples in refrigerator)	CLIA	nmol/L
57. Yu et al. (2016). Netherlands	longitudinal	externalizing problems	moderator	N = 358; 153 female; mean age 15.03 ± 0.45 at wave 3; Dutch; SES: 89.5% medium/ high, 10.5% low; BMI: NR	passive drool	three morning samples (immediately after awakening (Cort0), 30 min (Cort30) and 60 min (Cort60)) on one typical weekday during the school year	self-recorded collection time	annually	fridge, then at −20 °C	ECLIA	nmol/l
58. Yu et al. (2016). Netherlands	longitudinal	depression and violent outcomes	moderator	N = 358; 153 female; mean age 15 ± 0.5 at wave 3; Dutch; SES: 89.5% medium/high, 10.5% low; BMI: NR	passive drool	Three samples per day (immediately after awakening (Cort0), 30 min (Cort30) and 60 min (Cort60) on one typical weekday during the school year	self-recorded collection time	wave 3 to 5	fridge, then at −20 °C	ECLIA	nmol/l

Notes. RCT: randomized clinical trial; NR: not reported; NS: mentioned but not specify detailed information; NA: not applicable.
SES: social economic status; BMI: body mass index; EIA: Enzyme Immunoassay or ELISA: Enzyme-Linked Immunosorbent Assay; CLIA: chemiluminescent immunoassay; DELFIA: Dissociation-enhanced lanthanide fluoroimmunoassay; RIA: radioimmunoassay; ECLIA: electrochemiluminescence immunoassay; LC-MS: Liquid Chromatography-Mass Spectrometry.

cortisol collection and data analysis in a community setting, then identified gaps and highlighted areas where further research would be useful. Due to diverse study populations, behavioral components, salivary sampling protocol and different calculations of cortisol parameters, meta-analytical calculations or meaningful summaries of results were not undertaken.

3. Results

3.1. Literature search

The literature search yielded 1733 records, with 815 identified for review after removing duplicates. Titles and abstracts were screened for inclusion/exclusion criteria by the team and 446 were excluded. The full text of the remaining studies ($N = 87$) were screened for eligibility. Fifty-two papers were deemed to meet the inclusion criteria. Then, six extra papers were identified through backward and forward tracking. A total of fifty-eight articles were included in this review. The details of the selection procedure and the reason for excluding articles at each stage are displayed in the PRISMA flow diagram (Fig. 1). Scores on the CCAT ranged from 31 to 40 (possible score range 0–40) (See Supplement Table 4S), indicating that the overall quality of the included studies was satisfactory. Most studies demonstrated rigorous study designs, adequate sample sizes, and appropriate statistical methods, which contributed to the robustness of their findings. Furthermore, while the reporting of collection protocols could be improved (more details in discussion section below), many studies provided sufficient information regarding other critical methodological aspects, such as covariate considerations and cortisol analysis parameters. Thus, the collective strengths of these studies justify our assessment of their overall quality as satisfactory.

3.2. Study characteristics

The main characteristics of reviewed studies are presented in Table 1. The behavioral components examined in these studies encompass a wide range of domains, including quality of life, functional status, stress, physical activity, and sleep, alongside mental health factors such as depression, anxiety, and stress regulation strategies. Additionally, the studies explore diverse behavioral influences like sexual function, coping mechanisms, and health interventions, while also considering factors like life stress, fatigue, and social stressors in relation to both

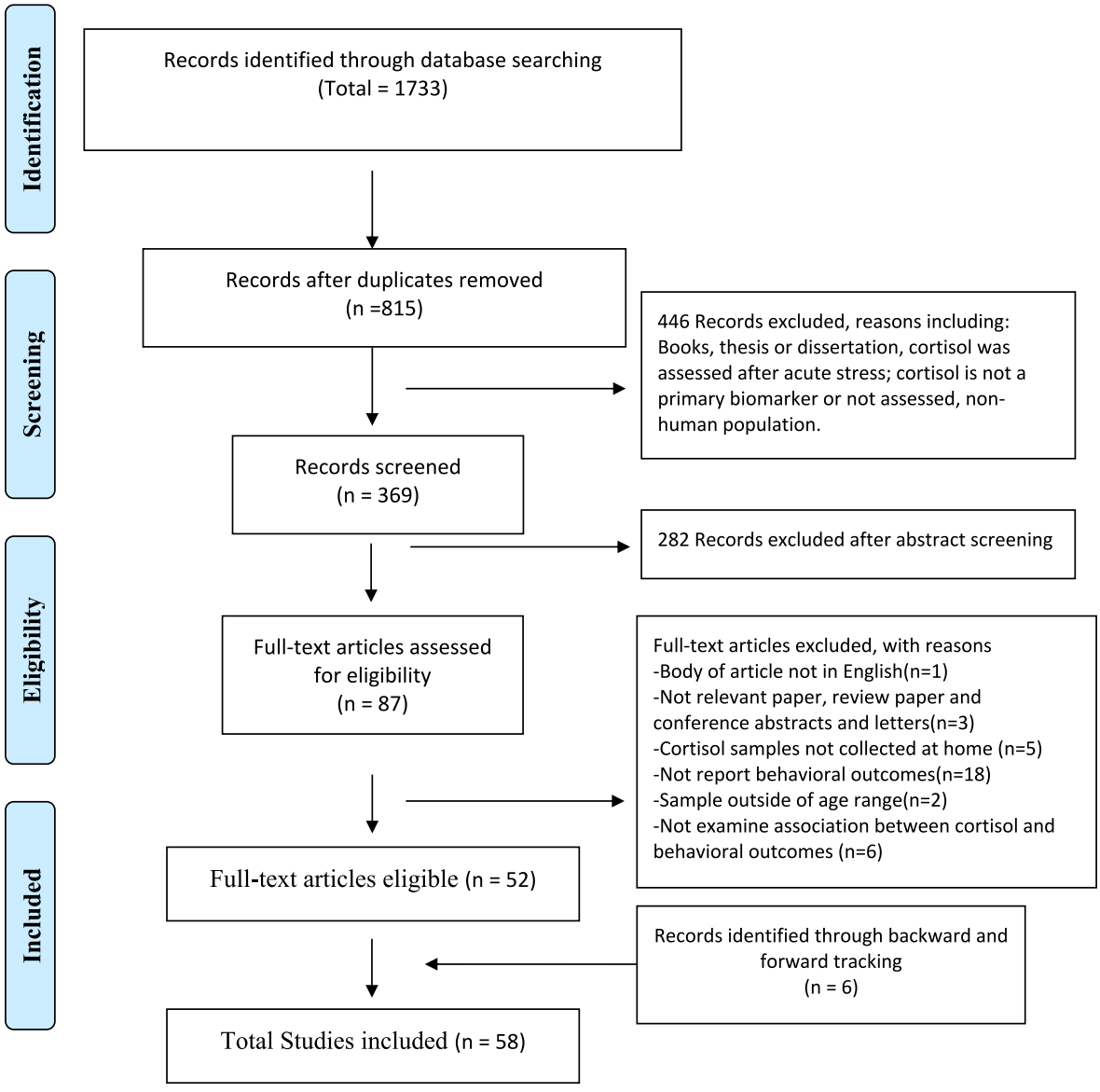


Fig. 1. Prisma Flowchart of studies included in the systematic review.

psychological and physical health outcomes. The sample size of these studies ranged from 17 to 1735. Among these studies, forty (articles 1–2, 4–9, 11–15, 17–24, 28, 30–32, 35, 36, 38, 40–43, 46, 47, 49–52, 55, 56) were cross-sectional studies utilizing retrospective data, seven (articles 3, 10, 26, 39, 48, 57, 58) were longitudinal studies with prospective data, nine (articles 25, 27, 33, 34, 37, 44, 45, 53, 54) were experimental design (seven out of nine were RCT), and two (articles 16, 29) were quasi-experimental studies. There were 30 studies from the United States (articles 1, 2, 3, 9, 11–13, 15, 17, 20–23, 29, 31, 34, 38–45, 47, 48, 50, 51, 54–56), and the remaining papers were from the Germany (4) (articles 6, 7, 19, 52), Spain (4) (articles 14, 36, 44, 46), Australia (3) (articles 8,16,3 7), Hong Kong, China (3) (articles 25, 26, 27), Canada (2) (articles 5, 33), Netherlands (2) (articles 57, 58), Denmark (1) (article 49), Finland (1) (article 28), Ireland (1) (article 18), Japan (1) (article 53), Poland (1) (article 32), Sweden (1) (article 35), Taiwan, China (1) (article 30), United Kingdom (1) (article 10), UK and Chile (1) (article 24), and Venezuela (1) (article 4).

Six studies (articles 3, 5, 15, 22, 23, 48) recruited exclusively females and two (articles 17, 46) studies included only males. Majority of the studies sampled adults while fifteen out of 58 studies sampled adolescents and youth aged 10 to 24 (articles 2, 6, 12, 17, 21, 23, 31, 36, 37, 47, 48, 51, 52, 57, 58). All of the studies reported the demographic information including age and gender, and over half (n = 36) of the studies reported race/ethnicity (articles 1, 2, 3, 5, 9, 11–15, 17, 18, 21–23, 26, 27, 29–31, 34, 35, 38, 40–42, 45–48, 51–55, 58), socioeconomic status (SES) (n = 25) (articles 4, 9, 11, 12, 20–23, 25, 26, 29, 31, 33, 40, 42, 43, 47, 50, 51, 52, 54–58), body mass index (BMI) (n = 24) (articles 2, 4, 6, 7, 9, 10, 12, 13, 14, 16, 18, 24, 31, 33, 35, 36, 38, 40, 42, 43, 47, 48, 52, 54), which were frequently reported to influence the cortisol levels. Most of the included studies (66%) used salivary cortisol as an outcome measure (n = 41) (articles 2, 3, 5–10, 12, 13, 15–20, 22, 24, 25, 27, 29, 31, 32, 34–36, 38, 40, 41, 42, 44, 45, 47, 49, 50–56), nine (articles 1, 4, 11, 23, 28, 30, 39, 46, 48) as predictor, one (article 37) as either outcome or predictor, four as mediator (article 21, 26, 33, 52), three as moderator (articles 51, 57, 58), and three (articles 14, 32, 43) used descriptive approaches.

3.3. Salivary cortisol collection protocol

Salivary cortisol collection protocols of the included studies are presented in Table 1. All the participants of the included studies collected their own saliva in the community setting (their home). The studies implemented a variety of salivary cortisol sampling protocols. Only half of studies (n = 28) (articles 2, 6, 7, 10, 11, 14, 19, 20, 22, 23, 24, 28, 30–33, 35, 38, 39, 42, 45, 47, 48, 50, 51, 55, 57, 58) clearly stated information of the salivary sampling collection protocol and method of collection, number and time of samples collected, as well as the storage prior to analysis. However, none of the studies specified whether protocol of collection and the method were successful or detailed data collection challenges.

Table 1 also gives an overview on cortisol collection protocols. Forty-four studies (articles 4, 6–8, 10–13, 15, 17–36, 38–43, 46–51, 53–55) used cotton-based saliva collection while seven studies (articles 2, 5, 16, 37, 52, 57, 58) used passive drool, five studies (articles 1, 3, 9, 44, 56) did not report collection device and two studies (articles 14, 45) directed the detailed collection protocol to another paper. There was some consistency in the collection method chosen as a large majority of the studies used cotton-based saliva collection, however, the level of procedural details also varied across different studies.

The studies sampled saliva at varying time points across days, ranging between one and ten time periods per day across one to eight days. Twenty-seven studies (articles 4, 8–10, 14, 16, 18, 20, 24–29, 32, 35, 36, 38, 41, 44, 46, 49,51, 53, 55, 57, 58) measured salivary cortisol within one day and three studies (articles 9, 53, 28) only took a single sample. For the facilitation of saliva collection time recording, among the 58 papers, 10 used a monitoring device only, 37 used self-report only, four used self-report in addition to a monitoring device, and seven studies did not specify saliva collection time was recorded. Among the studies that used self-recorded collection, researchers provided various ways to make sure participants followed the protocol, such as daily phone call or text reminder, instructing to do daily diary or log, using stamping booklet or stamping tube.

In terms of storage of the samples, there was a wide variety of approaches. Almost half of the included studies (n = 23) (articles 4, 6, 8, 9, 12, 13, 16, 19, 21, 22, 24, 27–31, 33, 35, 39, 41, 46, 48, 52, 55, 57, 58) reported the saliva samples were initially stored in a home freezer or in an iced bag or using dry ice. The majority of the studies did not mention the exact length of time the sample was stored in their home or before analysis. The temperature samples were stored at before analysis ranged from −15° to −80 °C.

In regards to laboratory tests used to measure the level of cortisol, techniques involving enzyme immunoassay (EIA) (articles 1, 9, 17, 21, 38, 39, 42, 43, 53) and enzyme-linked immunosorbent assay (ELISA) (articles 8, 14, 22, 25–27, 30–32, 35–37, 44–47, 52) yielded a total of 26 studies, making them the most frequently used techniques. The chemiluminescent immunoassay (CLIA) appeared in 15 studies (articles 3, 7, 10–12, 15, 20, 24, 33, 34, 40, 41, 50, 55, 56), while the dissociation-enhanced lanthanide fluorescent immunoassay (DELFIA) was reported in 10 studies (articles 2, 4, 5, 6, 13, 19, 23, 48, 51, 54). The electrochemiluminescent immunoassay (ECLIA) was cited in 3 studies (articles 49, 57, 58), and the radioimmunoassay (RIA) was mentioned in 2 studies (articles 28, 29). Liquid chromatography-mass spectrometry (LC-MS) was referenced only once (article 18), and one paper did not report on the technique for saliva cortisol measures (article 16).

3.4. Cortisol parameter assessment

The commonly used unit of cortisol is nmol/L or ug/dl in Table 1. Forty studies (articles 2–8, 10, 12, 15, 16, 18–20, 23–28, 32–36, 38–42, 44, 47–50, 54–58) reported cortisol values in a unit of "nmol/L″ and 10 studies (articles 1, 9, 21, 29, 30, 37, 43, 45, 46, 53) in "ug/dl". Three studies only mentioned the log transformed cortisol values (articles 11, 13, 31) and one study(article 17) reported in "ng/dl". Four studies did not report the unit. Fourteen studies (article 3, 13–15, 19, 22, 39, 47, 48, 51, 52, 56–58) did not report raw cortisol values.

Each of the included studies utilized one or several cortisol parameters (Table 2, see the studies and their references in Table 2) for their data analysis, including morning cortisol (n = 11), afternoon cortisol (n = 4), evening cortisol (n = 7), peak cortisol (n = 2), cortisol awaking response (CAR) (n = 30), diurnal cortisol slope (n = 28), cortisol awaking pulse (CAP) (n = 1), cortisol amplitude (n = 1), total daily cortisol output (n = 23), mean cortisol levels (n = 6), cortisol related ratios (n = 5), and cortisol raw values at each sampling point (n = 4). Table 2 summarized the definition and calculation methods of these cortisol parameters. In general, the calculations of the same cortisol parameters were different across studies.

3.5. Data cleaning and analysis approaches

The salivary cortisol data cleaning and analysis approaches are presented in Table 3. Although nearly half of the included studies (n = 26) (articles 1–4, 6, 12, 13, 20, 21, 23, 30, 31, 32, 34, 35, 37, 39, 42, 47, 48, 51, 55–58) briefly stated what they did to ensure data completeness, quality, and consistency, none of the studies provided the details on procedure or provided references. Only six studies (articles 12, 20, 31, 33, 54, 56) reported dealing with the impossible values. Twenty studies (articles 6, 11, 12, 15, 19, 20, 23, 25, 26, 27, 33, 34, 37, 43, 49, 50, 51, 54, 57, 58) reported dealing with the missing data, including listwise deletion, imputation use means or other not specified imputation approach or full information maximum likelihood. Eighteen studies (articles 3, 6, 12, 15, 19, 23, 26–28, 33, 37, 48–51, 54–56) reported dealing with the outliers defined as greater than either three or four

Table 2
Summary of the description of cortisol parameters in the included studies.

Parameters	Definition or Calculation for these parameters
Morning cortisol (n = 11)	Morning cortisol is defined as the cortisol level measured after waking up. When data was collected over multiple days, some studies specified averaging the results across days, while others did not. The exact wake-up times varied across studies, and not all studies reported this information. Studies that used the cortisol awakening response (CAR) as an indicator often described their morning cortisol collection methods. However, because these studies did not focus on morning cortisol as an independent indicator, their references are not included in detail here. (Basson et al., 2019; Ho, Lo et al., 2020; Holmqvist-Jamsen et al., 2017; Huynh et al., 2016; Keefe et al., 2019; Landau et al., 2021; Pace et al., 2020; Sampedro-Piquero et al., 2020; Starr et al., 2017; Tada, A., 2018; Wong and Shobo, 2017
Afternoon cortisol (n = 4)	Afternoon cortisol is defined as cortisol collected during the afternoon. The exact collection times varied across studies, and not all studies specified the precise timing. [1–5 pm (Boss et al., 2016); 12 pm and 4 pm (Keefe et al., 2019; Pace et al., 2020; Sampedro-Piquero et al., 2020)]
Evening cortisol (n = 7)	Evening cortisol is defined as being collected in the evening or at bedtime. However, the exact collection times varied across studies, and not all studies specified the time. [Basson et al., (2019); Chiang et al., (2016); bedtime at 21:30 or late afternoon at 19:30 (Ho, Lo et al., 2020); 8:00 pm (Keefe et al., 2019; Huynh et al., 2016; Landau et al., 2021; Sampedro-Piquero et al., 2020]
Peak cortisol (n = 2)	Peak cortisol is defined as highest cortisol level of each day or 30 min after waking. (Huang et al., 2020; Rosnick et al., 2016)
Cortisol Awake Response (CAR) (n = 30)	CAR is defined differently across studies. Included studies primarily used the following approaches: **1) Change in cortisol concentration**: The difference in cortisol levels between the waking sample and the second and/or third sample taken 30 min after waking, sometimes adjusted by dividing the difference by the time interval between the two measures (Fuentecilla et al., 2019; Huynh et al., 2016; Otto et al., 2018; Urizar et al., 2021; Anderson et al., 2021; Ayala-Grosso et al., 2021; Chiang et al., 2016; Darabos et al., 2019; Goldstein et al., 2017; Kristiansen et al., 2020). **2) Morning cortisol output (AUCi)**: Measurement of the area under the curve with respect to the increase (AUCi) (Abshire et al., 2018; Basson et al., 2019; Benz et al., 2019; Chian et al., 2016; Corominas-Roso et al., 2017; Herane-Vives et al., 2018; Jakuszkowiak-Wojten, 2016; Labad et al., 2018; Laures-Gore et al., 2018; Ramos-Quiroga et al., 2016; Schuler et al., 2017; Sin et al., 2017; Starr et al., 2017; Yu et al., 2016, 2019). **3) Modeling using statistical techniques**: a. Piecewise spline models, specifically linear splines, to represent the CAR (Charles et al., 2020). b. Mixed models (Garcia et al., 2017). c. CAR was assessed by calculating the area under the curve (AUC), peak, reactivity, and parameters of a regression line fitted through morning cortisol measurements (T0, T30, T60) (Doolin et al., 2017). **4) CAR increase threshold**: Defined as present when cortisol levels 30 or 45 min after awakening increased by 50% above the basal level at awakening (Ramos-Quiroga et al., 2016). **5) Delta measure**: The difference in cortisol concentration at the time of waking and 30 min post-awakening, calculated using the formula developed by Clow et al. and Kunz-Ebrecht et al. (Herane-Vives et al., 2018).
Total daily cortisol output	Total daily cortisol output is commonly defined as the area under the curve with respect to ground (AUCg) or AUC calculated over a specific test period, varies in calculation methods across studies. **1) Pruessner formula:** $AUC_G = \sum i = ln - 1(m(i+1)+mi)\ ti2$ with t_i denoting the individual time distance between measurements, m_i the individual measurement, and n the total amount of measures. Above formula is independent of the total number of measurements and can be used with any number of repetitions. This approach is independent of the total number of measurements and can accommodate any number of repetitions. For detailed information, refer to Pruessner's paper (2003). (Ayala-Grosso et al., 2021; Charles et al., 2020; Chiang et al., 2016; Darabos et al., 2019; Engert et al., 2018; Fuentecilla et al., 2019; Garcia, M.A. et al., 2021; Herane-Vives et al., 2018; Huynh et al., 2016; Johnson et al., 2020; Liu et al., 2017; Otto et al., 2018; Sampedro-Piquero et al., 2020; Schreier and Chen, 2017; Schuler et al., 2017; Seidenfaden et al., 2017; Urizar et al., 2021; Chin et al., 2017; Goldstein et al., 2017). **2)** Fekedulegn (2007) **formula**: The area under the regression line (AUR) was computed by using the estimated equation and integrating the resulting function as follows: $AUR = \int_0^{\Delta}(a+bx)dx = (a \times \Delta) + \left(\frac{b}{2} \times \Delta^2\right)$ where Δ is the time interval in minutes from the baseline measurement to the last measurement,x is the time from baseline (predictor variable), a is the intercept, and b is the slope of the fitted regression line. Used in some studies for calculating AUC. (D'Cunha et al., 2019; Walls et al., 2020). **3) Indexed by AUC with respect to ground (AUCg):** Used in studies for assessing total daily cortisol output. (Sin et al., 2017). **4) AUC for morning cortisol:** Focused on cortisol levels during the morning period.(Huang et al., 2020).
Mean cortisol over the day (n = 6)	Mean cortisol level is used 1) AUC (Ho, Lo, et al., 2020; Ho, Fong, Yau et al., 2020; Ho, Fong, Chan et al., 2020) 2) average score (Hooper, 2019; Huang et al., 2020; Morgan et al., 2017)
Change in cortisol (slope-DCS, Diurnal rhythm (DR) A type of slope, like DCS1, DCS2; Change any time point within a day, Change between days) (n = 28)	The diurnal slope is defined in various ways across studies, including the following: 1) **Modeling approaches**: Regression models are commonly used, with variations in the number of sampling time points and repeated days (Armer et al., 2018; Charles et al., 2020; Chin et al., 2017; Ho, Lo et al., 2020; Ho, Fong, Yau et al., 2020; Ho, Fong, Chan et al., 2020; Huang et al., 2020; Johnson et al., 2020; Mitchell et al., 2020; Schreier and Chen, 2017). 2) **Simple subtraction methods:** a. Diurnal rhythm dysregulation (Bitsika et al., 2017). b. Change scores: Calculated as the difference between: Wake to bedtime cortisol levels; 30 min after waking to bedtime levels; Waking to evening levels; Peak saliva levels to evening levels. (Chiang et al., 2016; Cuneo et al., 2017; Darabos et al., 2019; Engert et al., 2018; Fuentecilla et al., 2019; Huynh et al., 2016; Keefe et al., 2019; Labad et al., 2018; Landau et al., 2021; Otto et al., 2018; Pace et al., 2020; Schuler et al., 2017; Urizar et al., 2021; Walls et al., 2020). 3) **Cortisol Day Range (CDR):** Calculated as the difference between the day's highest and lowest log-transformed cortisol levels (Charles et al., 2020). 4) **Linear and quadratic slope:** Models fitted to represent the diurnal cortisol decline (Sin et al., 2017).
Cortisol amplitude, as a type of slope (n = 1)	Cortisol amplitude is calculated as the difference between the highest value of the two morning samples and the evening cortisol(Kristiansen et al., 2020)

(*continued on next page*)

Table 2 (*continued*)

Parameters	Definition or Calculation for these parameters
CAP and correlated Parameters (Benz et al., 2019) (n = 1)	CAP is calculated as the area under the curve with respect to the increase (AUCi), based on the total number of cortisol samples representing the first pulse after awakening for each individual. Since the duration of the CAP varies between individuals, this measure includes all cortisol samples from waking to the first trough. For each following pulse, the AUCi uses all cortisol samples from one trough to the next. The second measure, **amplitude**, is the difference between the peak value of the current pulse and the (detrended) mesor. The third measure, **peak-to-valley value**, is the difference between the peak value of an individual pulse and its successive trough (detrended). Finally, the duration of each pulse, in minutes, is the time from one trough to the next. For the first pulse, the duration is measured from waking to the first trough.
Cortisol related ratios (n = 5)	Cortisol-related ratios were calculated based on specific research aims. These included ratios of cortisol levels at different time points or comparisons of cortisol with other hormones: **1) Cortisol ratio** (Basson et al., 2019). **2) Ratios of cortisol at specific time points**: • Waking cortisol (M1) to cortisol 45 min after dinner (E) (M1/E). • Cortisol 30 min after waking (M2) to cortisol after dinner (E) (M2/E) (D'Cunha et al., 2019). **3) Cortisol/cortisone ratios**: Calculated at five time points to assess relative glucocorticoid levels (Doolin et al., 2017). **4) Cortisol suppression ratio in the dexamethasone suppression test (DSTR)**: Defined as the ratio of cortisol at 10:00 a.m. before dexamethasone (DEX) administration to cortisol at 10:00 a.m. after DEX administration (Labad et al., 2018). **5) Average cortisol levels**: • Morning cortisol ($Cort_{morn}$): Calculated by averaging consecutive morning saliva samples (Landau et al., 2021). • Evening cortisol ($Cort_{eve}$): Calculated similarly using evening saliva samples (Landau et al., 2021). • Morning Cort:CPR ratio ($Cort:CPR_{morn}$): Calculated by dividing untransformed $Cort_{morn}$ by untransformed CRP_{morn} values. • Evening Cort:CPR ratio ($Cort:CPR_{eve}$): Calculated in the same way using $Cort_{eve}$ and CRP_{eve} values (Landau et al., 2021).
Raw cortisol at each sampling point (n = 4)	Four studies used raw cortisol values for subsequent analysis: 1) **Multilevel modeling**: Six saliva samples were collected at specific times on a typical weekday: at waking, 30 min after waking, 2.5 h, 8 h, 12 h, and bedtime (Chandola et al., 2018). 2) **Mean** cortisol **levels**: Calculated at each sampling time point (Huang et al., 2020). 3) **Raw cortisol values**: Measured at two sampling points over three days (Wong and Shobo, 2017). 4) **Raw cortisol values**: Measured at six sampling points across four consecutive days (Strahler and Nater, 2018).

standard deviations, or winsorizing to a specific value. Twenty-nine studies (articles 3, 4, 5, 9, 10, 12–20, 25, 29, 31, 33–37, 41, 43, 44, 47, 50, 54, 56) transformed the cortisol for further interference analyses, including log-transformation (natural log, base 10 or using a specific formula) while 22 studies (articles 1, 2, 6, 7, 8, 11, 21, 24, 26, 27, 28, 26, 30, 32, 37, 38, 39, 42, 46, 48, 57, 58) used the raw value.

Statistical analysis methods also varied substantially within and between studies with most studies using more than one analysis method. Among the studies where cortisol parameters served as outcomes, the studies adopted various data analysis approaches, including intermediate statistical approaches, such as correlational analysis, t-tests methodologies, analyses of (co)variance (AN(C)OVA), regression analyses, Mann-Whitney U-tests, and advanced statistical models, e.g., generalized estimating equations, linear mixed model, multilevel growth curve modeling, hierarchical linear models (HLM).

4. Discussion

To our knowledge, this is the first systematic review focused on salivary collection for biobehavioral research conducted outside of a lab or clinical setting, and that includes a summary of the data cleaning and analysis approach taken by investigators. Fifty-eight studies were found to fulfill the inclusion and exclusion criteria for this systematic review. We found highly variable salivary sampling protocols, cortisol parameters measurements, and data cleaning and analysis approaches. Specifically, the studies showed pronounced heterogeneity in study populations, roles of cortisol with the biobehavioral measure, cortisol sampling time period, and the calculation and use of cortisol analysis parameters, and the adopted data cleaning and data analysis plan.

Key findings from the review include the following: 1) none of the studies specified whether protocol of collection and the method were successful; 2) the calculation of cortisol parameters were different across studies; 3) none of the included studies clearly stated the salivary cortisol data cleaning procedure; 4) various data analysis approaches were undertaken, and 5) only a small portion of studies (n = 4) treated cortisol as a potential mechanism. Our systematic review provides important information concerning the most frequently applied and promising study designs in this field and also raises issues of methodological considerations with regard to cortisol assessment, which should be addressed in future studies to enhance comparability of study results.

Salivary cortisol levels may vary across different populations, including individuals with different ages, genders, and races. Detailed demographics, beyond age and gender, such as race, socioeconomic status (SES), and body mass index (BMI), were not consistently reported and should be explicitly reported to inform biobehavioral research accurately. Accounting for these demographic factors is crucial for understanding the complexities of cortisol regulation and its implications for behavior and health outcomes across diverse populations. Furthermore, reporting on the influence of sociodemographic variables can enhance the generalizability and applicability of research findings, ensuring that interventions and policies are tailored to address the specific needs of different groups.

4.1. Heterogeneity in salivary cortisol sampling protocols and procedures

Our findings indicate that there is a large amount of variability in protocols used across studies for salivary collection. This review revealed the lack of a gold standard protocol for the method of salivary sampling collection, including the number and time of samples collected, the storage of saliva once it is obtained, as well as the techniques to analyze cortisol levels, which makes it difficult to compare findings and generalize the results and conclusions. As none of the studies report the success rate of obtaining samples that are suitable for analysis, it is difficult to make an informed decision on the protocols and the most successful guidelines to use in biobehavioral research.

Cortisol sampling timing varies across studies. It is not surprising that

Table 3
Data cleaning and analysis information of the included studies.

Study	Raw data preparation	Impossible values excluded	Missing data	Outlier	Analysis approach for cortisol data only	Statistical approach for the main research question
1. Abshire et al., (2018). United States	Data were checked for completeness, quality, and consistency.	NR	NR	NR	Original value	Nonparametric tests (including Mann–Whitney two-group comparisons) were used to examine the difference between implant strategy groups for continuous variables; categorical data comparisons were done using $\chi 2$ tests. A Spearman's rank correlation matrix was created to examine relationships between continuous psychological and physiological stress variables. Bivariate logistic regression modeling was used to explore relationships between physiological and psychological stress and dichotomized outcomes (high quality of life (QOL) and high functional status.
2. Anderson et al., 2021. United States	Participants were initially excluded from cortisol assays if they reported use of psychotropic or steroid-based medications (excluding birth control). Participants were excluded if there was no actigraphy or low actigraphy wear time (<80% wear time; excluded 36 participants), they did not have all saliva samples on the required days (excluded 23 participants), they did not have actigraphy data (including sleep) on the appropriate day to align with saliva (excluded 17 participants), or they did not have demographic data (excluded 1 participant); Only participants who had two complete consecutive days of data and saliva samples from the	NR	NR (Missing data was handled using mixed effect model)	NR	Original value	Multilevel linear models

(*continued on next page*)

Table 3 (*continued*)

Study	Raw data preparation	Impossible values excluded	Missing data	Outlier	Analysis approach for cortisol data only	Statistical approach for the main research question
	following morning were included in analysis					
3. Armer et al., 2018. United States	Before statistical analyses, sampling time outliers for cortisol were removed. Ranges of sampling times were determined to fit the maximum number of participants while maintaining homogeneity. Acceptable ranges were from 0400 to 0900 h for morning cortisol collection, from 1600 to 1830 h for afternoon cortisol collection, and from 2000 to 2400 h for nocturnal cortisol collection.	NR	NR	Cortisol values greater than 4 standard deviations (SD) beyond the mean for a particular time point were excluded.	log transformation (natural log)	General linear models controlling for patient age were used, and Bonferroni corrections were applied to allow for pairwise comparisons between time points. Longitudinal analyses included all 3 time points in trajectory calculation and used linear mixed-effects models with fixed slopes and participant intercept terms, Mediation model
4. Ayala-Grosso et al., 2021. Venezuela	Volunteers that failed in collecting the complete set of samples were excluded from the analysis.	NR	NR	NR	log transformation	Correlation
5. Basson et al., 2019. Canada	NR	NR	NR	NR	log transformation (log base 10)	Independent Samples t tests for group difference; simple linear regressions, ANOVA, linear mixed method
6. Benz et al., 2019. Germany	Recorded times from the MEMS caps were checked against the times written down on the protocol sheets to allow identification of discrepancies, visual inspection of raw data; Special occurrences noted on the protocol sheets like heavy exercise or sickness were used to discard individual observations.	NA	interpolation of missing values after visual inspection of raw data	winsorizing of outliers	raw data	Type III ANOVAs
7. Bernsdorf and Schwabe, 2018. Germany	NR	NR	NR	NR	raw data	Mixed model of ANOVA and correlations
8. Bitsika et al., 2017. Australia	NR	NR	NR	NR	raw data	MANOVA models
9. Boss et al., 2016. United States	NR	NR	NR	NR	log transformation (natural log)	Univariate analyses and multiple linear regression
10. Chandola et al., 2018. UK	NR	NR	NR	NR	log transformation (natural log)	Multilevel growth curve model
11. Charles et al., 2020. United States	NR	NR	Missing rate were low (this was mention for AL, to impute)	NR	raw data	Multi-level linear mixed effects model
12. Chiang et al., 2016. United States	Morning saliva samples that were considered noncompliant according to actigraphy-based estimations of wake time were also	Cortisol values greater than 60 nmol/L were set to missing	multiple imputation was conducted in order to minimize potential bias stemming from missing data. All study variables, potential confounds, and auxiliary variables were	After excluding outliers and cortisol values from noncompliant saliva samples, 217 out of the 316 participants had complete data on all computed variables of interest and covariates.	log transformed	multiple linear regressions (run both log transformed and raw values. and results reported based on raw values, using multiple imputation dataset)

(*continued on next page*)

Table 3 (*continued*)

Study	Raw data preparation	Impossible values excluded	Missing data	Outlier	Analysis approach for cortisol data only	Statistical approach for the main research question
	assigned as missing given that the estimation of CAR is sensitive to timing of samples relative to actual wake time (Dockray et al., 2008; Stalder et al., 2016). Samples were deemed non-compliant if they were provided past a 15-min window around the actigraph wake time, and around the 15- and 30-min mark after actigraphy wake time. On any given day, 43–84 adolescents provided at least one non-compliant morning sample		included in imputation models, and twenty datasets were generated.			
13. Chin et al., 2017. United States	In all cases, samples were only included for analysis if they were collected ±45 min of the scheduled collection time. This was based on our earlier work indicating we could maintain 95% or more of the data using this range and at the same time retain the normal diurnal rhythm (e.g., Janicki-Deverts et al., 2016; also, see http://www.cmu.edu/common-cold-project//combining-the-5-studies/variable-modifications.html). Samples collected outside of this window were treated as missing.	NR	NR (using missing data concept to define sufficient data, but not report how to deal with missing data)	NR	log transformation (log base 10)	hierarchical multiple linear regression with waking day cortisol AUC as outcome, and multilevel modeling waking daily cortisol slope as outcome
14. Corominas-Roso et al., 2017. Spain	NR	NR	NR	NR	log transformation (log base 10)	Pearson correlation
15. Cuneo et al., 2017. United States	NR	NR	Three participants missing afternoon cortisol values had slopes calculated from morning and bedtime samples, an approach consistent with recommendations from Kraemer et al., (2006).	Participants possessing cortisol values $\geq$ 4 SD from the mean at any time-point were also excluded (N = 1)	log transformation (natural log)	General linear models
16. D'Cunha et al., 2019. Australia	NR	NR	NR	NR	log transformation	Friedman test
17. Darabos et al., 2020. United States	NR	NR	NR	NR	log transformation	Multiple linear regression
18. Doolin et al., 2017. Ireland	NR	NR	NR	NR	log transformation	Mann-Whitney *U* test and correlation

(*continued on next page*)

Table 3 (*continued*)

Study	Raw data preparation	Impossible values excluded	Missing data	Outlier	Analysis approach for cortisol data only	Statistical approach for the main research question
19. Engert et al., 2018. Germany	NR	NR	Because salivary cortisol and experience sampling self-report data were eventually averaged acrosstwo sampling days, missing values were replaced for these repeatedlysampled variables	winsorization of outliers. non-parametricSpearman correlations in all analyses. Because Spearman's correlation limits an outlier to the value of its rank, outliers were included unwinsorized.	log transformation	Spearman Correlation, Network analysis
20. Fuentecilla et al., 2019. United States	Participants completed "five to seven daily diary interviews with a mean of 6.87 interviews (SD = 0.37) and provided saliva on average 3.99 (SD = 0.07) of the diary days. Given that waking up in the late afternoon is associated with cortisol output, the days in which participants woke up in the afternoon (n = 5 were excluded). Thus, of the total 563 valid days, 5 days were removed from the analysis, resulting in a total of 558 days.	Cortisol values were examined on a daily basis and removed if participants did not complete a daily interview, participants did not indicate time of sample collection, at least one cortisol value was over 60 nmol/L, participants were awake for less than 12 h or more than 20 h, or woke up past 12:00 noon. The entire day was excluded if there was less than 15 min or more than 60 min between the waking cortisol sample and the 30-min cortisol sample.	multilevel model can handle missing data	NR	The skew and kurtosis of each cortisol value was assessed. Due to the non- normal distribution of the cortisol levels, the natural log was calculated for all cortisol values and used for all analyses.	Multilevel modeling
21. Garcia A.F. et al., 2017. United States	To minimize the potential effects of exposure to stressful events during the sampling period, participants who were currently students were not sampled the week prior to scheduled class examinations. Inaddition, participants indicating daily hassles or exposure to stressful daily events or protocol non-compliance during sampling periods (teeth brushing, etc.) were excluded from the final analyses.	NR	NR	NR	the results based on raw score; but also use log transformed variables for modeling	Mixed effects regression model and path analysis.
22. Garcia M.A. et al., 2021. United States	NR	NR	NR	NR	NR	correlation and ANOVA
23. Goldstein et al., 2017. United States	Samples were excluded if the adolescent reported being sick; participants were only included in analyses if they had at least 1 day with all 3 samples meeting inclusion criteria.	NR	excluded participants with only one day of samples (this did not alter results)	the cortisol level was more than 3 SD above the mean for the cohort. Samples were also excluded if they fell outside the following time windows: waking samples taken more than 10 min after waking time,30-min samples taken less than 15 or more than 45 min after waking,and evening	Prior to conducting inferential statistics all individual cortisol samples were adjusted for sampling time since waking using regression	*t*-test, linear regression

(*continued on next page*)

Table 3 (*continued*)

Study	Raw data preparation	Impossible values excluded	Missing data	Outlier	Analysis approach for cortisol data only	Statistical approach for the main research question
				samples taken before 16:00 h or after 24:00 h.		
24. Herane-Vives et al., 2018. UK and Chile	NR	NR	NR	NR	raw data	ANOVA, linear regression and logistic regression
25. Ho, Lo et al., 2020. Hongkong, China	NR	NR	Missing data was handled using full information maximum likelihood under the missing-at-random assumption for the intent-to-treat analytic approach.	NR	log transformation	*t*-test, latent difference score approach
26. Ho, Fong, Yau et al., 2020. Hongkong, China	NR	NR	Missing data were handled via full information maximum likelihood under the missing-at-random assumption	Cortisol analysis was based on 838 valid samples (98.0%) after removing 17 outliers that deviated substantially (>3 standard deviations) from the mean.	raw data	structural equation modeling
27. Ho, Fong, Chan et al., 2020. Hongkong, China	NR	NR	Missing data were handled via full information maximum likelihood under the missing-at random assumption, which allowed the analysis of all of the available data under the standard intent-to-treat clinical approach	Preliminary screening of cortisol values winsorized outliers that deviated substantially (>3 SD) from the means. A total of 17, 13, 21, and 11 cortisol outliers were winsorized among the 853, 821, 761, and 678 samples at Time1, Time 2, Time 3, and Time 4, respectively.	raw data	Multigroup latent growth modeling
28. Holmqvist-Jamsen et al., 2017. Finland	NR	NR	NR	The cortisol values were winsorized to reduce the effect of potentially spurious outliers by setting outliers to 3 SD from the mean	raw data	GEE
29. Hooper, 2019. United states	NR	NR	NR only mention smoking status	NR	log transformation	Repeated measures ANOVA tested the effects of time of day, race/ethnicity, and their interactions on cortisol levels. Models controlled for income, education (continuous variables), and smoking status. Multivariate logistic regression models examined the odds of smoking relapse at the one-month follow-up by race/ethnicity, while controlling for (1) demographic covariates and (2) demographic covariates and baseline cortisol slope.
30. Huang et al., 2020. Taiwan, China	salivary cortisol data of 6 hepatocellular carcinoma patients were incomplete because the participants had forgotten to collect	NR	NR	NR		t tests to assess the difference in mean cortisol levels at each time point between the subgroups, GEE

(*continued on next page*)

Table 3 (*continued*)

Study	Raw data preparation	Impossible values excluded	Missing data	Outlier	Analysis approach for cortisol data only	Statistical approach for the main research question
	their saliva at certain time points.					
31. Huynh et al., 2016. United states	Adolescents provided three days of cortisol samples on different days of the week. Only weekday samples were included in the analyses	Samples with cortisol values over 60 (n = 14) were removed. Morning samples in which participants reported more than 30 min between sample 1 and sample 2 (n = 12) or more than 60 min between collecting sample 1 and sample 3 (n = 10) for a particular day were flagged. Analyses excluding these cases did not change the results, therefore these samples were not excluded from the final analyses. Above description is not clear that the exclusion is impossible value or treated as outlier.	NR	NR	log transformation	multiple regression
32. Jakuszkowiak Wojtenet al., 2016. Poland	Six subjects delivered incomplete sets of saliva samples and were excluded from the analysis	NR	NR	NR	raw data	Chi square; Pearson correlation
33. Johnson et al., 2020. Canada	NR	Cortisol values greater than 4 standard deviations above the sample mean for that timepoint were removed	The variables used in the analysis were examined for missing data using the MissMech package in R. The pattern of missing data as well as a non-significant Little's MCAR (missing completely at random) tests indicated that there was not enough evidence to reject the MCAR assumptions. Missing data were imputed using a multiple imputation with predictive mean matching method in the MICE package	Cortisol values greater than 4 standard deviations above the sample mean for that timepoint were removed.	To adjust for the non-normal distributions of the raw cortisol values, all values were transformed using a natural log transformation and the transformed values were used for all analyses	multilevel structural equation modeling framework
34. Keefe et al., 2018. United states	The average subject had 94.3% of pre-treatment measurements completed (mean = 11.3), and 92.8% of post-treatment measurements completed (mean = 11.1).	NR	All collected awakening and post-awakening measurements were used in the model, under the assumption that any given unobserved measurement was missing at random	NR	log transformation (log base 10)	mixed model
35. Kristiansen et al., 2020. Sweden	Only if there was a congruency between either exact time entries in the diary or event entries in the ECG with the movement pattern and	NR	NR	NR	log transformation (natural log)	Mann–Whitney *U* test

(*continued on next page*)

Table 3 (*continued*)

Study	Raw data preparation	Impossible values excluded	Missing data	Outlier	Analysis approach for cortisol data only	Statistical approach for the main research question
	increased heart rate (indicating awakening) were the morning samples included in the analysis. Based on this strict selection, 83% of the patients had acceptable cortisol samples and were included in the analysis (167 out of 201 individuals). Individuals with diabetes had a lower rate of successful sampling than controls (80% versus 88%), mostly due to low glucose levels in the morning that impeded cortisol sampling in some cases. Children had a lower rate of successful sampling than adults (80% versus 91%).					
36. Labad et al., 2018. Spain	NR	NR	NR	NR	Cortisol values were transformed to approximate a normal distribution, as suggested by recent expert consensus guidelines. The following power transformation was used: X' = (X0.26 − 1)/0.26	Pearson correlations (and Spearman correlations, when needed), GLM, Three separate multiple regression analyses
37. Landau et al., 2021. Australia	Consecutive morning saliva samples were averaged to create average Cortmorn and average CRPmorn values; evening saliva samples were calculated the same to create average Corteve and average CRPeve values. Morning Cort:CRP ratio (Cort: CRPmorn) was calculated by dividing untransformed Cortmorn values by untransformed CRPmorn values, and evening Cort: CRP ratio (Cort: CRPeve) was calculated in the same manner with Corteve and CRPeve values. Diurnal cortisol slopes were calculated by taking the difference	NR	Out of the 122 intention to treat sample at T1, a total of 107 participants (87.7% of the total sample) provided full or partial T3 (follow-up) data. Multiple Imputation was performed on the entire dataset. Predictive mean matching imputation was used for quantitative continuous data (e.g., saliva, questionnaires), and logistic regression was used for categorical data. Out of the 122 intentions to treat sample at T1, a total of 107 participants (87.7% of the total sample) provided full or partial T3 (follow-up) data. Little's Missing Completely at Random (MCAR) tests were used to test for patterns of missingness in the data prior to imputation. Little's MCAR results	Outliers > ±3 standard deviations (SD) above/ below the mean were investigated by log-transforming the values (ref to Laudau2019)., Saliva data outliers (n = 5 at T1 and n = 4 at T3) were winsorized to 0.01 μg/dL for cortisol values Outliers for questionnaire variables were not adjusted (as in Blake et al., 2016, 2017a, 2017b, 2018) because research has shown psychological variables are typically positively skewed in non-clinical populations with outliers to be expected due to the self-report nature of these measures.	raw data and log transformation (natural log)	Simple regression analyses; A series of analyses of covariance (ANCOVA)A series of multivariate linear and logistic regression analyses

(*continued on next page*)

Table 3 (*continued*)

Study	Raw data preparation	Impossible values excluded	Missing data	Outlier	Analysis approach for cortisol data only	Statistical approach for the main research question
	between natural-log transformed Cortmorn and Corteve values divided by time between sample collection. Saliva data outliers (n = 5 at T1 and n = 4 at T3) were winsorized to 0.01 μg/dL for cortisol values and 0.01 pg/mL for CRP values.		indicated non-significance (statistics not shown) suggesting MCAR and acceptability to multiple imputation. Multiple imputation was performed on the entire dataset using the 'Multiple Imputation by Chained Equations' (mice) package in RStudio with all variables included in the present study. Predictive mean matching imputation, considered more robust for use with non-normal data was used for quantitative continuous data (e.g., saliva, questionnaires), and logistic regression was used for categorical data. *Percentage of variables missing and other missingness assumptions are presented in* Supplemental Table 1.			
38. Laures-Gore et al., 2019. United States	NR	NR	NR	NR	raw data	Repeated measures ANOVA
39. Liu et al., 2017. United States	A saliva sample was invalid if: 1) the caregiver was awake for less than 12hr or greater than 20hr (n = 14), or 2) the caregiver woke up after 12pm (n = 0), or 3) for cortisol assay specifically, there was a greater than 10 nmol/L rise between the second (30 min after getting out of bed) and third sample (before lunch) (n = 11), or 4) the recorded collection time between the first (upon wakeup) and second sample (30 min after getting out of bed) is either less than 15min or greater than 60 min (n = 99).	NR	NR	NR	raw data	growth curve models
40. Mitchell et al. (2020). United States	NR	NR	Not specify, only mention to include who provide complete data.	NR	descriptive	Hierarchical general linear modeling
41. Morgan et al. (2017). United States	NR	NR	NR	NR	Cortisol Modeling: $Y_{ij} = f(t_{ij}) + \alpha_i + \epsilon_{ij}$ Yij: the log-transformed cortisol value for the jth sample from the ith respondent; tij: the time at which the sample was taken αi: a respondent-	Unadjusted and adjusted multiple linear regression. These models were fit using the survey weights distributed with the data set that accounts for differential probabilities of selection and

(*continued on next page*)

Table 3 (*continued*)

Study	Raw data preparation	Impossible values excluded	Missing data	Outlier	Analysis approach for cortisol data only	Statistical approach for the main research question
					level deviation from the mean with distribution N(0, σ2α). The error term ϵij is assumed to be independent with distribution N(0,σ2), log transformed average cortisol levels	differential nonresponse. Design-based standard errors were obtained using the linearization method46 as implemented in the Stata statistical software package version 13.1.47
42. Otto et al. (2018). United States	Days were excluded from the calculation of the cortisol indices if (1) saliva collection time stamps were missing, (2) the participant woke up after 12 pm, (3) the participant was awake <12 h or >20 h, or (4) if there was an indication of non-compliance with the saliva collection protocol such that <15 or >60 min elapsed between the first two measurements (Stawski, Cichy, Piazza and Almeida, 2013). The analytic sample sizes were 46 participants for DCS and 43 participants for CAR and AUCg.	NR	NR	NR	raw data	linear regression
43. Pace et al. (2021). United States	Success was defined as obtaining biomarker data from ≥85% of samples per protocol. Saliva concentrations of cortisol were averaged across collection days in morning, afternoon, or evening because an effect of day was not expected; We first examined biomarker and HRQOL variables by computing means and their standard errors by biomarker and time point (for cortisol only).	NR	not specify, only mentioned 96% and 92% of saliva samples were collected from survivors and caregivers	NR	Data that were not normally distributed (Shapiro–Wilk test) were naturallog transformed before any inferential testing	Examined the association between biomarker variables (CRP, AM cortisol, PM cortisol, and cortisol slope) and HRQOL domains by computing partial and semi partial correlation coefficients controlling for body mass index (BMI) and chemotherapy treatment (survivors) and Pearson product-moment correlation coefficients (caregivers). A Spearman's rank correlation coefficient was computed instead for associations where one or both outcomes were not normally distributed.
44. Ramos-Quiroga et al. (2016). Spain	NR	NR	NR	NR	Because the distribution of cortisol values was positively skewed, these data have been base-10	Chi-square test (χ2); repeated measures ANCOVA; Spearman-Rho correlations

(*continued on next page*)

Table 3 (*continued*)

Study	Raw data preparation	Impossible values excluded	Missing data	Outlier	Analysis approach for cortisol data only	Statistical approach for the main research question
					logarithmically transformed prior to any further analyses.	
45. Rosnick et al. (2016). United States	NR	NR	NR	NR	NR	GEE analysis was conducted to examine the between treatment group difference in peak cortisol change over time from pre- to post-augmentation.
46. Sampedro-Piquero et al. (2020). Spain	NR	NR	NR	NR	raw data	RM ANOVA and MANOVA, Pearson correlation
47. Schreier and Chen, 2017. United States	Cortisol data were unavailable for 17 adolescents who did not return useable samples. These adolescents did not differ from participants who returned useable samples with respect to age, BMI, chronic and acute stress ratings, ethnicity, and family income (ps > 0.10) but were more likely to be female ($\chi 2$ (1) = 6.184, p = .013). On average, adolescents completed 5.47 (±1.03) out of the 6 days.	NR	NR	NR	log transformation	hierarchical multiple regression analyses
48. Schuler et al., 2017. United States	Before testing hypotheses, cortisol data were inspected for outliers.	NR	NR	Four criteria were used to identify outliers, namely, (1) standardized cortisol values were bigger than three standard deviations from the mean; (2) adolescent participants were ill on a given sampling day (e.g., any illness symptoms indicated in the diary); (3) blood contamination (e.g., from cuts in the mouth); and (4) saliva samples deemed to be collected nonadherent to sampling instructions (i.e., participants ate or drank before collecting saliva samples or saliva samples were collected outside the instructed time)	raw data	a hierarchical multiple regression
49. Seidenfaden et al. (2017). Denmark	**NR**	NR	For series of samples with more than one sample missing, the AUC was not computed. If only one sample was missing, values were replaced by the mean of the two adjacent values, or, if the missing value were either the awakening or 11 pm sample, by the mean of the full sample for that time point.	Before computations, extreme values in each group for each time point (outside the 99th percentile) were excluded (30 out of a total of 658 determinations).	NR	repeated measures ANOVA

(*continued on next page*)

Table 3 (*continued*)

Study	Raw data preparation	Impossible values excluded	Missing data	Outlier	Analysis approach for cortisol data only	Statistical approach for the main research question
50. Sin et al. (2017). United States	NR	NR	Models were estimated using full information maximum likelihood estimation in SAS 9.4 PROC MIXED, which makes use of all available data in the estimation of parameters and can flexibly handle missing data	cortisol samples were excluded where the cortisol level was >60 nmol/L (1.46%), the time stamp was missing (1.28%), or the lunch sample was ≥10 nmol/L more than the 30-min post-waking sample (suggesting that participants ate before collecting their saliva, 1.82%). Further, cortisol samples were excluded from days when participants woke before 4 a.m. (3.14%) or after 12 pm (0.67%), or days when <15 or >60 min elapsed between the first two samples (indicators of noncompliance that influence assessment of the awakening response, 9.74%).	log transformation (natural log)	Multilevel modeling
51. Starr et al. (2017). United States	Of the original sample of 241, 12 were excluded from cortisol procedures for medical reasons, and 18 declined to participate in cortisol procedures or failed to return samples, leaving 211 participants with samples that were assayed. careful measures were taken to exclude values that might not accurately represent the CAR.	NR	Cortisol values at each sampling time were winsorized to correct for extreme outliers (>3SD; 5 data points for waking, 2 for +30 min, and 5 for +60 min)	Both variables were winsorized to 3 SD to correct for outliers	NR	Moderation analysis, linear regression
52. Strahler and Nater, 2018. Germany	NR	NR	NR	NR	NR	Hierarchical linear models
53. Tada, 2018.Japan	NR	NR	NR	NR	NR	Baseline data on POMS-SF and salivary biomarkers of both groups were compared using the Mann–Whitney *U* test. Wilcoxon signed-rank tests were used to compare differences in the groups' scores at baseline and 6-month follow-up. Correlations between changes in cortisol level and in POMS-SF "fatigue" score were assessed using Pearson correlation coefficients
54. Urizar et al. (2021). United States	veraging the cortisol values across the two saliva collection days at each study time point.	no impossible values based on no outliers	Missing cortisol samples for a particular collection day were estimated by using the participant's second day	No cortisol outliers (defined as being three standard deviations from the mean for each cortisol index) were identified in the current investigation;	log transformation (log base 10)	Pearson correlation, mixed effect linear model

(*continued on next page*)

Table 3 (*continued*)

Study	Raw data preparation	Impossible values excluded	Missing data	Outlier	Analysis approach for cortisol data only	Statistical approach for the main research question
			sample for that timepoint.			
55. Walls et al., 2020. United States	Single Sample Values were examined for possible measurement error	NR	NR	Single Sample Values were examined for possible measurement error and any outlier values that required deeper examination; We also performed separate t-tests to examine the influence of the largest discrepancies (i.e. outliers and extreme cases) on cortisol indices.	raw data	Pearson correlation, *t*-test
56. Wong and Shobo, 2017. United States	A set of criteria was used to determine the analytic sample. 235 did not provide saliva samples and were dropped. Individuals who did not follow the cortisol collection procedures (n = 10) and those who did not provide complete data on medication use (n = 79) were dropped.	Following the Winsorization statistical approach (Dixon and Yuen, 1974), salivary cortisol values higher than 60 nmol/L were recoded as 61 to minimize the influence of extreme outliers.	NR	Following the Winsorization statistical approach (Dixon and Yuen, 1974), salivary cortisol values higher than 60 nmol/L were recoded as 61 to minimize the influence of extreme outliers.	log transformation	Two-level multilevel models
57. Yu et al. (2016). Netherlands	All samples were checked for correctness of sampling. Cases were excluded from analyses if the cortisol data were of incorrect sampling time, unclear how it was sampled (i.e., not registered), contaminated (e.g., by smoking or brushing teeth), or of extreme values (i.e., >3 SD from average	NR	Reported attrition and little's MCAR test; applied Full Information Maximum Likelihood (FIML) in Mplus for the model estimations	NR	raw data	Multiple regression models incorporating latent growth models
58. Yu et al. (2019). Netherlands	NR	analyses of all variables used in this study revealed a normed χ^2 (χ^2/df) of 1.04, which indicates that the pattern of the missing data was not materially different from a missing completely at random pattern	NR	NR	raw data	mixed model

Notes. NR: not reported.
GEE: generalized estimating equations; ANOVA: Analyses of variance.

the timing of saliva collection when conducted at home by patients themselves is one of the most significant challenges. However, accurate timing is crucial for measuring diurnal cortisol rhythms, particularly the cortisol awakening response (CAR), which can be easily disrupted by delays or inconsistencies in sample collection (Adam, 2009). Without direct supervision, patient compliance and adherence to the exact timing of collections can vary, introducing potential biases and affecting the reliability of the data. Given the community setting of these studies, where participants may have different levels of health literacy and access to resources, ensuring compliance with timing protocols becomes even more challenging. Addressing these barriers in the protocol is essential for improving data reliability and ensuring that home-based saliva collection accurately reflects cortisol patterns in naturalistic environments. Additionally, although Adam (2009) believes passive drool is the gold standard for saliva collection, cotton-based saliva collection is suitable and more feasible for saliva collection in community settings as more than two thirds of included studies using cotton-based saliva collection.

A protocol for collection should be explicit and detailed to ensure comparable and replicable collection methods within and between

studies and to avoid contamination. We recommend investigators include instructions for participants and a training session with modelling of the sampling method. Instructions should also include restrictions of no food, drink, vigorous exercise, brushing teeth 30 min to 1 h prior to sampling if possible. Additionally, we recommend that participants record negative life events, health status, and any medication taken so this information can be reviewed and compared to the cortisol findings. It has repeatedly been shown that cortisol concentrations vary considerably between and within individuals over time (Adam et al., 2017; Hellhammer et al., 2007; Strahler et al., 2017). Although included studies have reported various approaches to monitor compliance to saliva collection and time recording, none of the studies were transparent with the whole process of the saliva collection and they did not provide a detailed protocol or report items like the success rate in obtaining the samples. Sampling protocols should be transparent and sample on several (at least two) consecutive days with multiple samples on each day while recording accurate collecting time to enhance reliability (Adam and Kumari, 2009).

As a result of our findings, we recommend that future research provides full details on the methodological choices, protocol of collection and storage, and success rate of obtaining salivary cortisol samples in the community settings. Potential strategies to record accurate time, especially among different age groups in community settings are also needed. Being more transparent will enable the establishment of a gold standard within the field. This would further inform research and ensure better, more consistent practice, leading to more robust and more comparable findings.

4.2. Cortisol parameters calculation and heterogeneity

Adam (2009) recommends reporting CAR, slope, and AUC parameters to enhance inter-study comparability of basal cortisol concentrations. This is especially important when other factors, such as study design and sample characteristics differ as much as in our sample of included studies. Of course, in order to report these parameters, sampling protocols need to be adapted accordingly. We recommend specifying the calculation of cortisol parameters and what these parameters reflect in order to optimize the reliability and validity of the cortisol measure.

4.3. Salivary cortisol data cleaning procedures

It is important for studies to be transparent about dealing with cortisol data, such as impossible values, missing data and outliers, before conducting any inferential statistics. For example, researchers have come to consensus on data cleaning procedure before reporting CAR (Stalder et al., 2016), including account for positively skewed of cortisol data, appropriate transformation techniques, addressing some extreme outlying cortisol values (including how to define them and dealing with them). Often these data are not missing at random (i.e., they are indicative of very low or very high values), which poses a unique challenge for data analyses. When dealing with the impossible/missing data/-outlier, included studies all reported different strategies. There is no recommendation for a promising approach for high out of range samples and presents an important area of future research effort.

4.4. Salivary data analysis approaches

Data analysis is an important step in interpreting salivary cortisol results and determining the significance of cortisol levels in different populations and conditions. It is important to consider the most appropriate data analysis method for each study design and population to ensure accurate and reliable results. Our findings proposed various data analysis methods to better interpret salivary cortisol data. Cortisol served as different roles, including outcome, predictor, mediator or moderator in the included studies. We call for future research to study the biological mechanism in behavioral research by considering cortisol as a potential mechanism.

None of the papers in this review applied statistical algorithms, such as machine learning algorithms, which have been found to be effective in detecting patterns in cortisol data (Riis et al., 2020). These algorithms can be used to identify changes in cortisol levels over time and predict cortisol levels based on specific environmental or physiological factors. Advanced statistical methods, such as mixed-effects models, can also account for the repeated measures nature of cortisol data over time and allow for the examination of between- and within-subjects effects.

To facilitate replication of research and to inform future studies, we urge researchers to make their data openly available whenever possible, or to at least provide descriptive statistics (e.g., mean and standard deviation) of baseline cortisol concentration (or by time point) to be comparable with participants with similar characteristics. As we navigate the nuances of cortisol's role in behavior, future research should prioritize addressing these methodological considerations to ensure robust and meaningful findings in biobehavioral research.

4.5. Limitations

Several methodological limitations need to be considered when interpreting the findings of this review. In order to be as inclusive as possible the search criteria were very broad as we did not seek to specify a behavioral component. This led to the great variability of our findings. Second, the search was conducted for publications in English and may have missed international studies with important implications. Finally, despite consultation with a medical librarian and hand searching of references, there may be missed publications in our search. Despite these limitations, our findings underscore the intricate interplay between cortisol dynamics and behavioral outcomes, shedding light on the complexities of biobehavioral research. The implications for this field are substantial, as the diverse approaches to cortisol assessment can significantly influence study outcomes and interpretations.

5. Conclusion

Inclusion of salivary cortisol as a biomarker in biobehavioral research is promising for understanding HPA function dynamics non-invasively. Future work is needed to elucidate a gold standard for salivary collection protocol, salivary parameter assessment and the reported data clean procedures and analysis plan in the community settings. We offer some recommendations for future studies ensuring the use of comparable study protocols and data clean and analysis approaches. Following these recommendations will ensure that future research is clear, replicable, and concise, with strong scientific rigor. The results will be generalizable and further research will be enabled to fill the knowledge gaps by conducting meta-analysis to better quantify the relationship between cortisol and behavioral components.

CRediT authorship contribution statement

Fanghong Dong: Writing – original draft, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Justine S. Sefcik:** Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Elizabeth Euiler:** Writing – review & editing, Formal analysis, Conceptualization. **Nancy A. Hodgson:** Writing – original draft, Validation, Supervision, Conceptualization.

Funding

This study was supported by the National Institutes of Health grant NR01522601. The funding resources had no further role in writing of this paper. J. S. Sefcik was funded by the National Institute of Nursing Research of the National Institutes of Health [K23 NR018673].

Declaration of competing interest

The Author(s) declare(s) that there is no conflict of interest.

Acknowledgements

We would like to thank Danielle Katsev, Amita Panicker, Esther Gordon, and Mohammed Alqurashi for their assistance.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbih.2024.100936.

Data availability

No data was used for the research described in the article.

References

Abshire, M., Bidwell, J.T., Page, G., Budhathoki, C., Davidson, P.M., Russell, S.D., Han, H.R., Desai, S., Dennison Himmelfarb, C., 2018. Physiological and psychological stress in patients living with a left ventricular assist device. Am. Soc. Artif. Intern. Organs J. 64 (6), e172–e180. https://doi.org/10.1097/MAT.0000000000000847 [Article].

Adam, E.K., Kumari, M., 2009. Assessing salivary cortisol in large-scale, epidemiological research. Psychoneuroendocrinology 34 (10), 1423–1436.

Adam, E.K., Quinn, M.E., Tavernier, R., McQuillan, M.T., Dahlke, K.A., Gilbert, K.E., 2017. Diurnal cortisol slopes and mental and physical health outcomes: a systematic review and meta-analysis. Psychoneuroendocrinology 83, 25–41.

Anderson, T., Corneau, G., Wideman, L., Eddington, K., Vrshek-Schallhorn, S., 2021. The impact of prior day sleep and physical activity on the cortisol awakening response. Psychoneuroendocrinology 126, 105131.

Armer, J.S., Clevenger, L., Davis, L.Z., Cuneo, M., Thaker, P.H., Goodheart, M.J., Bender, D.P., Dahmoush, L., Sood, A.K., Cole, S.W., Slavich, G.M., Lutgendorf, S.K., 2018. Life stress as a risk factor for sustained anxiety and cortisol dysregulation during the first year of survivorship in ovarian cancer. Cancer 124 (16), 3401–3408. https://doi.org/10.1002/cncr.31570 [Article].

Ayala-Grosso, C., Torrico, F., Ledezma-Ruiz, M., Busolo-Pons, M., 2021. Chronic stress in cognitive processes: cortisol dynamic range of secretion is associated with perception of unsafety environment in a Venezuelan population. J. Alzheim. Dis. 1–14.

Basson, R., O'Loughlin, J.I., Weinberg, J., Young, A.H., Bodnar, T., Brotto, L.A., 2019. Dehydroepiandrosterone and cortisol as markers of HPA axis dysregulation in women with low sexual desire. Psychoneuroendocrinology 104, 259–268. https://doi.org/10.1016/j.psyneuen.2019.03.001.

Bellagambi, F.G., Lomonaco, T., Salvo, P., Vivaldi, F., Hangouët, M., Ghimenti, S., Biagini, D., Di Francesco, F., Fuoco, R., Errachid, A., 2020. Saliva sampling: methods and devices. An overview. TrAC, Trends Anal. Chem. 124, 115781.

Benedict, C., 2013. Biobehavioral mechanisms. In: Gellman, M.D., Turner, J.R. (Eds.), Encyclopedia of Behavioral Medicine. Springer, New York, NY. https://doi.org/10.1007/978-1-4419-1005-9_153.

Benz, A., Meier, M., Mankin, M., Unternaehrer, E., Pruessner, J.C., 2019. The duration of the cortisol awakening pulse exceeds sixty minutes in a meaningful pattern. Psychoneuroendocrinology 105, 187–194. https://doi.org/10.1016/j.psyneuen.2018.12.225 [Article].

Bernsdorf, M., Schwabe, L., 2018. Cortisol response to awakening in prepubertal children and adults: magnitude and variability. Psychophysiology 55 (12), e13273.

Bitsika, V., Sharpley, C.F., Andronicos, N.M., Agnew, L.L., 2017. What worries parents of a child with Autism? Evidence from a biomarker for chronic stress. Res. Dev. Disabil. 62, 209–217. https://doi.org/10.1016/j.ridd.2017.02.003.

Boss, L., Branson, S., Cron, S., Kang, D.H., 2016. Biobehavioral examination of religious coping, psychosocial factors, and executive function in homebound older adults [Article]. Religions 7 (5). https://doi.org/10.3390/rel7050042. Article 42.

Chandola, T., Rouxel, P., Marmot, M.G., Kumari, M., 2018. Retirement and socioeconomic differences in diurnal cortisol: longitudinal evidence from a cohort of British civil servants. J. Gerontol. B Psychol. Sci. Soc. Sci. 73 (3), 447–456. https://doi.org/10.1093/geronb/gbx058.

Charles, S.T., Mogle, J., Piazza, J.R., Karlamangla, A., Almeida, D.M., 2020. Going the distance: the diurnal range of cortisol and its association with cognitive and physiological functioning. Psychoneuroendocrinology 112. https://doi.org/10.1016/j.psyneuen.2019.104516 [Article].

Chiang, J.J., Tsai, K.M., Park, H., Bower, J.E., Almeida, D.M., Dahl, R.E., Irwin, M.R., Seeman, T.E., Fuligni, A.J., 2016. Daily family stress and HPA axis functioning during adolescence: the moderating role of sleep. Psychoneuroendocrinology 71, 43–53. https://doi.org/10.1016/j.psyneuen.2016.05.009 [Article].

Chin, B., Murphy, M.L., Janicki-Deverts, D., Cohen, S., 2017. Marital status as a predictor of diurnal salivary cortisol levels and slopes in a community sample of healthy adults. Psychoneuroendocrinology 78, 68–75.

Corominas-Roso, M., Armario, A., Palomar, G., Corrales, M., Carrasco, J., Richarte, V., Ferrer, R., Casas, M., Ramos-Quiroga, J.A., 2017. IL-6 and TNF-α in unmedicated adults with ADHD: relationship to cortisol awakening response. Psychoneuroendocrinology 79, 67–73. https://doi.org/10.1016/j.psyneuen.2017.02.017 [Article].

Crowe, M., Sheppard, L., 2011. A general critical appraisal tool: an evaluation of construct validity. Int. J. Nurs. Stud. 48 (12), 1505–1516.

Cuneo, M.G., Schrepf, A., Slavich, G.M., Thaker, P.H., Goodheart, M., Bender, D., Cole, S.W., Sood, A.K., Lutgendorf, S.K., 2017. Diurnal cortisol rhythms, fatigue and psychosocial factors in five-year survivors of ovarian cancer. Psychoneuroendocrinology 84, 139–142. https://doi.org/10.1016/j.psyneuen.2017.06.019 [Article].

D'Cunha, N.M., McKune, A.J., Isbel, S., Kellett, J., Georgousopoulou, E.N., Naumovski, N., Garrido, S., 2019. Psychophysiological responses in people living with dementia after an art gallery intervention: an exploratory study. J. Alzheim. Dis. 72 (2), 549–562. https://doi.org/10.3233/JAD-190784.

Darabos, K., Hoyt, M.A., 2020. Emotional processing coping methods and biomarkers of stress in young adult testicular cancer survivors. J. Adolesc. Young Adult Oncol. 9 (3), 426–430. https://doi.org/10.1089/jayao.2019.0116.

Doolin, K., Farrell, C., Tozzi, L., Harkin, A., Frodl, T., O'Keane, V., 2017. Diurnal hypothalamic-pituitary-adrenal axis measures and inflammatory marker correlates in major depressive disorder [Article]. Int. J. Mol. Sci. 18 (10). https://doi.org/10.3390/ijms18102226. Article 2226.

Engert, V., Kok, B.E., Puhlmann, L.M.C., Stalder, T., Kirschbaum, C., Apostolakou, F., Papanastasopoulou, C., Papassotiriou, I., Pervanidou, P., Chrousos, G.P., Singer, T., 2018. Exploring the multidimensional complex systems structure of the stress response and its relation to health and sleep outcomes. Brain Behav. Immun. 73, 390–402. https://doi.org/10.1016/j.bbi.2018.05.023.

Fuentecilla, J.L., Liu, Y., Huo, M., Kim, K., Birditt, K.S., Zarit, S.H., Fingerman, K.L., 2019. Midlife adults' daily support to children and parents: implications for diurnal cortisol [article]. J. Aging Health. https://doi.org/10.1177/0898264319863994.

Garcia, A., Wilborn, K., Mangold, D., Garcia, A.F., Mangold, D.L., 2017. The cortisol awakening response mediates the relationship between acculturative stress and self-reported health in Mexican Americans. Ann. Behav. Med. 51 (6), 787–798. https://doi.org/10.1007/s12160-017-9901-5.

Garcia, M.A., Li, X., Allen, P.A., Delahanty, D.L., Eppelheimer, M.S., Houston, J.R., Johnson, D.M., Loth, F., Maleki, J., Vorster, S., 2021. Impact of surgical status, loneliness, and disability on interleukin 6, C-reactive protein, cortisol, and estrogen in females with symptomatic type I chiari malformation. Cerebellum 1–15.

Goldstein, B.L., Perlman, G., Kotov, R., Broderick, J.E., Liu, K., Ruggero, C., Klein, D.N., 2017. Etiologic specificity of waking Cortisol: links with maternal history of depression and anxiety in adolescent girls [Article]. J. Affect. Disord. 208, 103–109. https://doi.org/10.1016/j.jad.2016.08.079.

Hellhammer, D.H., Wüst, S., Kudielka, B.M., 2009. Salivary cortisol as a biomarker in stress research. Psychoneuroendocrinology 34 (2), 163–171.

Hellhammer, J., Fries, E., Schweisthal, O., Schlotz, W., Stone, A., Hagemann, D., 2007. Several daily measurements are necessary to reliably assess the cortisol rise after awakening: state-and trait components. Psychoneuroendocrinology 32 (1), 80–86.

Herane-Vives, A., de Angel, V., Papadopoulos, A., Wise, T., Chua, K.C., Strawbridge, R., Castillo, D., Arnone, D., Young, A.H., Cleare, A.J., 2018. Short-term and long-term measures of cortisol in saliva and hair in atypical and non-atypical depression [Article]. Acta Psychiatr. Scand. 137 (3), 216–230. https://doi.org/10.1111/acps.12852.

Higgins, J.P.T., Thomas, J., Chandler, J., Cumpston, M., Li, T., Page, M.J., Welch, V.A., 2019. Cochrane Handbook for Systematic Reviews of Interventions. John Wiley & Sons.

Ho, R.T.H., Fong, T.C.T., Chan, W.C., Kwan, J.S.K., Chiu, P.K.C., Yau, J.C.Y., Lam, L.C.W., 2020a. Psychophysiological effects of dance movement therapy and physical exercise on older adults with mild dementia: a randomized controlled trial. J. Gerontol. B Psychol. Sci. Soc. Sci. 75 (3), 560–570. https://doi.org/10.1093/geronb/gby145 [Article].

Ho, R.T.H., Lo, H.H.M., Fong, T.C.T., Choi, C.W., 2020b. Effects of a Mindfulness-based Intervention on diurnal cortisol pattern in disadvantaged families: a randomized controlled trial. Psychoneuroendocrinology 117, 7. https://doi.org/10.1016/j.psyneuen.2020.104696.

Ho, R.T., Fong, T.C., Yau, J.C., Chan, W.C., Kwan, J.S., Chiu, P.K., Lam, L.C., 2020c. Diurnal cortisol slope mediates the association between affect and memory retrieval in older adults with mild cognitive impairment: a path-analytical study. Front. Aging Neurosci. 12, 35.

Hodgson, N.A., Granger, D.A., 2013. Collecting saliva and measuring salivary cortisol and alpha-amylase in frail community residing older adults via family caregivers. J. Vis. Exp. 82, e50815. https://doi.org/10.3791/50815.

Holmqvist-Jämsén, S., Johansson, A., Santtila, P., Westberg, L., von der Pahlen, B., Simberg, S., 2017. Investigating the role of salivary cortisol on vocal symptoms. J. Speech Lang. Hear. Res. 60 (10), 2781–2791.

Hooper, M.W., 2019. Racial/ethnic differences in physiological stress and relapse among treatment seeking tobacco smokers [Article]. Int. J. Environ. Res. Publ. Health 16 (17). https://doi.org/10.3390/ijerph16173090. Article 3090.

Huang, T.W., Cheung, D.S.T., Xu, X., Loh, E.W., Lai, J.H., Su, W.W., Wu, S.S., Lin, C.C., 2020. Relationship between diurnal cortisol profile and sleep quality in patients with hepatocellular carcinoma. Biol. Res. Nurs. 22 (1), 139–147. https://doi.org/10.1177/1099800419881195 [Article].

Huynh, V.W., Guan, S.S.A., Almeida, D.M., McCreath, H., Fuligni, A.J., 2016. Everyday discrimination and diurnal cortisol during adolescence. Horm. Behav. 80, 76–81. https://doi.org/10.1016/j.yhbeh.2016.01.009 [Article].

Jakuszkowiak-Wojten, K., Landowski, J., Wiglusz, M.S., Cubała, W.J., 2016. Cortisol awakening response in drug-naïve panic disorder. Neuropsychiatric Dis. Treat. 12, 1581–1585. https://doi.org/10.2147/NDT.S107547 [Article].

Johnson, J.A., Subnis, U., Carlson, L.E., Garland, S.N., Santos-Iglesias, P., Piedalue, K.-A. L., Deleemans, J.M., Campbell, T.S., 2020. Effects of a light therapy intervention on diurnal salivary cortisol in fatigued cancer survivors: a secondary analysis of a randomized controlled trial. J. Psychosom. Res. 139, 110266.

Keefe, J.R., Guo, W., Li, Q.S., Amsterdam, J.D., Mao, J.J., 2018. An exploratory study of salivary cortisol changes during chamomile extract therapy of moderate to severe generalized anxiety disorder. J. Psychiatr. Res. 96, 189–195. https://doi.org/10.1016/j.jpsychires.2017.10.011 [Article].

Kristiansen, E., Wanby, P., Åkesson, K., Blomstrand, P., Brudin, L., Thegerström, J., 2020. Assessing heart rate variability in type 1 diabetes mellitus-Psychosocial stress a possible confounder. Ann. Noninvasive Electrocardiol. 25 (5), 1–12. https://doi.org/10.1111/anec.12760.

Labad, J., Armario, A., Nadal, R., Solé, M., Gutiérrez-Zotes, A., Montalvo, I., Moreno-Samaniego, L., Martorell, L., Sánchez-Gistau, V., Vilella, E., 2018. Clinical correlates of hypothalamic-pituitary-adrenal axis measures in individuals at risk for psychosis and with first-episode psychosis [Article]. Psychiatr. Res. 265, 284–291. https://doi.org/10.1016/j.psychres.2018.05.018.

Landau, E., Raniti, M., Blake, M., Waloszek, J., Blake, L., Simmons, J., Schwartz, O., Murray, G., Trinder, J., Allen, N., 2021. The ratio of morning cortisol to CRP prospectively predicts first-onset depression in at-risk adolescents. Soc. Sci. Med. 114098.

Laures-Gore, J., Cahana-Amitay, D., Buchanan, T.W., 2019. Diurnal cortisol dynamics, perceived stress, and language production in aphasia. Journal of Speech, Language & Hearing Research 62 (5), 1416–1426. https://doi.org/10.1044/2018_JSLHR-L-18-0276.

Liu, Y., almeida, D.M., Rovine, M.J., Zarit, S.H., 2017. Care transitions and adult day services moderate the longitudinal links between stress biomarkers and family caregivers' functional health. Gerontology 63 (6), 538–549. https://doi.org/10.1159/000475557.

Mitchell, H.R., Kim, Y., Carver, C.S., Llabre, M.M., Ting, A., Mendez, A.J., 2020. Roles of age and sources of cancer caregiving stress in self-reported health and neuroendocrine biomarkers. Psychol. Health. https://doi.org/10.1080/08870446.2020.1800009 [Article].

Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L.A., 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. Syst. Rev. 4 (1), 1.

Morgan, E., Schumm, L.P., McClintock, M., Waite, L., Lauderdale, D.S., 2017. Sleep characteristics and daytime cortisol levels in older adults. Sleep 40 (5). https://doi.org/10.1093/sleep/zsx043.

O'Connor, D.B., Thayer, J.F., Vedhara, K., 2021. Stress and health: a review of psychobiological processes. Annu. Rev. Psychol. 72, 663–688.

Otto, L.R., Sin, N.L., Almeida, D.M., Sloan, R.P., 2018. Trait emotion regulation strategies and diurnal cortisol profiles in healthy adults. Health Psychol. 37 (3), 301–305. https://doi.org/10.1037/hea0000564 [Article].

Pace, T.W.W., Badger, T.A., Segrin, C., Sikorskii, A., Crane, T.E., 2021. The relationship between health-related quality of life and saliva C-reactive protein and diurnal cortisol rhythm in latina breast cancer survivors and their informal caregivers: a pilot study. J. Transcult. Nurs. 32 (4), 326–335. https://doi.org/10.1177/1043659620926537 [Article].

Padilla, G.A., Calvi, J.L., Taylor, M.K., Granger, D.A., 2020. Saliva collection, handling, transport, and storage: special considerations and best practices for interdisciplinary salivary bioscience research. Salivary Bioscience 21–47.

Ramos-Quiroga, J., Corominas-Roso, M., Palomar, G., Ferrer, R., Valero, S., Corrales, M., Richarte, V., Casas, M., 2016. Cortisol awakening response in adults with attention deficit hyperactivity disorder: subtype differences and association with the emotional lability. Eur. Neuropsychopharmacol 26 (7), 1140–1149.

Riis, J.L., Chen, F.R., Dent, A.L., Laurent, H.K., Bryce, C.I., 2020. Analytical strategies and tactics in salivary bioscience. Salivary bioscience: foundations of interdisciplinary saliva research and applications 49–86.

Rosnick, C.B., Wetherell, J.L., White, K.S., Andreescu, C., Dixon, D., Lenze, E.J., 2016. Cognitive-behavioral therapy augmentation of SSRI reduces cortisol levels in older adults with generalized anxiety disorder: a randomized clinical trial [Article]. J. Consult. Clin. Psychol. 84 (4), 345–352. https://doi.org/10.1037/a0040113.

Sampedro-Piquero, P., Vicario, S., Pérez-Rivas, A., Venero, C., Baliyan, S., Santín, L.J., 2020. Salivary cortisol levels are associated with craving and cognitive performance in cocaine-abstinent subjects: a pilot study [Article]. Brain Sci. 10 (10), 1–13. https://doi.org/10.3390/brainsci10100682. Article 682.

Schreier, H.M.C., Chen, E., 2017. Low-grade inflammation and ambulatory cortisol in adolescents: interaction between interviewer-rated versus self-rated acute stress and chronic stress. Psychosom. Med. 79 (2), 133–142. https://doi.org/10.1097/PSY.0000000000000377 [Article].

Schuler, K.L., Ruggero, C.J., Goldstein, B.L., Perlman, G., Klein, D.N., Kotov, R., 2017. Diurnal cortisol interacts with stressful events to prospectively predict depressive symptoms in adolescent girls. J. Adolesc. Health 61 (6), 767–772. https://doi.org/10.1016/j.jadohealth.2017.06.005.

Seidenfaden, D., Knorr, U., Soendergaard, M.G., Poulsen, H.E., Fink-Jensen, A., Jorgensen, M.B., Jorgensen, A., 2017. The relationship between self-reported childhood adversities, adulthood psychopathology and psychological stress markers in patients with schizophrenia. Compr. Psychiatr. 72, 48–55. https://doi.org/10.1016/j.comppsych.2016.09.009.

Shamseer, L., Moher, D., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., Stewart, L.A., 2015. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. Bmj 349.

Sin, N.L., Ong, A.D., Stawski, R.S., Almeida, D.M., 2017. Daily positive events and diurnal cortisol rhythms: examination of between-person differences and within-person variation. Psychoneuroendocrinology 83, 91–100. https://doi.org/10.1016/j.psyneuen.2017.06.001 [Article].

Stalder, T., Kirschbaum, C., Kudielka, B.M., Adam, E.K., Pruessner, J.C., Wüst, S., Dockray, S., Smyth, N., Evans, P., Hellhammer, D.H., Miller, R., Wetherell, M.A., Lupien, S.J., 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.

Starr, L.R., Dienes, K., Stroud, C.B., Shaw, Z.A., Li, Y.I., Mlawer, F., Huang, M., 2017. Childhood adversity moderates the influence of proximal episodic stress on the cortisol awakening response and depressive symptoms in adolescents [Article]. Dev. Psychopathol. 29 (5), 1877–1893. https://doi.org/10.1017/S0954579417001468.

Strahler, J., Nater, U.M., 2018. Differential effects of eating and drinking on wellbeing—an ecological ambulatory assessment study. Biol. Psychol. 131, 72–88.

Strahler, J., Skoluda, N., Kappert, M.B., Nater, U.M., 2017. Simultaneous measurement of salivary cortisol and alpha-amylase: application and recommendations. Neurosci. Biobehav. Rev. 83, 657–677.

Tada, A., 2018. Psychological effects of exercise on community-dwelling older adults. Clin. Interv. Aging 13, 271–276. https://doi.org/10.2147/CIA.S152939 [Article].

Urizar, G.G., Miller, K., Saldaña, K.S., Garovoy, N., Sweet, C.M.C., King, A.C., 2021. Effects of health behavior interventions on psychosocial outcomes and cortisol regulation among chronically stressed midlife and older adults. Int. J. Behav. Med. 1–14.

Walls, M., Dertinger, M., Unzen, M., Forsberg, A., Aronson, B., Wille, S., al'Absi, M., 2020. Assessment of feasibility and outcomes of a salivary cortisol collection protocol in five American Indian communities. Stress 23 (3), 265–274. https://doi.org/10.1080/10253890.2019.1675628 [Article].

Wong, J.D., Shobo, Y., 2017. The influences of daily stressors on morning cortisol levels in midlife and older retirees: the moderating roles of age and gender. J. Aging Health 29 (5), 858–879. https://doi.org/10.1177/0898264316645551.

Yaribeygi, H., Panahi, Y., Sahraei, H., Johnston, T.P., Sahebkar, A., 2017. The impact of stress on body function: a review. EXCLI Journal 16, 1057.

Yu, R., Branje, S., Meeus, W., Cowen, P., Fazel, S., 2019. Depression, violence and cortisol awakening response: a 3-year longitudinal study in adolescents. Psychol. Med. 49 (6), 997–1004. https://doi.org/10.1017/S0033291718001654 [Article].

Yu, R., Nieuwenhuis, J., Meeus, W., Hooimeijer, P., Koot, H.M., Branje, S., 2016. Biological sensitivity to context: cortisol awakening response moderates the effects of neighbourhood density on the development of adolescent externalizing problem behaviours. Biol. Psychol. 120, 96–107. https://doi.org/10.1016/j.biopsycho.2016.08.004.

References (Included studies references showed in Table 1):

Pruessner, J.C., Kirschbaum, C., Meinlschmid, G., Hellhammer, D.H., 2003. Two formulas for computation of the area under the curve represent measures of total hormone concentration versus time-dependent change. Psychoneuroendocrinology 28 (7), 916–931.

Fekedulegn, D.B., Andrew, M.E., Burchfiel, C.M., Violanti, J.M., Hartley, T.A., Charles, L. E., Miller, D.B., 2007. Area under the curve and other summary indicators of repeated waking cortisol measurements. Psychosom. Med. 69 (7), 651–659.