

Stress in Action Wearables Database: A database of non-invasive wearable monitors with systematic technical, [reliability](#), [validity](#) and [usability](#) information

Myrte Schoenmakers^{†1,2}, Melisa Saygin^{†1,2}, Magdalena Sikora^{†3}, Thomas Vaessen^{3,4},
Matthijs Noordzij³, Eco de Geus^{1,2}

[†]These authors contributed equally to this work and share first authorship. The author order was determined using a random list generator (<https://www.random.org/lists/>).

¹ Department of Biological Psychology, VU Amsterdam, Amsterdam, The Netherlands

² Amsterdam Public Health research institute, Amsterdam UMC, Amsterdam, The Netherlands

³ Department of Psychology, Health and Technology, University of Twente, Enschede, The Netherlands

⁴ Center for Contextual Psychiatry, Department of Neurosciences, KU Leuven, Leuven, Belgium

Corresponding author: Prof. Dr. Eco de Geus, Vrije Universiteit Amsterdam, Van der

Boechorststraat 7, 1081 BT Amsterdam, Netherlands, eco.de.geus@vu.nl, [+31 20 59 88813](tel:+31205988813)

Abstract

Ambulatory wearable monitoring of human physiology is increasingly utilized in the fields of psychology, movement sciences, and medicine. With the rapid growth of available consumer and research-oriented wearables, researchers are faced with a multitude of devices to choose from. It is timewise unfeasible for researchers to determine all relevant technical specifications, available signals, signal sampling details, (raw) data availability, *and* conduct a search of studies regarding the **reliability, validity and usability** of wearables. Thus, selection of wearables for a given study proves highly challenging and will often be unsystematic and uninformed. The ten-year research program Stress in Action initiated a publicly accessible database of wearable ambulatory monitoring devices. We outline the genesis and final structure of the first version of Stress in Action Wearables Database (SiA-WD), and a summary of the characteristics of the wearables it currently contains. Furthermore, one short-term (two days) and one long-term (three months) scenario from the field of stress research are provided with walkthroughs of how the SiA-WD can help select the optimal wearable for a specific research project. Insights gathered include the scarceness of studies testing wearable user-friendliness, inconsistencies in reported validity statistics, and imprecise manufacturer documentation on recorded physiological data such as sampling rate (or window) of signals and parameter extraction. The SiA-WD is the first open-access database to simultaneously include physiological sampling information and technical specifications along with a systematic **reliability, validity, usability** search. It will be iteratively expanded to facilitate informed and time-efficient wearable selection.

For access to the database: <https://osf.io/umgvp/>

Keywords Wearable · Database · Physiology · Ambulatory · Validation · Reliability · Usability

Introduction

Ambulatory wearable devices are being used in various fields, such as psychology, movement sciences and medicine for the continuous monitoring of a multitude of physiological signals in daily life (Pevnick et al., 2018). Ambulatory wearable devices, shortened to wearables hereafter, can be used by (clinical) researchers (Patel et al., 2012; Patel et al., 2021) as well as the public to track physical parameters like activity level, sleep duration, respiration rate, and heart rate variability. These wearables may come in forms such as smart garments with embedded sensors (e.g., Wearable Wellness System by Smartex, Hexoskin by Carré Technologies), smartwatches (e.g., Fitbit Sense 2), rings (e.g., Oura Ring by Oura Health, ResRing by BIOPAC), stick-on electrodes connected to a carry-on central processing unit (VU-AMS 5fs by Vrije Universiteit Amsterdam), stretchable belts (e.g., Equivital eq02+), or headbands (Muse 2 by Interaxon) (Majumder et al., 2017; Peake et al., 2018). As the already extensive availability of ambulatory wearables keeps increasing rapidly, it becomes challenging to find the optimal [wearable](#) for [a given research question](#) (Haddad et al., 2020). While validity and reliability are still the most prominent factors to consider in choosing a measurement device, other criteria have considerable pragmatic implications such as user-friendliness to reduce drop-out rates, data security to avoid data breaches (Areia et al., 2020), and low cost to allow larger-scaled research (Haddad et al., 2020).

In the field of stress research, the ability to monitor continuous physiological outcomes is becoming increasingly important to improve ecological and predictive validity (De Geus & Gevonden, 2024). However, making informed between-device comparisons and selecting the optimal wearable for physiological stress monitoring has proven challenging (Giurgiu et al.,

2022; Pevnick et al., 2018). This is both due to the vast number of devices and the time investment associated with examining all [reliability, validity and usability](#) papers per device. Furthermore, the list of devices that meet the required conditions for a study (e.g., specific set of [recorded](#) signals) [is](#) not always easy to identify through rudimentary web or literature searches. This may, for example, lead researchers to find out that a device does not allow for raw data extraction only after the purchase of the device or collection of data. Selection of devices is often based on recommendations of colleagues or by relying on the most often used wearables in the existing literature. This leads to the same devices being used time and again, which hampers the adoption of newer technology that may have better signal quality or offer new measures. In the current paper, we consider both physiological signals and parameters. Signals refer to continuous time-series data such as a photoplethysmography or an electrodermal activity sensor recording. Parameters are derived from these signals over particular windows of time, and examples include heart rate and skin conductance [level](#).

An overview of wearables in the form of a database systematically compiling the [reliability, validity and usability](#), amongst other technical information may assist researchers in choosing the optimal solution to perform continuous measurement of psychophysiological signals in daily life. To [date](#), such a systematic and regularly updated overview is currently lacking. While there have been attempts to provide an overview of available wearables [relevant](#) for a given research field, often in the form of systematic reviews (Iqbal et al., 2016; Lu et al., 2023; Vijayan et al., 2021), these often provided [no](#) systematic information on the [wearables' reliability, validity or usability](#). They also tend to have a narrow scope on a subset of parameters. For example, in a systematic review of ambulatory monitoring devices for measuring the cardiovascular activity in community-dwelling adults, a list of devices that measured one or more cardiovascular parameters [were provided](#) but important information on

devices' broader functionality and any other physiological parameters measured [were](#) left out (Lu et al., 2023). [Furthermore](#), existing databases are usually not updated iteratively. Just as in printed overviews in journal publications, [this static](#) information [can](#) become outdated quickly. This problem is particularly salient in view of the ongoing rapid expansion of new (versions of) wearables.

Recognizing these limitations, several attempts have been made to create online databases that would provide a more usable overview of wearable devices for a specific research field or for commercial purposes. In academic context, the CHIMERA database (Paredes et al., 2021) has been developed to facilitate access to and exchange of information on a wide range of concepts related to wearable technologies. Its main goal was to support a multidisciplinary discourse and collaboration between institutes and companies developing wearables, [and](#) to help researchers select the optimal device for their research. However, the database mostly served the development of wearables, not their final use, and is [now](#) no longer available online.

As a second example, [Henriksen and colleagues](#) compiled a [database](#) with 423 consumer-based fitness trackers and watches measuring physical activity. Such a database allows for easier comparison between the devices included because it is possible to filter the data using programming or directly in spreadsheet software. However, information that would be important for daily-life stress research is missing, including information on which stress-related physiological signals are measured, the battery life, availability of raw data, [and measurement](#) reliability and validity. [Therefore](#), direct comparison between devices on these aspects is hampered.

A third example of an [open-access database](#) is the Library of Digital Measurement Products, which has gathered validity and usability evidence [of the wearables included](#). This database provides a dynamic overview of both wearable and ambient (e.g., Wi-Fi sensing) technologies but the scope of information is still limited. The technical specifications are restricted to *Form Factor* and *Wear Location* with no further details [\(e.g., available signals\)](#) provided. This again makes comparison between devices difficult.

There are also overviews of wearables that are not geared to researchers but are entirely consumer oriented. These are usually broader in scope and rely predominantly on affiliate marketing models where revenue is generated from readers engaging with their reviews and subsequently making purchases through provided links. These outlets are better at regularly updating their resources and meeting the pace of new developments. Most commonly, such online resources (e.g. <https://www.wareable.com/> or <https://www.techradar.com/>) include reviews of consumer devices, provide an option to compare them using relevant filters, as well as signaling ongoing trends and new incoming wearables. An example of such an online comparison tool is the Vandrico [database](#) which provides information on over four hundred devices that can be used for industrial decision support and facilitate workplace productivity and automatization. However, due to the strong focus on the consumer markets, such websites are not the optimal resource for researchers, as research-oriented ambulatory monitoring devices, as well as more the detailed device information relevant to researchers are commonly not reported. Additionally, they usually concentrate on the technical specifications and the user-based reviews of perceived performance, [which](#) cannot be treated as scientific evidence of actual performance.

A gap can be seen between the listings of devices in scientific articles which easily get outdated, and the up-to-date, online listings which lack scientific focus. The recently started ten-year research program Stress in Action (stress-in-action.nl) therefore set as [one of](#) its [goals](#) to create a database of wearable ambulatory monitoring devices, including both consumer and research-oriented devices, with a comprehensive overview of aspects relevant for research. Online access to the Stress in Action wearables database (SiA-WD) will be made available at no costs, and it will facilitate the comparison between different devices through a set of criteria-based filters. The SiA-WD will be continuously updated [for the duration of the](#) Stress in Action [project](#).

The primary focus of the SiA-WD is stress research, and the inclusion of the devices is based on the signals related to the physiological stress response. [Our goal is achieving an overview of information comprehensive enough to optimally support stress researchers in selecting the device not only in regard to technical aspects, but that best facilitate answering their research question. The SiA-WD aims to support all researchers drawing inference based on the autonomic stress reactivity; from diagnostic efforts around biomarkers of psychopathology development \(e.g., as central to the work of Beauchaine and Gatzke-Kopp \(2012\) on cardiac control and impulsivity or emotion regulation\) to those focused on efficacy of interventions \(e.g., changes in HRV as reflective of autonomic regulation and cardiovascular risk in interventions like exercise therapy, in accordance with the neurovisceral integration theory \(see for example de Oliveira Matos et al. \(2020\)\). Considering the heterogenous nature of stress research, studies can encompass a wide range of methodological prerequisites and come with different requirements that a device needs to satisfy, including different parameters of interest, which in turn require different measurement techniques. Therefore, the SiA-WD covers a wide range of devices recording physiology at different levels of scientific](#)

detail supporting the diversity inherent to stress research. To illustrate the use of the database, we end the paper with two research scenarios (focused on threat-challenge and cross-stressor adaptation hypotheses) showcasing how research questions guide device selection. Moreover, SiA-WD aims to facilitate the selection of the most suitable devices for cohort studies measuring stress in daily life as well as to provide best candidate devices for subsequent in-depth validation studies. However, the database can also facilitate researchers to select wearables for a range of other research topics such as sleep, physical activity, and cardiovascular health. In short, this database will be a scientific resource on available wearable devices for ambulatory assessment of physiology, and will be open-access, applicable to many different research fields, and periodically updated every six months for the coming ten years.

In this paper, we begin by outlining the methods used to structure and populate the database. Then, the database is presented together with its relevant components and we provide two research scenarios demonstrating how researchers can use the database to optimally choose wearables given their research project and resources. We end by discussing the insights gathered while creating the first version of the SiA-WD, as well as outlining the future maintenance of the database.

Methods

Currently, the SiA wearables database is implemented as a Microsoft Excel file that is publicly available at <https://osf.io/umgvp/> [the database will be added to this public OSF page upon acceptance, for now please see Supplementary 1]. The excel format ensures easy transfer to other formats. The maximal amount of data, even with foreseen future growth, is

considered manageable in this format. The information in the cells of this excel database was structured in a way that would enable easy automated searching, sorting, and analyzing the database with programs such as Python or R. Most of the time, numerical fields or text field from dropdown menus were used to standardize the input. Rarely, open text fields were used to allow additional clarification. For columns containing multiple components, a standardized format was created with semicolons used to separate the different components. For example, a device with an accelerometer with three axes sampling at 1000 Hz and positioned on the hip is displayed as “1; 3; 1000; hip”; whereas a cell of a device without an accelerometer will display as “0”.

Below, we first give a description of our approach to select wearables to be included into the database (each *row* of the database contains the information of one device), followed by the type of information on these devices to include in the database (structured into the *columns* of the database), and finally the methods to populate the database with the device-specific information (filling in the *cells*). Since the [reliability, validity and usability](#) of a wearable are essential aspects for researchers, separate subsections describe our strategies for finding, summarizing, and adding this information on the wearables to the database.

[Wearables included in the SiA wearables database](#)

The starting point to select wearables to be included into the database was a list of physiological signals and the parameters derived from these signals that have been widely used to study the physiological human stress response (Geus & Gevonden, 2024; Geus et al., 1995; Eckberg, 2003; El-Hamad et al., 2023; Grossman & Svebak, 1987; Henley et al., 2018; Kim et al., 2018; Klimek et al., 2023; Malm et al., 2004; Mukkamala et al., 2015; Neumann & Blanton, 1970; Osei et al., 2024; Rahma et al., 2022; Steptoe et al., 2000; Treadwell et al.,

2010; Ward et al., 2012; Wilhelm et al., 2003). Based on the vast body of physiological stress research, we included electrocardiography (ECG), impedance cardiography (ICG), respiration, photoplethysmography (PPG), electrodermal activity (EDA), and blood pressure as our main signals; for more details see Table 1.

Table 1

Basic explanation of the signals commonly used in ambulatory stress research

Signal	Details
Electrocardiography (ECG)	ECG records the electrical activity of the heart; i.e., the depolarization and repolarization of the heart muscle. Commonly, a 12-lead ECG is recorded in clinical settings (McStay, 2019), but for ambulatory recordings a 2-lead solution is often used (Krittanawong et al., 2021)
Impedance cardiography (ICG)	ICG records the changes in impedance in the thorax, e.g. caused by respiration and fluctuations in blood volume and flow during contractions of the heart. It can provide information regarding the cardiac and respiratory system (and fluid states) (Parry & McFetridge-Durdle, 2006).
Photoplethysmography (PPG)	PPG signals are based on absorption and/or reflection of light (green, red or infrared), commonly measured at the wrist (e.g., smartwatches) or finger (e.g., pulse oximeter). The signal can be used to derive cardiac measures such as the heart rate, and – when using multiple wavelengths – oxygen saturation (Alian & Shelley, 2014).
Electrodermal activity (EDA)	EDA measure the conductance of the skin using two electrodes. EDA captures the activity of the eccrine sweat glands and is considered to be a pure measure of the sympathetic nervous system (Boucsein et al., 2012).

From these signals, sometimes by combining them, physiological parameters can be extracted and reported by various wearables. Several parameters typically used in physiological stress research are heart rate (HR) (Steptoe et al., 2000) and other parameters measuring cardiac performance like the Pre-Ejection Period, Stroke Volume, and Cardiac Output from the combined ECG and ICG (El-Hamad et al., 2023; Henley et al., 2018; Malm et al., 2004; Treadwell et al., 2010), respiration rate (RR) and tidal volume (Vt) from respiratory stretch

belts, thoracic impedance, or inductance plethysmography (de Geus et al., 1995; Pattyn et al., 2010; Wilhelm et al., 2003), skin conductance level (SCL) and the frequency of non-specific skin conductance responses (ns.SCR) from the EDA (Klimek et al., 2023; Neumann & Blanton, 1970; Rahma et al., 2022) and a number of heart rate variability measures like the standard deviation of heart period intervals (SDNN), the root mean square of successive differences (RMSSD), or high frequency spectral power (HF-HRV) from the ECG or PPG signals (Kim et al., 2018; Osei et al., 2024) sometimes derived in combination with respiration signals to obtain peak-valley respiratory sinus arrhythmia (RSA) (Eckberg, 2003; Grossman & Svebak, 1987). Lastly, measures like systolic (SBP), mean arterial (MAP) and diastolic (DBP) blood pressure are obtained through oscillometer cuff-based methods or by using estimation through Pulse Transit Time assessment (Mukkamala et al., 2015; Vrijkotte et al., 2000; Ward et al., 2012).

Apart from these physiological measures it has been recommended to additionally measure posture, physical activity and ambient noise level and temperature, as they can confound the ambulatory assessment of many of the above physiological signals and parameters (E. De Geus & Gevonden, 2024) but can also by themselves reveal effects of stress on behavioral activation (Giakoumis et al., 2012; O'Brien et al., 2017; Sano & Picard, 2013). Therefore, when a device co-records the accelerometer and/or gyroscope signals and skin or ambient temperature this information is also included in the SiA-WD. Wearable solutions also exist for a number of other signals and parameters [such as](#) sleep duration and quality (Irwin, 2023; Romeijn et al., 2012; Sadeghi et al., 2019), ambient light (Akinwande & Kireev, 2019), electrooculography (Moon et al., 2023), electromyography (Ngo et al., 2022), or core body temperature sensing (Dolson et al., 2022). To keep the SiA-WD concise we did not include

wearables that *only* measured these other signals but included wearables that co-record them *in addition* to the physiological parameter(s) of interest in stress research.

Currently, there are hundreds of wearable devices that measure the physiological signals listed above. A pragmatic approach was used to select a subset of these devices for the first iteration of the SiA-WD database. Based on domain-knowledge of the authors we selected a set of ten well-known wearables already used in research: the Empatica E4, Empatica EmbracePlus, VU-AMS 5fs, Oura Ring, Hexoskin Proshirt, Mindware Mobile, Novacor Diasys 3 plus, Apple Watch Series 6, Garmin vivosmart 5, and Fitbit charge 5. Adding these first ten devices to the database bootstrapped the iterative process of selecting the relevant device information to be included in the database, described in more detail below.

Subsequently, we used a systematic approach to select *both often-used and newer* devices through keyword-based searches by:

(1) Conducting database searches at PubMed, IEEE Xplore, Web of Science, Scopus and APA PsycInfo (through Ebscohost) and ACM Digital Library, with the sorting set to ‘Most recent,’ while using a search string containing keywords on wearable monitors, stress, and physiological signals. The physiological keywords included both those on signal types and parameters that can be derived using these signals. The search string was as follows: (photoplethysmography OR electrocardiography OR “heart rate” OR “electrodermal activity” OR “skin conductance” OR ns.SCR OR “blood pressure” OR electromyography OR “skin temperature” OR “stress level” OR “stress detection”) AND (wearable).

(2) Examining website articles on newly released wearables via *dcrainmaker.com*, *wearable.com*, *techradar.com*, *vandrico.com*, and *wired.com/tag/wearables*

(3) Consulting colleagues and other researchers who work in the field and asking them for nominating wearables, *especially newer ones*.

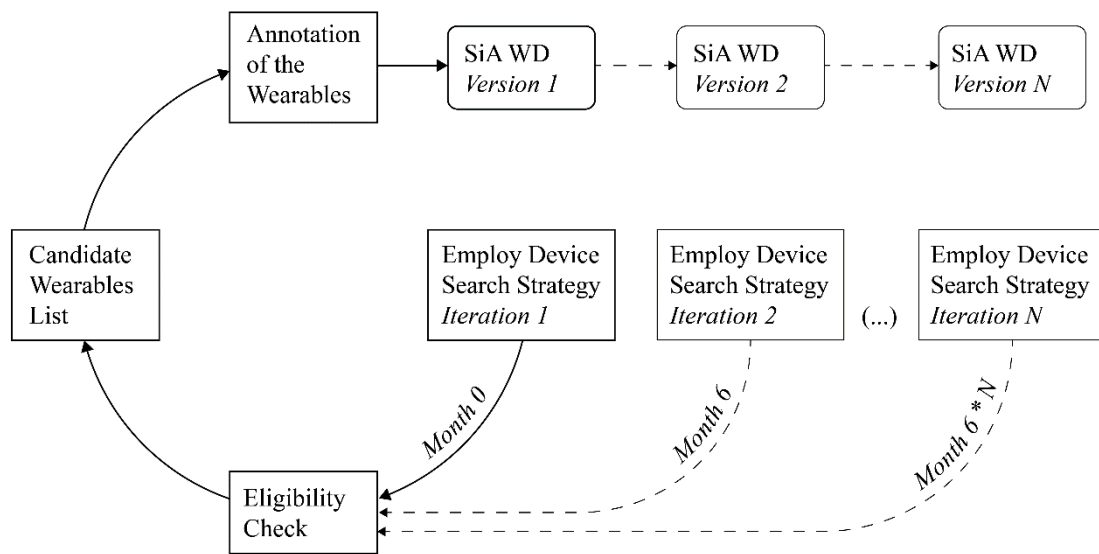
The list of wearables obtained in this way was filtered by the inclusion and exclusion criteria listed in Table 2.

Table 2

Eligibility criteria for wearable device inclusion in SiA-WD

Inclusion criteria
<ul style="list-style-type: none"> • The device is available on the market or was recently discontinued but is still widely used in research • The device measures physiological parameters relevant to stress research • The device is wearable, mobile, and can be used in daily life • Specified information on the device and its components is available in English, either on the manufacturer's website or in a form of a digital manual
Exclusion criteria
<ul style="list-style-type: none"> • The device is in development and prototype phase and not readily available • The device has a singular focus on measuring gross body movement • The device is predominantly intended for medical use

This process led to the first 172 candidate wearables listed in *Supplementary File 1*. As illustrated in Figure 1, the above procedure will be replicated every six months for at least the duration of the Stress in Action project, resulting in updated versions of candidates for the SiA-WD. Using a majority voting procedure the authors selected a subset of 54 wearables for the purpose of this paper, yielding version 1 of the SiA-WD. In selecting this subset we considered a wide variety of devices to obtain a database structure covering all aspects required for adding various device types (e.g., including various device types – e.g., watch, ring and patch – and cover the signals included in the database). We then focused on devices that combine various signals that have been repeatedly used in stress research, , and had relative high usability or relatively high levels of established reliability/validity.

Figure 1*Iterative process of populating the SiA Wearables Database*

Device information included in the SiA wearables database

For all devices, the database columns present relevant information on aspects of the device, its measurement capabilities and performance. A critical piece of information is the signals recorded by a device and the calculation of certain parameters. Accordingly, we report on both the actual signals (e.g., the columns PPG, ECG, BP, Respiration) that are recorded and on parameters that are provided using these signals. For the respiration column, it is important to note that a device is considered as measuring respiration signal (i.e., noted as 1 in the database) if it captures both the rate and the depth of breathing continuously. Example techniques would include thoracic and/or abdominal piezoresistive belts, strain gauge belts, respiratory inductance plethysmography, and impedance pneumography. Thus, a wearable only extracting respiration rate from the PPG signal would have a 0 in its respiration signal column, but respiration rate would be added as a parameter in the provided parameters column. Regarding the BP column, all wearables that produce blood pressure (BP) values are considered to measure BP and annotated accordingly in the database. The employed method

is specified; e.g., cuff-based measures such as auscultatory and oscillometry are specified as well as cuffless methods based on (machine learning) algorithms.

To decide which other columns to add to the SiA-WD, we reviewed the extant literature on relevant criteria for the selection of a wearable. Existing recommendations (Byrom et al., 2018; Kleckner et al., 2021; Pantelopoulos & Bourbakis, 2008, 2010; Polhemus et al., 2020) demonstrate that a range of aspects should be considered before selecting a wearable. For example, the intended use of a wearable, be it clinical, research, or consumer, could be a selection criterion when assessing a patient population with a clinical grade device being required. Other general information, such as the price, is also essential, including one-time-purchase costs (e.g., the device and software) as well as additional monthly subscriptions (e.g., for application or data server use). Such aspects were grouped in the category of *general device information*. All technical information about the device was combined in the *technical specifications* category, including battery life or charging method and duration, which determines whether a participant can measure constantly during the study period or if they must charge or swap the device (Boateng et al., 2019; Liu & Han, 2022). Additionally, it is important to consider the availability of (raw) data and the form in which it is available, e.g., at the signal level or merely as time series of provided parameters. It is also useful to know where the data is stored, if it is stored securely and what software is required for data processing. These, among other data related aspects were grouped in the *data access* category (Kleckner et al., 2021; Siboni et al., 2016). The last set of columns contains information on *reliability, validity and usability*, which are important criteria for selecting a wearable from a researcher perspective.

After having decided on these categories and their aspects, we fine-tuned the structure by actively seeking out the information for the first ten devices in the SiA-WD. Based on the issues encountered we added a number of aspects or changed how we reported on them. We extended this fine-tuning after consultation with experts on wearables research within the Stress in Action consortium. This led to the final column structure shown in Table 3.

Table 3

Device aspects included in the SiA-WD, grouped per category

Category	Column	Details
General	Device	Name of the device
Device	Manufacturer	Name of the manufacturer
Information	Website	Link to the webpage of the device
	Release Date	Official release date of the device on the market
	Market Status	Upcoming/Current/Discontinued
	Main use	Research/Consumer/Clinical or their combinations
	Device Costs	One-time purchase price (EUR) and additional costs
	Wearable type	Type of the device (e.g., watch, CPU + electrodes)
	Location	Location at which the device is worn (e.g. wrist, chest)
	Size	Device dimensions in mm
	Weight	Device weight in gram
Signals	PPG	Photoplethysmography
	ECG	Electrocardiogram
	ICG	Impedance Cardiography
	EMG	Electromyography
	Respiration	Respiration (and how it is derived – e.g., through ICG)
	EDA	Electrodermal Activity (Galvanic Skin Response)
	BP	Blood Pressure
	Accelerometer	Accelerometer
	Gyroscope	Gyroscope
	GPS	Global Positioning System
	Skin Temp	Skin Temperature
	Other signals	All other signals the device can record
Technical specifications	Water resistance	In terms of depth in meters and time in minutes
	Battery life	Maximum battery life as specified by the manufacturer
	Charging method	Device charger or disposable/rechargeable batteries
	Charging duration	Time needed in minutes to fully recharge the battery
	Bio-cueing	Options to cue (e.g., vibration) users based on their physiology
	Bio-feedback	Access users get into their physiology (e.g., via device display)







Data access	Raw data available	If signal-level data can be exported for analysis
	Provided parameters	The parameters automatically generated by device
	Sampling window of parameters	Approximate time window around which each parameter is calculated, listed for known parameters
	Data transfer method	All methods of transfer (e.g., Bluetooth, SD-card)
	Compatibility	System compatibility of the device (mobile and PC)
	Required software	Software required to record and/or extract the data
	Additional software	Additional (e.g., analytical) software available
	Internal storage method	Internal storage availability (e.g., SD-card)
	Device storage capacity	Hours of data that can be recorded and stored internally
	Server data storage	Data stored on external servers (including their location)
	GDPR compliance	Compliance with the General Data Protection Regulation Act
	FDA approval/clearance	FDA approval or clearance for the device or its components
	CE approval/label	Device has been assessed to meet the EU market regulations
Reliability, validity, usability	Highest level of validation evidence	Highest level of validation available (external/internal/no validation)
	Number of validity and reliability studies reviewed	Number of the relevant validity and reliability studies found in the search for a given device
	Studied parameters	List of the parameters included in the reviewed VRU studies
	General validity and reliability synthesis	Short synthesis statement of the validity and reliability results of the reviewed studies
	Number of usability studies reviewed	Number of the relevant usability studies found in the search for a given device
	General usability synthesis	Short synthesis statement of the usability results of the reviewed studies
	Hyperlink to the device VRU page	Hyperlink to the separate VRU page with detailed reliability, validity, usability information for the device
Curators' expert scores	SiA short-term usefulness score	Average score based on the perceived usefulness of a device for the future short-term SiA studies as assessed by three raters
	SiA long-term usefulness score	Average score based on the perceived usefulness of a device for the future short-term SiA studies as assessed by three raters

The database contains a total of 53 columns. The columns represent the aspects relating to the five categories (General Device Information, Signals, Technical specifications, Data access, [Reliability, validity and usability](#)). The last two columns include the SiA expert scores for

short- and long-term assessment. As a brief illustration of the database, Figure 2 shows eleven example columns for six different devices selected to represent the different categories of wearables (e.g. more research or consumer oriented).

Figure 2

A section of the SiA-WD showing 10 of the 53 columns for six devices

DEVICE	DEVICE COSTS (€)	WEARABLE TYPE	PPG	ECG	RAW DATA AVAILABLE	DEVICE STORAGE CAPACITY (HOURS;MB)	BATTERY LIFE (HOURS)	HIGHEST LEVEL OF VALIDATION EVIDENCE	NUMBER OF VALIDITY AND RELIABILITY STUDIES	GENERAL VALIDITY AND RELIABILITY SYNTHESIS
 EmbracePlus	1.984; three year bundle per a device + software (25% academic discount available)	Watch	1; reflection; 26-208; wrist; green, red and infrared	0	1	24; 128	48	Internal	1	Reliable and valid for use in clinical SpO2 measurement under static conditions, for both male and female subjects as well as for different skin tones.
 VU-AMS Core	5995; one-time purchase price including a device + software	CPU + electrodes	0	1; 3; 1000; electrode; lead I; lead II; aVF	1	NP; 32000	28	No validation	0	No validation studies yet. This device is the successor of the VU-AMS 5fs device.
 Fitbit Charge 5	150	Watch	1; reflection; NP; wrist; red and infrared	0	0	NP; NP	168	External	1	Good HR detection for sinus rhythm, poor for atrium fibrillation. Irregular rhythm notification acceptable for guiding decision regarding atrial fibrillation patients.
 Oura Ring Gen3	314; one-time purchase price including a device but not the required monthly app subscription (€6)	Ring	1; reflection, transmission; 250; finger; green, red and infrared	0	0	168	168	External	4	High validity for HR and RMSSD and acceptable accuracy for nocturnal AVNN, pNNS0, HF and SDNN in average-per-night tests with high error rates for LF and LF:HF ratio. Participant compliance was high over several weeks.
 Hexoskin Proshirt	600	CPU + vest	0	1; 1; 256; touch; lead I	1	2400	36	External	12	Strong evidence is present that respiratory rate (RR) and heart rate (HR) can both be validly estimated with Hexoskin Proshirt during a variety of light, moderate, and intense activities, whether controlled or naturalistic. Only two studies one of which were naturalistic showed HR to not correspond with reference device values [...]
 Research Ring	2020; one-time purchase price of a device, option to interface up to 9 rings with AcqKnowledge Ring software (€5284)	Ring	1; reflection; 400; finger; green	1; 1; 800; MIR; lead I	1	Bluetooth	10	No validation	0	NA

Note. In these columns, 0 stands for Not Available, 1 for Available, and NP for Not Provided. The SiA-WD itself does not contain the images of the devices, but are entered in this figure for illustrative purposes.

1 Retrieving the relevant information for a wearable

2 To fill the cells of the database (i.e. retrieving the relevant information in each column for
3 each of the selected wearables) [for the general and technical device specifications as well as](#)
4 [the information regarding the physiological measurement](#), we started by using the
5 manufacturer's website and device manuals in English. This meant searching the web using
6 the device name – and if known already, the manufacturer name – to find the manufacturer's
7 website and looking for the product specifics' page and the device manual. In addition, we
8 collected data from publications on studies that utilized the wearable. When any information
9 on the relevant aspects was missing, a standardized email inquiring about the missing
10 information was sent to the email address specified on the [manufacturer's](#) website.

12 Establishing the validity and reliability of devices in the SiA wearables database

13 Besides completing the *general device information, signal, technical specifications, and data*
14 *access* columns, it is important to provide researchers with an understanding of the device's
15 measurement accuracy and consistency, that is of whether a device measures what one
16 expects it to measure and if it maintains doing so over extended periods of time. This can be
17 [sufficiently](#) achieved by focusing on convergent validity and test-retest reliability (Hopkins,
18 2000). Convergent validity refers to how well a device's measurements agree with that of a
19 gold standard or other reference device (RD), under the same conditions while recording
20 concurrently. Test-retest reliability assesses the extent to which the device, under similar
21 conditions, produces the same results at different points in time (Kottner et al., 2011).

23 Assessment of convergent validity can be done at the signal, parameter, and event levels (van
24 Lier et al., 2020). While we agree with van Lier and colleagues that wearables should ideally
25 be validated at *all* these three levels, most validation papers to date do not report this

extensive validation information. Furthermore, in non-diagnostic ambulatory monitoring, the raw time-series signal itself in general is not the focus of interpretation, but the parameters (e.g., HR, SDNN, RMSSD) derived from the signals frequently are (van Lier et al., 2020). Thus, when presenting validity information, our focus will be on convergent validity at the parameter level (e.g., to what extent RMSSD of a new wearable comply with that from a gold standard) and not, for example, on the signal-to-signal cross-correlations. For the current version of the database, we did not include studies that investigate a wearable's construct validity (e.g., if parameters such as heart rate indeed differ between baseline and intense exercise) or ability to make predictions using machine learning models (e.g., predictive validity of the PPG signals for subjective stress detection).

Establishing usability of devices in the SiA wearables database

Apart from validity and reliability, the research value of a wearable also depends on its usability, as this is a main contributor to participant compliance in ambulatory studies. High usability can also enhance optimal use by participants and thereby increase signal quality and reduce data loss. Among participants who initially agree to put on wearable monitors for a longitudinal study, the majority may remove it before study's termination due to getting irritated, uncomfortable, overwhelmed, or unwell (Jeffs et al., 2016), and may perceive the monitor as impeding their usual activity (Areia et al., 2020; Ehmen et al., 2012). Dias and colleagues (2009) furthermore asserted that skin reactions may occur with certain wearables clenched tightly onto the skin. These can result not only in a reduction of quality data but also in ethical and medical concerns. Thus, to get an understanding of usability (including user-friendliness and user-acceptance), we identified papers that conducted quantitative or qualitative (e.g., posing systematic questionnaires or open interviewing) research on these aspects.

Compiling the reliability, validity and usability information of devices

The **reliability, validity and usability** of the wearables were established by a literature search of papers on *original* experimental studies using the wearable (i.e., not meta-analyses or reviews) that were published in English in peer-reviewed journals or conference proceedings. No filtering was performed based on year of publication. As the search strategy was to include all papers that investigated a given device's **reliability, validity or usability**, the search string included terms relevant to all three aspects along with the device name :

“((Device Name) AND (valid* OR reliab* OR compar* OR accur* OR verif* OR usab* OR "user experience" OR "user friend*" OR user-friend*))”. If a device had multiple names for the given version, they were all added as a device name using the OR operator.

Papers were only considered to supply the necessary validity information when they mentioned the use of a reference device (RD) in their abstract, such as the following:

“Measurements were recorded simultaneously using the Hexoskin and Polar Team Pro” (Haddad et al., 2020). Furthermore, the RD had to be a proven gold standard (i.e., showing greater established precision than the wearable being validated), such as the COSMED K5 for respiratory volumes or an ambulatory device consistently known to be accurate for heart rate, such as the Polar H10 Band. If two wearables were merely being compared to one another without a clear rationale for one of them to be considered as the golden standard/ground truth, the paper was excluded. See Table 4 for a list of eligibility criteria for papers to be accepted as providing device-specific information.

Table 4*Eligibility criteria for device-specific studies on validity and reliability*

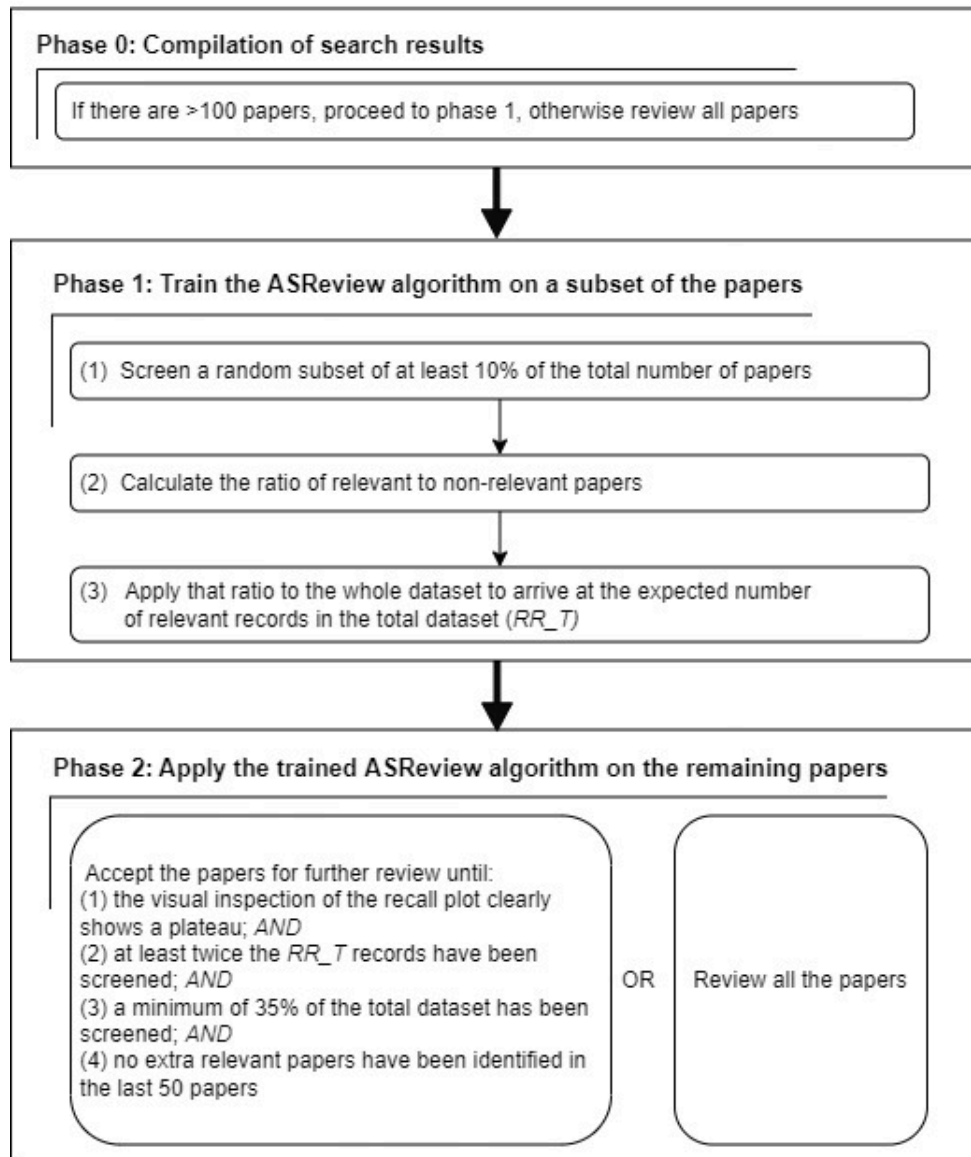
Inclusion Criteria	Exclusion Criteria
<ul style="list-style-type: none"> Assessment of parameter level convergent validity, test-retest reliability, and/or usability. Assessment of convergent validity Peer-reviewed articles and conference proceedings published in English 	<ul style="list-style-type: none"> Studies on construct validity only Studies on machine learning based detection of secondary outcomes (e.g., perceived stress) Meta-analyses and reviews Theses, grey literature, other text that was not peer reviewed

Increasing the efficiency of the literature search

For several wearables, the literature review produced a large amount of hits (e.g., 530 papers for Empatica E4 device). Given the rapid expansion of this literature, this challenge protracts into the future maintenance of the database. We therefore developed a procedure to efficiently extract relevant papers, using the ASReview version 1.6.2 (<https://asreview.nl/>) screening tool, used when there were more than 100 papers found for the search for a device. ASReview can greatly reduce the time needed to select the relevant records through a so-called active learning method, in which a machine learning model continuously rearranges the items based on the decisions made by a researcher regarding their relevance (Van De Schoot et al., 2021). The two phases of the literature review were based on the SAFE procedure (Boetje & Van De Schoot, 2024) and are shown in Figure 3. Further details on the approach can be found in *Supplementary 2*.

1 **Figure 3**

2 *Using ASReview to select papers for a detailed review of **reliability, validity and usability** of*
 3 *a wearable*



4

5

6 Data extraction of the RVU (reliability, validity, usability) information

7 For each wearable in the database, detailed information on **reliability, validity and usability**

8 were extracted manually from the papers identified. This information was collated in device

9 specific worksheets referred to as the **RVU** worksheet and were maintained separately from

10 the database. **RVU** represents the **reliability, validity and usability** information of a device, as

extracted from the papers identified in the search. A direct hyperlink to these device-specific RVU worksheets is provided in the database. Validity and reliability data were extracted at the level of the provided parameters (e.g., heart rate), as papers can focus on different parameters from the same device and can report differential validity and reliability performance for these parameters. The information extracted for the RVU per device, per parameter is listed in Table 5. Below, we first explain which data is extracted per paper, and then clarify how the information of all papers for a device (as found in the RVU worksheet) is used by the SiA-WD curators to create the synthesis statements in the database.

From each paper of a device, first, the overall conclusion of the authors who evaluated the device reliability, validity and/or usability were extracted as direct quotations (e.g., “Hexoskin was able to correctly measure tidal volume in healthy subjects during various tasks [...] by Mannée and colleagues (2021)”), typically taken from the abstract. This column in the RVU worksheet was called the “Authors’ conclusion,” referring to the authors who empirically tested the device, and not to the curators (the three first co-authors of the current paper). If the test-retest reliability was also examined in a study, the authors’ conclusion also included the reliability statement per parameter tested. If specified, the time interval between two testing points was entered into the RVU worksheet as a separate column.

Table 5

Data extraction form for reliability, validity and usability of each device, gathered in the device specific information page

Data category	Column name	Description
Metadata	Study citation	In APA 7 th edition format
	Year of publication	YYYY
	Publication journal	Journal name

Study characteristics	External/internal	Was the study performed internally (i.e., by device manufacturers) or externally
	Population	Population type (e.g. elderly)
	Sample size	Initial and after exclusion
	Age	Mean and SD
Methods (reliability , validity and usability of the device)	Reference device	The gold-standard reference device used for comparison, e.g. standard clinical ECG
	Number of events	Number of tasks included
	Included events	Types of tasks included
	Studied parameters	All the parameters studies (e.g., HRV and SCL)
	Time between test-retest	Time after which authors made a second measurement with the device under same circumstances to calculate internal consistency
Findings and conclusions	Authors' conclusion	Excerpt of authors' (who empirically assessed the device) conclusion
	Validity and reliability summary	Summary of validity/reliability findings written by the curators
	Usability summary	Summary of usability findings written by the curators
	Final verdict (Negative/Neutral/Positive)	Curator's final verdict per physiological parameter based on all available RVU information

1

2

3

4

5

6

7

8

Apart from the authors' conclusion on convergent validity and test-retest reliability, information on the independence of the validation work, the reference device used, and the scope of conditions and populations considered in the validation studies were entered (see Table 5 *Metadata, Study characteristics and methods*). We determined the nature of a study, external or internal, by reviewing conflict-of-interest disclosures. External validations are those conducted independently from the device manufacturers, and typically show less bias than internal studies. In line with the INTERLIVE consortium (Johnston et al., 2021; Molina-

Garcia et al., 2022; Mühlen et al., 2021) that proposed recommendations for systematic validation of commercial devices, we argue that variables including the sample size, type of population, duration of data collection, and the testing conditions (i.e., tasks and the presence of daily life components) matter in determining the trustworthiness of a validity study. The relevant information was located by scanning the full text of papers and entered in the [RVU](#) worksheets to help future researchers using the database assess if a device has been sufficiently validated for the tasks, populations, and measurement durations of their interest.

The authors' conclusion may sometimes lack important methodological information on the study. Thus, after reviewing a given paper, curators wrote a 'Validity and reliability summary' and a 'Usability summary' whenever applicable. The summaries aim to succinctly capture the most important methods, findings, and unique sample characteristics in a paper. An example 'Validity and reliability summary' is "[...] The 95% Limits of Agreement were -3.89 to 3.77 (mean bias 0.06) beats per minute for HR and -173 to 171 (mean bias -1) for IBIs. Results were comparable across all subgroups (i.e., different skin type, hair density, age, BMI and gender)". An example usability summary is "87% of participants successfully and consistently wore the ring. Gender, age and weight did not influence the adherence. 56% were willing to continue wearing the ring after the study [...]." After considering all information extracted from a paper, curators made a final verdict for each studied parameter: positive, neutral or negative (e.g., HR: positive; RMSSD: positive; HF-HRV: negative).

Synthesis of [reliability](#), [validity](#) and [usability](#) studies for SiA Wearables Database

After completing the review of the [RVU](#) studies for a device, a synthesis statement for validity and reliability were written in the SiA wearable database by the curators. This synthesis combines all the available validity and reliability studies in the device-specific

RVU page – and another for usability – combining all the usability studies in the device-specific RVU page. An example excerpt of a general validity-reliability synthesis is as follows: “Most studies comparing E4 to a reference device (RD) suggest high validity of HR under static and dynamic conditions (one study found opposite results) and often report HRV parameters as valid but only in static conditions, although results seem conclusive only for RMSSD and SDNN [...]”. An example usability synthesis is “Hexoskin is perceived by participants to be highly comfortable when used in a laboratory procedure. It can be used in those with chronic obstructive pulmonary disease [...] Use in an ambulatory study of a week, however, is problematic [...]”. Such synthesis statements allow researchers to review all validation or usability conducted for a device. Thus, they can make expedited and systematic between-device comparisons in *reliability, validity and usability*. In addition to the validity-reliability and usability syntheses, key information from the RVU worksheet of a device were also entered into the SiA-WD. The full account of these columns are listed in Table 6.

Table 6

Reliability, validity and usability information on the wearable in SiA-WD

Column name	Entry type	Explanation
Highest level of Validation	No validation/Internal/External	External validation ranks higher than an internal one.
Number of validity-reliability studies reviewed	Numeric	Number of validity and/or reliability studies resulting from the search
Studied parameters	List	All the parameters present in all of the studies
General validity and reliability synthesis	Text	A comprehensive but short written overview of all the different study results. This includes both validity and reliability information.

Number of usability studies	Numeric	Number of usability studies resulting from the search
General usability synthesis	Text	A condensed overview of the device usability based on the reviewed studies. This includes not only user experience but also aspects such as adherence.

1

2 [Wearable selection for applications in stress research](#)

3 After populating the SiA-WD, the curators created a scoring system to rank order the

4 wearables based on their usefulness for stress research. In this scoring, the curators [were](#)

5 strongly guided by the perceived usefulness of the wearables for the future large-scale cohort

6 studies within the [Stress in Action](#) project. These cohort studies seek to answer questions on

7 both short-term and long-term relationships between physiological, emotional, cognitive, and

8 behavioral responses to stress in within-person studies and the predictive value of these

9 responses for disease outcomes in between-person studies. Within-person studies require

10 long-term assessment across weeks or months of at least one physiological indicator of stress

11 in the same participant, whereas between-person studies to predict disease require more

12 intensive measurements of physiological stress systems – i.e., measuring various signals and

13 output parameters – for one to three full 24-h recordings, possibly on selected days of the

14 week. While a high level of validity and reliability is always essential independent of wear

15 time, high levels of usability become increasingly more important for longer wear times.

16 Accordingly, two separate scores were given by the three curators to each device reflecting

17 its perceived research usefulness for both short (i.e., approximately two days) and long (i.e.,

18 at least two weeks) measurement durations.

19

Since scoring the devices is an inherently subjective process, a set of criteria had to be outlined to achieve standardization of the rating by the [three](#) curators. The criteria list came from the columns in the database (e.g., raw data availability, validity, usability). Separate scoring criteria were used for a typical short-term (approximately 48 hours) and long-term (2+ weeks) ambulatory study. The finalized scoring criteria per term of use is listed in Table 7. Blinded to each other's scores, each curator scored each device on a scale of 1 (least optimal) to 10 (most optimal) for both short-term and long-term use. The scores of the three curators were averaged into the final expert scores for short- and long-term studies. These were included in the database. No score was entered for devices that had not been on the market for [less than](#) 8 months. This avoids unfair scoring of devices which were just released on the market, for which little to no external validation research will be available.

Table 7

List of scoring criteria ordered by their importance for short-term and long-term studies

Short-term importance of criteria		Long-term importance of criteria	
• GDPR approval	High	• GDPR approval	High
• CE approval	High	• CE approval	High
• All Reliability-Validity Criteria	High	• Price	High
• Number and type of different signals that can be measured by the device (e.g., PPG, ECG, accelerometer)	High	• Number of different physiological parameters that can be measured by the device (e.g., HR, Skin temperature)	High
• Raw data availability	High	• Wearable type, location, and weight and size	High
• Provided parameters	Medium	• Usability outcome (if available)	High
• Parameter sampling window	Medium	• Battery life	High

• Price	Medium	• All Validity-Reliability Criteria	High/Medium
• Wearable type, location, and weight and size	Medium	• Parameter sampling window	High/Medium
• Usability outcome (if available)	Medium	• Data Storage Capacity	High/Medium
• Data Storage Capacity	Medium/Low	• Data Transfer Method	Medium
• Data Transfer Method	Medium/Low	• Data Storage Method	Medium
• Data Storage Method	Medium/Low	• Charging duration	Medium
• Battery life	Low	• Charging method	Medium
• Charging duration	Low	• Raw data availability	Medium
• Charging method	Low	• Data Transfer Compatibility	Medium
• FDA approval	Low	• Waterproof	Medium/Low
• Waterproof	Low	• Bio-cueing	Medium/Low
• Bio-cueing	Low	• Biofeedback	Medium/Low
• Biofeedback	Low	• FDA approval	Low

Many devices regularly produce updated versions. This is especially true for consumer devices which are often upgraded for commercial purposes with the focus on aesthetics or usability changes. For newer versions that did not introduce new signals or a different form factor, we used the information on the previous version of the same line of product to guide the scoring. For example, studies on the Fitbit Charge 4 could guide the scoring for Fitbit Charge 5, but the Fitbit Sense cannot be used in the scoring of the Fitbit Charge 5. It was assumed that in the absence of large changes to its hardware or software (although this was a subjective judgement of the curators), the validity and reliability findings on the previous version of a given device would be indicative of the validity and reliability of the newer version.

Results

Descriptives of wearables included in the Database

General device Information

The first version of the database contains 54 devices that were primarily intended for Consumers (35), followed by Research (18) and Clinical use (10), including nine devices intended for multiple user fields (e.g., Research and Clinical). The average price of a wearable is €1013 with substantial differences in device costs ($SD = €1378$) caused by a big gap between prices of consumer and research/clinical devices. Wearables intended for consumers cost on average €347 ($SD = €175$), whereas prices of clinical and research-oriented wearables average out to €1489 ($SD = 1403$) and €2082 ($SD = €1713$), respectively. For six wearables, all intended for clinical settings, the price is only available on request. Moreover, eleven out of fifty-four devices have additional software or subscription costs. The most common type of wearables included in the database is a *watch* (40.7%) followed by a *ring* (16.7%) and *CPU with external electrodes* (16.7%). Other form factors were straps – sometimes combined with a CPU – worn on the upper arm, wrist or with flexible positioning possibilities. The location where most wearables are worn is the wrist (44.4%), followed by finger (18.5%) and chest (18.5%).

Signals

Regarding the physiological signals, that most often measured by the devices were PPG (64.8%), ECG (50.0%), Skin temperature (35.2%), blood oxygen saturation level (SpO2) (31.5%) and EDA (31.5%). The respiration signal, either derived from ICG or respiration belts, is measured by 5.6% of the devices. For some physiological parameters it is needed to record multiple physiological signals. For example, to obtain peak-valley RSA both an ECG and a respiration signal are needed. 79.6% of the devices are equipped with sensors for

multiple physiological signals. All three devices which record ICG measure ECG simultaneously.

Of the co-recorded signals that can account for confounding by posture, physical activity and ambient noise level and temperature, accelerometry was the most common (90.7% of the devices). A further 20.4% of devices also offered GPS functionality, through a built-in sensor in the device itself. 37.0% of wearables included a gyroscope. Among other signals identified, the most common one was SpO2, with 31.5% of devices providing this option, followed by ambient light detection in 25.9% of the devices.

Technical specifications

83.3% of devices were water resistant. The average battery life was 261 hours for continuously measuring devices, ranging from 10 hours for BIOPAC Research Ring and Plux respiBAN BLE to as much as 30 days for Withings ScanWatch 2. The majority (83.3%) of these wearables have a battery life of at least 24h.

Data access

All devices except two have internal storage capacity, with options to transfer the data either through Bluetooth (92.6% of the wearables), via a (micro-)SD card (7.5%) and/or using a cable (16.7%). Only the BIOPAC Research Ring and cosinuss c-med alpha do not have any internal storage capacity and require direct Bluetooth transfer of the data to the manufacturer's application, therefore requiring continuous connection between the wearable and the phone.

Regarding the data output, 35.2% of the devices provide raw data, all intended for clinical or research purposes. Among the consumer-oriented parameters provided by the devices, the most common ones are heart rate (92.6%), sleep staging (75.0%) and physical activity scores

(64.8%). ‘Stress’ as a separate parameter is provided by 33.3% of the devices, commonly referred to as *Stress Management Score*. A parameter that represents essentially the same might be named differently by manufacturers. To illustrate this issue, physical activity related scores fall under the following names: *Activity counts*, *Hourly Activity*, *Activity Score*, *Active Time*, *Active Zone Minutes*, *Exercise Tracking*, *Activity Detection*, and simply *Activity*. Even after contacting manufacturers for more detailed information on the provided parameters, we did not obtain the information needed to standardize the terminology. We, therefore, opted to include the names of parameters provided by the manufacturer.

Missing data

Twenty-four columns contain all the required information about all devices. However, not all information is provided by the manufacturers at all instances, even after requesting this information through email. This resulted in 284 NP (i.e., not provided by the manufacturer) values in the database. On average, columns contain 5.9 NP values, and most frequently the sampling rates of consumer wearables were not provided. After missing sampling rates, the column of the *Device storage capacity* contains the highest number of NPs (49). For many devices (22), details on their water resistance level are missing, and if provided lacking standardization between manufacturers in the way this information is reported.

Reliability, validity and usability

Of all devices combined, there were 96 papers in the database regarding validity and reliability and 18 for usability. The Polar H10 had the highest number of relevant validity and reliability papers (17), and the Empatica E4 had the most usability papers (5). For 31 devices no reliability, validity and usability papers were identified. 22 of the 54 devices had external validation, and 3 devices had only internal validation.

Curators' expert scores

Inter-rater reliability of the perceived usefulness of the wearables for the future large-scale cohort studies within the SiA project was calculated between all curators. The interrater agreement between each pair of curators was calculated via Pearson's correlation. Using the *psych* package in RStudio, the two-way random effects intraclass coefficient ICC (2, k) was also calculated to represent the absolute agreement between all three curators' scores (Koo & Li, 2016). Three out of the fifty-four devices in the database were not scored as 8 months have not passed since their release date. The short-term and long-term usefulness scores were provided to 51 devices. The average ICC for long-term usefulness across three curators was .85, 95% CI = [.76, .91], $F(50, 100) = 6.6, p < .001$. For the long-term usefulness scores, Pearson's r was .58 between curator 1 (Saygin) and 2 (Schoenmakers), .72 between rater 1 and 3 (Sikora), and .68 between rater 2 and 3. The average ICC for short-term usefulness across three curators was .87, 95% CI = [.78, .92], $F(50, 100) = 8.0, p < .001$. For short-term usefulness scores, the Pearson's r between curator 1 and 2 was .59, between curator 1 and 3 was .76, and between 2 and 3 was .77. Statistically significant and good (Koo & Li, 2016) absolute agreement was present across curators in both the short- and long-term usefulness scores of wearables.

The device with the highest short-term usefulness score was VU-AMS 5fs (8.7 out 10), followed by movisens EcgMove 4 (8.0), a tie between Hexoskin ProShirt and VU-AMS Core (7.7), and MindWare Mobile, Plux BioSignal kit, and Polar H10 (all three scoring 7.0). Devices with a moderately high score included the ambulatory blood pressure monitors Novacor Diasys 3 plus (6.3) and Spacelabs OnTrak (6.7), as well as BioHarness 3.0 (6.5) and Equivital Eq02+ Lifemonitor (6.8). All other devices scored less than 6 on short-term usefulness. The highest long-term use score of 8 was achieved by Empatica EmbracePlus and

Fitbit Sense 2. They were followed by Polar H10 and Garmin vivosmart 4 (7.5), Fitbit Sense with 7.3, Oura Ring gen3 with 7.3, and Garmin vivosmart 5 with 7.0. There were a number of devices that had moderately high scores for long-term use, including Garmin vivoactive 5 , WHOOP 3.0, Withings Scanwatch (all with 6.8), Empatica E4 (6.7), Fitbit Charge 5 (6.7), Corsano Cardiowatch 287-2 (6.7), WHOOP 4.0 (6.7), Google Pixel Watch 2 (6.3), Garmin vivoactive 4 (6.3), Corsano Cardiowatch 287-1 (6.0), Apple Watch Series 8 (6.0), and NOWATCH (6.0).

Illustration of the use of the SiA-WD in two different stress research scenarios

Depending on the aim of a given ambulatory study, human physiology can be recorded for durations ranging from several hours to years. In addition, the type and combination of parameters, along with the underlying signals included in different ambulatory studies vary widely. Numerous other criteria including raw data extraction, availability of provided parameters, and costs can also be fundamental decision points. Although many wearables are available for research, they may be unsuitable after considering the study plan, duration, resources, hypotheses, and theoretical framework. Our Stress in Action wearables database provides detailed information to assist in these decisions. Below, we present two different research scenarios, one for a short-term (Research Scenario A) and one for a long-term study scenario (Research Scenario B) on the activity of physiological stress systems in daily life. We first illustrate the steps one might take in using the SiA-WD to get to a shortlist of wearables that fulfil the study requirements and then review which information in the overall database as well as the RVU worksheets might be used to make a final decision. Nonetheless, researchers might also benefit from using the database in a more pragmatic manner. For instance, to check how a device they already have, or a recommended one, performs in comparison to other wearables. Therefore, device-based comparison is facilitated by the

online filtering tool, but a more structured approach to device selection, as illustrated by the scenarios below, is recommended to best support the study purpose and its design characteristics.

Research Scenario A: Short-term study of how cardiovascular threat-challenge response patterns impact health outcomes

Research Team A is interested in the impact of perceived threat and challenge in response to daily life stressors on cardiovascular health outcomes. Informed by the threat-challenge hypothesis (Blascovich & Tomaka, 1996; Tomaka et al., 1993), the team seeks to continuously record left-ventricular contractility (indexed by the pre-ejection period, PEP), cardiac output (CO), and total peripheral resistance (Wormwood et al., 2019). According to this hypothesis, a threat should be characterized by a dominant vascular response (increased total peripheral resistance), whereas challenge generates a cardiac response (shortened pre-ejection period, increased cardiac output). Thus, ambulatory impedance cardiography plus electrocardiography are required in conjunction with a cuff-based ambulatory blood pressure monitor. The main goal for this research scenario is disease prediction, and so it has a largely between-subject design. Specifically, the researchers want to investigate whether the effect of the differences of trait anxiety (as assessed by a validated questionnaire) on the health outcome of cardiometabolic/immunologic ‘allostatic load’ risk profile (as assessed by future blood sampling) (Robertson et al., 2017; Seeman et al., 2001) is mediated by a predominance of cardiovascular threat reactivity in daily life. They aim to measure perceived threat and challenge via mobile ecological momentary assessment. As a secondary research question, Team A is interested in whether physiological levels measured during sleep, in particular of cardiac vagal control, are linked to threat or challenge reactivity in daily life (Mendes et al.,

2011). To index cardiac vagal [control](#), they seek to use respiratory sinus arrhythmia (de Geus & Gevonden, 2024).

A total of 110 participants with moderate generalized anxiety disorder are scheduled to be measured within 30 weeks. Each participant's recording will take place over two days within the same week: one working and one leisure day, including sleep. As the participants will return to the lab in between their recording days for device replacement, a minimum battery life of 24 hours is [needed](#). High reliability and validity of the physiological measurements under naturalistic conditions are considered crucial. They have a budget for purchasing devices of fourteen thousand euros. The researchers have a strong preference for retrieving raw data [to](#) derive the parameters themselves. See Table 8 for an overview of the requirements for [Research Scenario A](#).

Table 8

Description of the requirements for [Research Scenario A](#)

Requirements for Research Scenario A	
• Signals	Raw ICG and ECG data (optional respiration signal)
• Parameters	PEP, RSA (optional RR), HR, BP, CO, TPR
• Participants	110 anxiety patients
• Project duration	30 weeks
• Assessment duration per participant	2 days within a week
• Maximal costs per device	6000 Euro
• Minimum battery life	24 hours
• Validity & Reliability	High

[After](#) opening the SiA Wearable Database tool to find suitable devices, they needed to decide between selecting either the *signals* (e.g., PPG or ECG, ICG, and BP) or the *parameters provided by device* (e.g., CO, PEP, BP, RSA, RR) for the first stage of filtering. The researchers opted for providing a list of *signals* they require a device to have, as they wanted

to extract the needed parameters – CO, total peripheral resistance, PEP, RSA – using their own data-analytic software rather than obtaining them directly from the device. In this case, the device would need to have ECG and ICG and, if possible with the same device, also the ability for BP recording. Because thorax impedance signals also provide a respiration proxy signal (dZ) it can be used in combination with the ECG to calculate peak-valley RSA. So, the researchers do not specify respiration as an additional required signal, but select devices that simultaneously record ICG, ECG, and BP.

The SiA-WD contained no devices that can simultaneously measure the ICG, ECG, and BP. The researchers, therefore, searched the database anew, now looking for a combination of two devices, one measuring ECG/ICG and one measuring cuff-based ambulatory BP. Their budget allows them to buy a minimum of two of the same devices that record the ICG and ECG and two that record blood pressure. This aligns with the logistic capacity and time frame of the study (4 people per week to be equipped for two days). Consequently, ICG and ECG as required signals, 24 as the minimum battery life, and six thousand euros as the maximum device cost are entered (two thousand euros are set apart for BP monitor costs to be used in the second search). As it will be a short-term study, they decide on not filtering based on the form of the wearable (e.g. restricting to wrist placement). Choosing to view their results sorted by the devices' SiA-expert score on short-term usefulness column (alternative sorting options are available such as based on the number of validity studies or device cost) they obtain the following three devices, listed from the one highest to lowest scoring: VU-AMS 5fs, MindWare Mobile, VU-AMS Core.

Following, researchers examine the raw data availability of these three devices in the database. They all allow raw data extraction. Then, each device's overall reliability-validity conclusions are read, and the device specific RVU sheet (consisting of all reliability, validity, usability papers of a device and detailed information) is investigated to understand if the particular parameters of interest were validated for a given device, along with the specifics of such validation. As they see VU-AMS 5fs was already shown to be valid for all parameters of interest, is CE approved, and has a comparatively lower cost including the data analysis software, it becomes the device of choice. They then perform a search for only ambulatory blood pressure monitors that are in-market, entering "BP" to the signal required without further filtering and obtain Novacor Diasys 3 plus, Garmin Index BPM, and SpaceLabs On Trak. As the Garmin blood pressure monitor is the only one matching the budget criteria, is FDA-cleared, and has some validation, they decided on using the Garmin Index BPM in conjunction with the VU-AMS 5fs.

Research Scenario B: A long-term study on the relations between physical activity and physiological stress-reactivity

Research team B is interested in testing the cross-stressor adaptation hypothesis of physical activity in the daily life context. The cross-stressor adaptation hypothesis posits that regular exercise leads to adaptations in stress response systems that leads to reduced physiological reactivity in response to psychological stressors (von Haaren et al., 2016). Current empirical evidence has been mixed, and no studies have conducted a prolonged assessment of the relations between physical activity and physiological stress-reactivity in daily life (Van Der Mee et al., 2023). The researchers are mainly interested in the effects of moderate-to-vigorous activity periods, which typically occur with low frequency in a general population, on stress-reactivity. Regular exercisers are expected to harvest a larger benefit in terms of more strongly attenuated stress-reactivity after a period of moderate-to-vigorous activity than

non-exercisers. In short, they seek to record physical activity and physiological reactivity across a longer period of three months in a sample of at least 300 participants, with the focus on the within-subject relations between physical activity and physiological stress-reactivity. The default method to assess stress levels in daily life is ecological momentary assessment involving repeated self-reports by smartphone beeping which is burdensome. Alternatively, this stress response can be measured via physiological reactivity after accounting for periods of activity. Accordingly, the researchers use the ‘additional heart rate’ approach (Brouwer et al., 2018). In this approach, increases in heart rate are flagged only when co-recorded physical activity indicate that the heart rate response was not simply part of homeostatic regulation in response to changed hemodynamic and metabolic demands (Verkuil et al., 2016).

Table 9

Description of the requirements for Research Scenario B

Requirements for Research Scenario B	
• Signals	PPG or ECG, and Accelerometer
• Parameters	HR, ns.SCR, and Physical Activity related parameters (optional HRV) – provided at least every 5 minutes
• Participants	≥300
• Assessment duration per participant	3 months
• Minimum battery life	48 hours
• Usability	High

Therefore, they require a wearable that can make prolonged recordings of physical activity (measured by an accelerometer) and measures of physiological arousal, like heart rate and electrodermal activity (optionally HRV). Although the reliability and validity of the physiological measurements are important, it is crucial to have many recording days within a

1 single person. As participants will be tracked [over](#) three months, they only allow wearables in
 2 the form of rings and smartwatches to increase adherence. The wearable should give a
 3 minute-by-minute index for physical activity (e.g., vector magnitude or the number of steps)
 4 and for general physiological arousal (5-minute epoch with an increase in HR or ns.SCR of
 5 more than 20% compared to previous epoch). See Table 9 for an overview of the
 6 requirements for [Research Scenario B](#). The researchers filter based on the parameters
 7 provided by the device. Considering that combining increased HR and increased ns.SCR may
 8 improve estimation of the physiological arousal level, they opt for selecting devices that can
 9 readily provide *both* of these parameters for at most 5 min time windows along with physical
 10 activity related parameters. As the researchers think having the participants charge their
 11 wearable every two days is a reasonable frequency, minimum battery life is set at 48 hours,
 12 and the resulting devices were sorted based on their SiA long-term usefulness scores, from
 13 highest to lowest scoring, resulting in: Fitbit Sense 2 (8 points), Empatica EmbracePlus (8)
 14 Fitbit Charge 5 (6.7), and NOWATCH (6). All devices provide Distance, Steps, and Active
 15 Zone Minutes parameters, which can be used to index physical activity. The details [of the](#)
 16 SiA-WD show the Fitbit charge 5 does not provide continuous but manually initiated scan-
 17 based values, which does not suit the requirements and is therefore excluded. Similarly,
 18 NOWATCH measures EDA but does not provide parameters derived from the signal.
 19 Empatica EmbracePlus has one validity study, which only examined the blood oxygen
 20 saturation levels. Fitbit Sense 2 lacks [reliability, validity or usability](#) studies. When checking
 21 the General Validity and Reliability synthesis column, it is seen that although Sense 2 does
 22 not have validation studies, its predecessor device Fitbit Sense's HR was found to have
 23 overall acceptable validity under static as well as active conditions, and its skin conductance
 24 level had significant positive correlation with the reference device parameters. Because the
 25 Fitbit Sense 2 is defined as a successor of this previous release, the researchers assume the

latest version to perform up to similar validity standards as the previous version. As the study will be carried out in Europe, they check the GDPR column to make sure the device's way of storing data meets the data protection regulations required by law and choose Fitbit Sense 2 for the study.

Discussion

Selecting the right wearable for stress research can be challenging in face of the large number of physiological wearables on the market paired with the lack of a comprehensive, systematic, and iteratively updated overview of the relevant wearable characteristics (Connelly et al., 2021; Dobson et al., 2023). The Stress in Action Wearables Database (SiA-WD) provides such an overview, enabling stress researchers to compare a host of wearable devices on a number of research-informed characteristics with detail. Compared to other overviews of wearable devices such as systematic reviews and databases (Iqbal et al., 2016; Lu et al., 2023; Paredes et al., 2021; Vijayan et al., 2021), the SiA-WD has several major advantages. First, the database is comprehensive regarding the information on general device information, technical specifications, and data access. Second, the technical, practical, and importantly, physiological details were obtained with a *systematic* review including the validity and reliability of the measured physiological parameters. This is – to the best of our knowledge – a first in the field. Third, where other overviews may have a specific focus, [such as](#) the focus on cardiovascular parameters by Lu and colleagues (2023), the SiA-WD has a broad scope of signals and devices relevant for physiological research focusing on autonomic nervous activity. Thus, the database applies to a relatively larger audience of researchers in the field. Fourth, this systematic search included a usability assessment also with a systematic search of studies that might have investigated its user-friendliness. The curators' summary

statements on [reliability, validity and usability](#) papers make an easy and time-efficient selection of devices possible, while the device-specific [reliability, validity, usability](#) (RVU worksheets) allow access to the evidence supporting these summary statements. And fifth, where other overviews get outdated, the SiA Wearables Database will be [consistently extended and updated](#).

Challenges encountered

Populating the database was not without its challenges. As some manufacturers did not provide the required information, they were sent a standardized email requesting these details. Most often, information on the particulars of physiological sensors, a comprehensive list of parameters provided by the device, and the sampling windows of parameters (e.g., the number of seconds over which each heart rate is calculated) were lacking. In the case that no response addressing the question(s) was received, information fields had to be filled with *NP* (*not provided by the manufacturer*). For example, columns specifying the signals' sampling frequency or the device storage capacity contained frequent NPs, because this information could not be retrieved. Another challenge in populating the database concerns the lack of clear definitions of the output parameters of the wearable systems. For example, heart rate can be calculated continuously per five-minute intervals, per day, per activity, [and so on](#) but the used time scale [was](#) not always available for data extraction. Moreover, upon filling in the database it became apparent that researchers conducting long term ambulatory studies may use external applications as a solution, as it facilitates more detailed [data](#) extraction and storing of the recorded signals and parameters; e.g., the Heart Rate Variability Logger application. Finally, conceptually similar parameters go by many different names, as exemplified by the heterogeneous terminology used for physical activity.

1

2 The SiA Wearables Database version 1.0 includes more devices that are consumer grade
3 rather than research grade devices. This probably explains why more than half of the
4 wearable systems do not provide raw data access and instead focus on providing parameters
5 such as HR, sleep stages, stress scores and activity values. Wristwatch was the most used
6 wearable type, and ECG and PPG were the signals most often recorded, with most devices
7 also co-recording accelerometry. A watch is a convenient and generally accepted way of
8 wearing a device, and the integration of PPG sensors and accelerometers can provide
9 parameters (e.g., HR values and activity scores) that are informative for a broad audience.
10 Due to its relatively lower burden (for example, as compared to a vest), it also enables higher
11 adherence and comfort for populations with higher risk factors. Consumer grade wearables
12 are substantially cheaper and [more](#) user-friendly than research grade wearables. Compared to
13 research-grade wearables, consumer-based wearables are therefore much more suitable for
14 research questions that need to assess many people and/or for an extended period.

15

16 However, many parameters produced by consumer devices lack validation as shown by the
17 overview of the validity and reliability syntheses. In terms of established validity and
18 reliability, research-grade wearables clearly outperform consumer wearables. Research-grade
19 devices also more often provide raw data and record multiple different signals, such as
20 continuous ECG and ICG. A full continuous ECG is rarely recorded in consumer wearables,
21 and those that record ECG do so by having the user manually initiate a short ECG recording
22 by placement of the hand on top of the device (e.g., to the side of a watch). Furthermore, the
23 participants are typically asked to refrain from any physical movement which might create
24 stress [because](#) of multiple failed recordings (Seshadri et al., 2020). A research-grade device in

a low-burden wearable format such a wristwatch **could** have high utility for research interested in assessing autonomic and physical activity parameters for more fine-tuned windows and over longer periods of time.

The results highlighted the lack of both validation and usability research on existing wearables, as well as the need for a more systematic approach to such studies, especially when one considers the high tempo with which devices get discontinued and introduced to the market. Although systematic validation studies are performed for some wearables, the reported statistics vary widely (e.g., Bland-Altman plots, intra-class correlations, linear regression, half-split reliability) which complicates comparisons between studies and across devices. This constitutes a challenge for both the curators and users of the database to make a fine-tuned comparison between those devices that are deemed valid. We hope that the gaps in validation studies identified by the SiA Wearables Database will create incentives for studies assessing the **reliability, validity and usability** of wearable technology in a more consistent manner with regards to study design and statistical reporting (Keogh et al., 2021; Liang et al., 2018; Shei et al., 2022; van Lier et al., 2020).

The two selected research scenarios where the SiA Wearables Database illustrate the need for such a database. There is rarely a clear ‘winner’ device that fits a study design. Instead, the optimal device choice depends on a consideration of many aspects of the research study, including the theoretical relevance of the physiological parameters to be measured, the total number of participants, the duration of the sampling, the burden threshold that can be tolerated by the participants, and financial and logistical constraints of the research team. Filtering by the relevant columns in the SiA-WD based on their study requirements should

provide researchers with a good first selection of wearables. Afterwards, they can narrow it down to the optimal wearable by further inspection of details, for example that on validity.

The two presented [scenarios](#) showed substantially different research goals, one of obtaining detailed information on the daily operation of physiological systems during a short measurement to predict future disease risk, the other to obtain insight into the temporal dynamics of physical activity, affect and physiological arousal over a prolonged period within-individuals. Suitability of a wearable for these typical short-term and long-term scenarios was added as a separate score to the SiA-WD to assist researchers in selecting the potential wearables for their research. Intermediate research scenarios exist that may require researchers to carefully formulate what is of importance.

[Limitations of the SiA-WD](#)

A major limitation of the current version 1.0 of the SiA-WD is that it contains a modest selection of [54](#) devices, whereas over 160 devices were identified by our search. The [54](#) devices were selected for version 1.0 based on repeated use of a device in past stress research, input obtained from experts within the SiA consortium, and our aim to cover different types of devices (i.e., focusing on signals, wearable types, and consumer, research and clinical grade devices). While we succeeded to build a structure suitable for the wide variety of devices on the market, this first version may suffer from selection bias. The SiA-WD version 1.0 only includes wearables measuring autonomic nervous system reactivity, and no wearables capturing central nervous system activity. The gradual expansion of the database in the upcoming renewal cycles will at some point add these devices too, and also gradually reduce other selection biases. The continued renewal also ensures that we can pick up new promising devices that move from prototype to product in this rapidly moving field of

technology and highlight the need for their subsequent validation, promoting the use of new technology. For example, we are especially awaiting to include in the database non-invasive wearables assessing continuous blood cortisol levels which may interest many stress and health researchers (Parlak et al., 2018).

It is not untypical in ambulatory research to extract the raw data recorded by the wearable, and subsequently conduct preprocessing and parameter extraction in an open-access third party platform such as Python or R. Upon the extraction of raw data, one might either use a more generalist package such as *neurokit2* that can be implemented to data from many wearables (Makowski et al., 2021) or utilize a more streamlined toolkit that was built for particular wearables, like the [Wearables International repository](#). To facilitate this, we report in the SiA-WD whether the wearables enable raw data extraction as specified by the manufacturer.

Another limitation is that SiA-WD is now maintained by a relatively modest-sized team. Not all candidate devices can be annotated for the database at once, as substantial work is involved per device (the time investment ranges from .5 to 5 workdays). Paired to the rapid expansion of the total pool of wearables, this means that the curators may not keep up and will need to prioritize which devices first to annotate. Basic rules for this prioritization are that a device, listed in no order of importance (1) has a novel or attractive feature (e.g., new way to monitor a widely used physiological parameter, large extension of battery life), (2) combines multiple physiological signals relevant to stress research (e.g., a device measures continuous ECG, BP and EDA), (3) has the potential to substantially decrease participant burden (e.g., a ring rather than a patch), or (4) has been repeatedly used for peer-reviewed stress research and is still in-market. Also, the SiA-WD will maintain a clear focus on non-invasive wearables, therefore excluding ingestibles, implants and domotic sensors. However,

just like through the use of AS Review, we will stay open to implementing technologies that could benefit the maintenance of the database. For example, with the continuous improvements in the field of AI, Large Language Models (LLMs) could potentially be implemented in the future to synthesize the RVU information. Nonetheless, with the variety of tasks involved in populating the SiA-WD, we see the human-in-the-loop principle as central to the decision-making process. We believe that in this way sustainability of the database will not be dependent on a given technology and as a result SiA-WD will be more responsive to change, including adoption of new, beneficial tools.

Conclusion

The Stress in Action Wearables Database (SiA-WD) is a comprehensive and well-sustained database of physiological wearable devices that have application potential in behavioral research, in particular stress research. It provides a large amount of information that a researcher would look for such as the general device information, recorded signals, technical specifications and data access, combined with a systematic reliability, validity and usability review of the available literature on a device. The SiA-WD will be iteratively expanded and the information, including that for devices already existing in the database, updated for an period of at least ten years. A user-friendly tool will enable researchers to conveniently select the most suitable wearable for their study. The wearable database will remain to be moderated by a team of SiA researchers, but a future goal is to allow a broader group of researchers to actively contribute to this effort. They could propose devices to be added, point to additional identified reliability, validity and usability studies, and suggest other points of information to be added to the database based on their user experience.

1 Funding

2 This work is funded by Stress in Action. The research project ‘Stress in Action’: [www.stress-](http://www.stress-in-action.nl)
3 [in-action.nl](http://www.stress-in-action.nl) is financially supported by the Dutch Research Council and the Dutch Ministry of
4 Education, Culture and Science (NWO gravitation grant number 024.005.010).

5 Conflict of interest

6 Prof. dr. Eco de Geus owns 4.9% of the shares in VU Ambulatory Monitoring Solutions Ltd.

7 Ethical approval

8 Not applicable.

9 Consent to participate

10 Not applicable.

11 Consent for publication

12 All authors provided consent for the publication of the paper.

13 Availability of data and materials

14 Upon request the individual scores of the curators along with the code used to calculate the
15 inter-rater reliability can be provided. The database is openly accessible at
16 <https://osf.io/umgvp/> [the database will be added to this public OSF page upon acceptance,
17 for now please see Supplementary 1].

18 Code availability

19 Not applicable.

20 Authors’ contributions

21 Myrte Schoenmakers: conceptualization, methodology, formal analysis, data collection, data
22 curation, Writing – Original Draft, Writing – Review & Editing, Visualization

23 Melisa Saygin: conceptualization, methodology, formal analysis, data collection, data
24 curation, Writing – Original Draft, Writing – Review & Editing, Visualization

25 Magdalena Sikora: conceptualization, methodology, formal analysis, data collection, data
26 curation, Writing – Original Draft, Writing – Review & Editing, Visualization

- 1 Thomas Vaessen: Writing – Review & Editing, Visualization
 2 Matthijs Noordzij: conceptualization, Writing – Review & Editing, Visualization,
 3 Supervision, Funding acquisition
 4 Eco de Geus: conceptualization, Writing – Review & Editing, Visualization, Supervision,
 5 Funding acquisition

6 Open Practices Statement

- 7 The work and materials presented in the current paper were not preregistered.
 8 All data and materials, including the database, are openly available at <https://osf.io/umgvp/>
 9 [the database will be added to this public OSF page upon acceptance, for now please see
 10 Supplementary 1].

11 References

- 12 Akinwande, D., & Kireev, D. (2019). Wearable graphene sensors use ambient light to monitor health.
 13 *Nature*, 576(7786), 220–221. <https://doi.org/10.1038/d41586-019-03483-7>
 14 Alian, A. A., & Shelley, K. H. (2014). Photoplethysmography. *Best Practice & Research. Clinical*
 15 *Anaesthesiology*, 28(4), 395–406. <https://doi.org/10.1016/j.bpa.2014.08.006>
 16 Areia, C., Young, L., Vollam, S., Ede, J., Santos, M., Tarassenko, L., & Watkinson, P. (2020).
 17 Wearability Testing of Ambulatory Vital Sign Monitoring Devices: Prospective Observational
 18 Cohort Study. *JMIR mHealth and uHealth*, 8(12), e20214. <https://doi.org/10.2196/20214>
 19 Beauchaine, T. P., & Gatzke-Kopp, L. M. (2012). Instantiating the multiple levels of analysis
 20 perspective in a program of study on externalizing behavior. *Development and*
 21 *Psychopathology*, 24(3), 1003–1018. <https://doi.org/10.1017/S0954579412000508>
 22 Blascovich, J., & Tomaka, J. (1996). The biopsychosocial model of arousal regulation. In *Advances in*
 23 *experimental social psychology*, Vol. 28 (pp. 1–51). Academic Press.
 24 [https://doi.org/10.1016/S0065-2601\(08\)60235-X](https://doi.org/10.1016/S0065-2601(08)60235-X)
 25 Boateng, G., Motti, V. G., Mishra, V., Batsis, J. A., Hester, J., & Kotz, D. (2019). Experience: Design,
 26 Development and Evaluation of a Wearable Device for mHealth Applications. *Proceedings of*
 27 *the .. Annual International Conference on Mobile Computing and Networking. International*

Conference on Mobile Computing and Networking, 2019, 31.

<https://doi.org/10.1145/3300061.3345432>

Boucsein, W., Fowles, D. C., Grimnes, S., Ben-Shakhar, G., Roth, W. T., Dawson, M. E., Filion, D. L., &

Society for Psychophysiological Research Ad Hoc Committee on Electrodermal Measures.

(2012). Publication recommendations for electrodermal measurements. *Psychophysiology*,

49(8), 1017–1034. <https://doi.org/10.1111/j.1469-8986.2012.01384.x>

Brouwer, A.-M., Van Dam, E., Van Erp, J. B. F., Spangler, D. P., & Brooks, J. R. (2018). Improving Real-

Life Estimates of Emotion Based on Heart Rate: A Perspective on Taking Metabolic Heart

Rate Into Account. *Frontiers in Human Neuroscience*, 12, 284.

<https://doi.org/10.3389/fnhum.2018.00284>

Byrom, B., Watson, C., Doll, H., Coons, S. J., Eremenco, S., Ballinger, R., Mc Carthy, M., Crescioni, M.,

O'Donohoe, P., & Howry, C. (2018). Selection of and Evidentiary Considerations for Wearable

Devices and Their Measurements for Use in Regulatory Decision Making: Recommendations

from the ePRO Consortium. *Value in Health*, 21(6), 631–639.

<https://doi.org/10.1016/j.jval.2017.09.012>

Connelly, K., Molchan, H., Bidanta, R., Siddh, S., Lowens, B., Caine, K., Demiris, G., Siek, K., & Reeder,

B. (2021). Evaluation framework for selecting wearable activity monitors for research.

mHealth, 7, 6–6. <https://doi.org/10.21037/mhealth-19-253>

De Geus, E., & Gevonden, M. (2024). Chapter 6: Acquisition and analysis of ambulatory autonomic

nervous system data. In *Mobile sensing in psychology: Methods and applications*. The

Guilford press.

De Geus, E. J. C., Willemsen, G. H. M., Klaver, C. H. A. M., & Van Doornen, L. J. P. (1995). Ambulatory

measurement of respiratory sinus arrhythmia and respiration rate. *Biological Psychology*,

41(3), 205–227. [https://doi.org/10.1016/0301-0511\(95\)05137-6](https://doi.org/10.1016/0301-0511(95)05137-6)

- de Oliveira Matos, F., Vido, A., Garcia, W. F., Lopes, W. A., & Pereira, A. (2020). A Neurovisceral Integrative Study on Cognition, Heart Rate Variability, and Fitness in the Elderly. *Frontiers in Aging Neuroscience*, 12, 51. <https://doi.org/10.3389/fnagi.2020.00051>
- Dias, A., Fisterer, B., Lamla, G., Kuhn, K., Hartvigsen, G., & Horsch, A. (2009). Measuring physical activity with sensors: A qualitative study. *Studies in Health Technology and Informatics*, 150, 475–479.
- Dobson, R., Stowell, M., Warren, J., Tane, T., Ni, L., Gu, Y., McCool, J., & Whittaker, R. (2023). Use of Consumer Wearables in Health Research: Issues and Considerations. *Journal of Medical Internet Research*, 25, e52444. <https://doi.org/10.2196/52444>
- Dolson, C. M., Harlow, E. R., Phelan, D. M., Gabbett, T. J., Gaal, B., McMellen, C., Geletka, B. J., Calcei, J. G., Voos, J. E., & Seshadri, D. R. (2022). Wearable Sensor Technology to Predict Core Body Temperature: A Systematic Review. *Sensors*, 22(19), 7639. <https://doi.org/10.3390/s22197639>
- Eckberg, D. L. (2003). The human respiratory gate. *The Journal of Physiology*, 548(Pt 2), 339–352. <https://doi.org/10.1113/jphysiol.2002.037192>
- Ehmen, H., Haesner, M., Steinke, I., Dorn, M., Gövercin, M., & Steinhagen-Thiessen, E. (2012). Comparison of four different mobile devices for measuring heart rate and ECG with respect to aspects of usability and acceptance by older people. *Applied Ergonomics*, 43(3), 582–587. <https://doi.org/10.1016/j.apergo.2011.09.003>
- El-Hamad, F., Ernst, H., Schmidt, M., Shahrabaki, S. S., & Baumert, M. (2023). Contributors to beat-to-beat stroke volume variability during acute mental stress in healthy volunteers. *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, 1–4. <https://doi.org/10.1109/EMBC40787.2023.10340083>

- Giakoumis, D., Drosou, A., Cipresso, P., Tzovaras, D., Hassapis, G., Gaggioli, A., & Riva, G. (2012). Using Activity-Related Behavioural Features towards More Effective Automatic Stress Detection. *PLoS ONE*, 7(9), e43571. <https://doi.org/10.1371/journal.pone.0043571>
- Giurgiu, M., Timm, I., Becker, M., Schmidt, S., Wunsch, K., Nissen, R., Davidovski, D., Bussmann, J. B. J., Nigg, C. R., Reichert, M., Ebner-Priemer, U. W., Woll, A., & Von Haaren-Mack, B. (2022). Quality Evaluation of Free-living Validation Studies for the Assessment of 24-Hour Physical Behavior in Adults via Wearables: Systematic Review. *JMIR mHealth and uHealth*, 10(6), e36377. <https://doi.org/10.2196/36377>
- Grossman, P., & Svebak, S. (1987). Respiratory Sinus Arrhythmia as an Index of Parasympathetic Cardiac Control During Active Coping. *Psychophysiology*, 24(2), 228–235. <https://doi.org/10.1111/j.1469-8986.1987.tb00284.x>
- Haddad, M., Hermassi, S., Aganovic, Z., Dalansi, F., Kharbach, M., Mohamed, A. O., & Bibi, K. W. (2020). Ecological Validation and Reliability of Hexoskin Wearable Body Metrics Tool in Measuring Pre-exercise and Peak Heart Rate During Shuttle Run Test in Professional Handball Players. *Frontiers in Physiology*, 11, 957. <https://doi.org/10.3389/fphys.2020.00957>
- Henley, B. C., Shokouhi, M., Mahajan, A. Y., Inan, O. T., & Hajjar, I. (2018). Cardiovascular Response to Mental Stress in Mild Cognitive Impairment and its Association with Cerebral Perfusion. *Journal of Alzheimer's Disease*, 63(2), 645–654. <https://doi.org/10.3233/JAD-180036>
- Hopkins, W. G. (2000). Measures of Reliability in Sports Medicine and Science: *Sports Medicine*, 30(1), 1–15. <https://doi.org/10.2165/00007256-200030010-00001>
- Iqbal, M. H., Aydin, A., Brunckhorst, O., Dasgupta, P., & Ahmed, K. (2016). A review of wearable technology in medicine. *Journal of the Royal Society of Medicine*, 109(10), 372–380. <https://doi.org/10.1177/0141076816663560>
- Irwin, M. R. (2023). Sleep disruption induces activation of inflammation and heightens risk for infectious disease: Role of impairments in thermoregulation and elevated ambient

temperature. *Temperature: Multidisciplinary Biomedical Journal*, 10(2), 198–234.

<https://doi.org/10.1080/23328940.2022.2109932>

Jeffer, E., Vollam, S., Young, J. D., Horsington, L., Lynch, B., & Watkinson, P. J. (2016). Wearable monitors for patients following discharge from an intensive care unit: Practical lessons learnt from an observational study. *Journal of Advanced Nursing*, 72(8), 1851–1862.

<https://doi.org/10.1111/jan.12959>

Johnston, W., Judice, P. B., Molina García, P., Mühlen, J. M., Lykke Skovgaard, E., Stang, J., Schumann, M., Cheng, S., Bloch, W., Brønd, J. C., Ekelund, U., Grøntved, A., Caulfield, B., Ortega, F. B., & Sardinha, L. B. (2021). Recommendations for determining the validity of consumer wearable and smartphone step count: Expert statement and checklist of the INTERLIVE network. *British Journal of Sports Medicine*, 55(14), 780–793.

<https://doi.org/10.1136/bjsports-2020-103147>

Keogh, A., Argent, R., Anderson, A., Caulfield, B., & Johnston, W. (2021). Assessing the usability of wearable devices to measure gait and physical activity in chronic conditions: A systematic review. *Journal of NeuroEngineering and Rehabilitation*, 18(1), 138.

<https://doi.org/10.1186/s12984-021-00931-2>

Kim, H.-G., Cheon, E.-J., Bai, D.-S., Lee, Y. H., & Koo, B.-H. (2018). Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature. *Psychiatry Investigation*, 15(3), 235–245.

<https://doi.org/10.30773/pi.2017.08.17>

Kleckner, I. R., Feldman, M. J., Goodwin, M. S., & Quigley, K. S. (2021). Framework for selecting and benchmarking mobile devices in psychophysiological research. *Behavior Research Methods*, 53(2), 518–535. <https://doi.org/10.3758/s13428-020-01438-9>

Klimek, A., Mannheim, I., Schouten, G., Wouters, E. J. M., & Peeters, M. W. H. (2023). Wearables measuring electrodermal activity to assess perceived stress in care: A scoping review. *Acta Neuropsychiatrica*, 1–11. <https://doi.org/10.1017/neu.2023.19>

- 1 Koo, T. K., & Li, M. Y. (2016). A Guideline of Selecting and Reporting Intraclass Correlation
2 Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155–163.
3 <https://doi.org/10.1016/j.jcm.2016.02.012>
- 4 Kottner, J., Audigé, L., Brorson, S., Donner, A., Gajewski, B. J., Hróbjartsson, A., Roberts, C., Shoukri,
5 M., & Streiner, D. L. (2011). Guidelines for Reporting Reliability and Agreement Studies
6 (GRRAS) were proposed. *Journal of Clinical Epidemiology*, 64(1), 96–106.
7 <https://doi.org/10.1016/j.jclinepi.2010.03.002>
- 8 Krittanawong, C., Rogers, A. J., Johnson, K. W., Wang, Z., Turakhia, M. P., Halperin, J. L., & Narayan, S.
9 M. (2021). Integration of novel monitoring devices with machine learning technology for
10 scalable cardiovascular management. *Nature Reviews. Cardiology*, 18(2), 75–91.
11 <https://doi.org/10.1038/s41569-020-00445-9>
- 12 Liang, J., Xian, D., Liu, X., Fu, J., Zhang, X., Tang, B., & Lei, J. (2018). Usability Study of Mainstream
13 Wearable Fitness Devices: Feature Analysis and System Usability Scale Evaluation. *JMIR*
14 *mHealth and uHealth*, 6(11), e11066. <https://doi.org/10.2196/11066>
- 15 Liu, Y., & Han, M. (2022). Determining the Key Factors of Wearable Devices Consumers' Adoption
16 Behavior Based on an MADM Model for Product Improvement. *IEEE Transactions on*
17 *Engineering Management*, 69(6), 4036–4051. <https://doi.org/10.1109/TEM.2019.2960499>
- 18 Lu, J. K., Sijm, M., Janssens, G. E., Goh, J., & Maier, A. B. (2023). Remote monitoring technologies for
19 measuring cardiovascular functions in community-dwelling adults: A systematic review.
20 *GeroScience*, 45(5), 2939–2950. <https://doi.org/10.1007/s11357-023-00815-4>
- 21 Majumder, S., Mondal, T., & Deen, M. (2017). Wearable Sensors for Remote Health Monitoring.
22 *Sensors*, 17(1), 130. <https://doi.org/10.3390/s17010130>
- 23 Makowski, D., Pham, T., Lau, Z. J., Brammer, J. C., Lespinasse, F., Pham, H., Schölzel, C., & Chen, S. H.
24 A. (2021). NeuroKit2: A Python toolbox for neurophysiological signal processing. *Behavior*
25 *Research Methods*, 53(4), 1689–1696. <https://doi.org/10.3758/s13428-020-01516-y>

- 1 Malm, S., Frigstad, S., Sagberg, E., Larsson, H., & Skjaerpe, T. (2004). Accurate and reproducible
2 measurement of left ventricular volume and ejection fraction by contrast echocardiography.
3 *Journal of the American College of Cardiology*, 44(5), 1030–1035.
4 <https://doi.org/10.1016/j.jacc.2004.05.068>
- 5 Mannée, D., De Jongh, F., & Van Helvoort, H. (2021). The Accuracy of Tidal Volume Measured With a
6 Smart Shirt During Tasks of Daily Living in Healthy Subjects: Cross-sectional Study. *JMIR*
7 *Formative Research*, 5(10), e30916. <https://doi.org/10.2196/30916>
- 8 McStay, S. (2019). Recording a 12-lead electrocardiogram (ECG). *British Journal of Nursing*, 28(12),
9 756–760. <https://doi.org/10.12968/bjon.2019.28.12.756>
- 10 Molina-Garcia, P., Notbohm, H. L., Schumann, M., Argent, R., Hetherington-Rauth, M., Stang, J.,
11 Bloch, W., Cheng, S., Ekelund, U., Sardinha, L. B., Caulfield, B., Brønd, J. C., Grøntved, A., &
12 Ortega, F. B. (2022). Validity of Estimating the Maximal Oxygen Consumption by Consumer
13 Wearables: A Systematic Review with Meta-analysis and Expert Statement of the INTERLIVE
14 Network. *Sports Medicine (Auckland, N.Z.)*, 52(7), 1577–1597.
15 <https://doi.org/10.1007/s40279-021-01639-y>
- 16 Moon, K. S., Lee, S. Q., Kang, J. S., Hnat, A., & Karen, D. B. (2023). A Wireless Electrooculogram (EOG)
17 Wearable Using Conductive Fiber Electrode. *Electronics*, 12(3), 571.
18 <https://doi.org/10.3390/electronics12030571>
- 19 Mühlen, J. M., Stang, J., Lykke Skovgaard, E., Judice, P. B., Molina-Garcia, P., Johnston, W., Sardinha,
20 L. B., Ortega, F. B., Caulfield, B., Bloch, W., Cheng, S., Ekelund, U., Brønd, J. C., Grøntved, A.,
21 & Schumann, M. (2021). Recommendations for determining the validity of consumer
22 wearable heart rate devices: Expert statement and checklist of the INTERLIVE Network.
23 *British Journal of Sports Medicine*, 55(14), 767–779. [https://doi.org/10.1136/bjsports-2020-](https://doi.org/10.1136/bjsports-2020-103148)
24 103148

- 1 Mukkamala, R., Hahn, J.-O., Inan, O. T., Mestha, L. K., Kim, C.-S., Töreyn, H., & Kyal, S. (2015).
2 Toward Ubiquitous Blood Pressure Monitoring via Pulse Transit Time: Theory and Practice.
3 *IEEE Transactions on Bio-Medical Engineering*, 62(8), 1879–1901.
4 <https://doi.org/10.1109/TBME.2015.2441951>
- 5 Neumann, E., & Blanton, R. (1970). The early history of electrodermal research. *Psychophysiology*,
6 6(4), 453–475. <https://doi.org/10.1111/j.1469-8986.1970.tb01755.x>
- 7 Ngo, C., Munoz, C., Lueken, M., Hülkenberg, A., Bollheimer, C., Briko, A., Kobelev, A., Shchukin, S., &
8 Leonhardt, S. (2022). A Wearable, Multi-Frequency Device to Measure Muscle Activity
9 Combining Simultaneous Electromyography and Electrical Impedance Myography. *Sensors*,
10 22(5), 1941. <https://doi.org/10.3390/s22051941>
- 11 O’Brien, J. T., Gallagher, P., Stow, D., Hammerla, N., Ploetz, T., Firbank, M., Ladha, C., Ladha, K.,
12 Jackson, D., McNaney, R., Ferrier, I. N., & Olivier, P. (2017). A study of wrist-worn activity
13 measurement as a potential real-world biomarker for late-life depression. *Psychological*
14 *Medicine*, 47(1), 93–102. <https://doi.org/10.1017/S0033291716002166>
- 15 Osei, J., Vaccarino, V., Wang, M., Shah, A. S., Lampert, R., Li, L. Y., Ko, Y.-A., Pearce, B. D., Kutner, M.,
16 Garcia, E. V., Piccinelli, M., Raggi, P., Bremner, J. D., Quyyumi, A. A., Sun, Y. V., Ahmed, H.,
17 Haddad, G., Daaboul, O., Roberts, T., ... Shah, A. J. (2024). Stress-Induced Autonomic
18 Dysfunction is Associated With Mental Stress–Induced Myocardial Ischemia in Patients With
19 Coronary Artery Disease. *Circulation: Cardiovascular Imaging*, 17(6), e016596. Scopus.
20 <https://doi.org/10.1161/CIRCIMAGING.124.016596>
- 21 Pantelopoulos, A., & Bourbakis, N. (2008). A survey on wearable biosensor systems for health
22 monitoring. *2008 30th Annual International Conference of the IEEE Engineering in Medicine*
23 *and Biology Society*, 4887–4890. <https://doi.org/10.1109/IEMBS.2008.4650309>

- 1 Pantelopoulos, A., & Bourbakis, N. G. (2010). A Survey on Wearable Sensor-Based Systems for Health
2 Monitoring and Prognosis. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*
3 *(Applications and Reviews)*, 40(1), 1–12. <https://doi.org/10.1109/TSMCC.2009.2032660>
- 4 Paredes, L., McMillan, C., Chan, W. K., Chandrasegaran, S., Singh, R., Ramani, K., & Wilde, D. (2021).
5 CHIMERA: Supporting Wearables Development across Multidisciplinary Perspectives.
6 *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 5(4),
7 1–24. <https://doi.org/10.1145/3494974>
- 8 Parlak, O., Keene, S. T., Marais, A., Curto, V. F., & Salleo, A. (2018). Molecularly selective nanoporous
9 membrane-based wearable organic electrochemical device for noninvasive cortisol sensing.
10 *Science Advances*, 4(7), eaar2904. <https://doi.org/10.1126/sciadv.aar2904>
- 11 Parry, M. J. E., & McFetridge-Durdle, J. (2006). Ambulatory Impedance Cardiography: A Systematic
12 Review. *Nursing Research*, 55(4), 283.
- 13 Patel, S., Park, H., Bonato, P., Chan, L., & Rodgers, M. (2012). A review of wearable sensors and
14 systems with application in rehabilitation. *Journal of NeuroEngineering and Rehabilitation*,
15 9(1), 21. <https://doi.org/10.1186/1743-0003-9-21>
- 16 Patel, V., Orchanian-Cheff, A., & Wu, R. (2021). Evaluating the Validity and Utility of Wearable
17 Technology for Continuously Monitoring Patients in a Hospital Setting: Systematic Review.
18 *JMIR mHealth and uHealth*, 9(8), e17411. <https://doi.org/10.2196/17411>
- 19 Pattyn, N., Migeotte, P.-F., Neyt, X., Den Nest, A. V., & Cluydts, R. (2010). Comparing real-life and
20 laboratory-induced stress reactivity on cardio-respiratory parameters: Differentiation of a
21 tonic and a phasic component. *Physiology & Behavior*, 101(2), 218–223.
22 <https://doi.org/10.1016/j.physbeh.2010.04.037>
- 23 Peake, J. M., Kerr, G., & Sullivan, J. P. (2018). A Critical Review of Consumer Wearables, Mobile
24 Applications, and Equipment for Providing Biofeedback, Monitoring Stress, and Sleep in

- Physically Active Populations. *Frontiers in Physiology*, 9, 743.
<https://doi.org/10.3389/fphys.2018.00743>
- Pevnick, J. M., Birkeland, K., Zimmer, R., Elad, Y., & Kedan, I. (2018). Wearable technology for cardiology: An update and framework for the future. *Trends in Cardiovascular Medicine*, 28(2), 144–150. <https://doi.org/10.1016/j.tcm.2017.08.003>
- Polhemus, A. M., Novák, J., Ferrao, J., Simblett, S., Radaelli, M., Locatelli, P., Matcham, F., Kerz, M., Weyer, J., Burke, P., Huang, V., Dockendorf, M. F., Temesi, G., Wykes, T., Comi, G., Myin-Germeys, I., Folarin, A., Dobson, R., Manyakov, N. V., ... Hotopf, M. (2020). Human-Centered Design Strategies for Device Selection in mHealth Programs: Development of a Novel Framework and Case Study. *JMIR mHealth and uHealth*, 8(5), e16043.
<https://doi.org/10.2196/16043>
- Rahma, O. N., Putra, A. P., Rahmatillah, A., Putri, Y. S. K. A., Fajriaty, N. D., Ain, K., & Chai, R. (2022). Electrodermal Activity for Measuring Cognitive and Emotional Stress Level. *Journal of Medical Signals and Sensors*, 12(2), 155–162. https://doi.org/10.4103/jmss.JMSS_78_20
- Robertson, T., Beveridge, G., & Bromley, C. (2017). Allostatic load as a predictor of all-cause and cause-specific mortality in the general population: Evidence from the Scottish Health Survey. *PLOS ONE*, 12(8), e0183297. <https://doi.org/10.1371/journal.pone.0183297>
- Romeijn, N., Raymann, R. J. E. M., Møst, E., Te Lindert, B., Van Der Meijden, W. P., Fronczek, R., Gomez-Herrero, G., & Van Someren, E. J. W. (2012). Sleep, vigilance, and thermosensitivity. *Pflugers Archiv: European Journal of Physiology*, 463(1), 169–176.
<https://doi.org/10.1007/s00424-011-1042-2>
- Sadeghi, R., Banerjee, T., Hughes, J. C., & Lawhorne, L. W. (2019). Sleep quality prediction in caregivers using physiological signals. *Computers in Biology and Medicine*, 110, 276–288.
<https://doi.org/10.1016/j.compbiomed.2019.05.010>

- 1 Sano, A., & Picard, R. W. (2013). Stress Recognition Using Wearable Sensors and Mobile Phones.
2 *2013 Humaine Association Conference on Affective Computing and Intelligent Interaction*,
3 671–676. <https://doi.org/10.1109/ACII.2013.117>
- 4 Seeman, T. E., McEwen, B. S., Rowe, J. W., & Singer, B. H. (2001). Allostatic load as a marker of
5 cumulative biological risk: MacArthur studies of successful aging. *Proceedings of the*
6 *National Academy of Sciences*, 98(8), 4770–4775. <https://doi.org/10.1073/pnas.081072698>
- 7 Seshadri, D. R., Bittel, B., Browsey, D., Houghtaling, P., Drummond, C. K., Desai, M. Y., & Gillinov, A.
8 M. (2020). Accuracy of Apple Watch for Detection of Atrial Fibrillation. *Circulation*, 141(8),
9 702–703. <https://doi.org/10.1161/CIRCULATIONAHA.119.044126>
- 10 Shei, R.-J., Holder, I. G., Oumsang, A. S., Paris, B. A., & Paris, H. L. (2022). Wearable activity trackers—
11 advanced technology or advanced marketing? *European Journal of Applied Physiology*,
12 122(9), 1975–1990. <https://doi.org/10.1007/s00421-022-04951-1>
- 13 Siboni, S., Shabtai, A., Tippenhauer, N. O., Lee, J., & Elovici, Y. (2016). Advanced Security Testbed
14 Framework for Wearable IoT Devices. *ACM Transactions on Internet Technology*, 16(4), 1–
15 25. <https://doi.org/10.1145/2981546>
- 16 Steptoe, A., Cropley, M., & Joeke, K. (2000). Task demands and the pressures of everyday life:
17 Associations between cardiovascular reactivity and work blood pressure and heart rate.
18 *Health Psychology*, 19(1), 46–54. <https://doi.org/10.1037/0278-6133.19.1.46>
- 19 Stuyck, H., Dalla Costa, L., Cleeremans, A., & Van den Bussche, E. (2022). Validity of the Empatica E4
20 wristband to estimate resting-state heart rate variability in a lab-based context.
21 *International Journal of Psychophysiology*, 182, 105–118.
22 <https://doi.org/10.1016/j.ijpsycho.2022.10.003>
- 23 Tomaka, J., Blascovich, J., Kelsey, R. M., & Leitten, C. L. (1993). Subjective, physiological, and
24 behavioral effects of threat and challenge appraisal. *Journal of Personality and Social*
25 *Psychology*, 65(2), 248–260. <https://doi.org/10.1037/0022-3514.65.2.248>

- Treadwell, M. J., Alkon, A., Quirolo, K. C., & Boyce, W. T. (2010). Stress reactivity as a moderator of family stress, physical and mental health, and functional impairment for children with sickle cell disease. *Journal of Developmental and Behavioral Pediatrics: JDBP*, 31(6), 491–497.
<https://doi.org/10.1097/DBP.0b013e3181e2830f>
- Van De Schoot, R., De Bruin, J., Schram, R., Zahedi, P., De Boer, J., Weijdemans, F., Kramer, B., Huijts, M., Hoogerwerf, M., Ferdinands, G., Harkema, A., Willemsen, J., Ma, Y., Fang, Q., Hindriks, S., Tummers, L., & Oberski, D. L. (2021). An open source machine learning framework for efficient and transparent systematic reviews. *Nature Machine Intelligence*, 3(2), 125–133.
<https://doi.org/10.1038/s42256-020-00287-7>
- Van Der Mee, D. J., Gevonden, M. J., Westerink, J. H. D. M., & De Geus, E. J. C. (2023). Cardiorespiratory fitness, regular physical activity, and autonomic nervous system reactivity to laboratory and daily life stress. *Psychophysiology*, 60(4), e14212.
<https://doi.org/10.1111/psyp.14212>
- Van Lier, H. G., Pieterse, M. E., Garde, A., Postel, M. G., De Haan, H. A., Vollenbroek-Hutten, M. M. R., Schraagen, J. M., & Noordzij, M. L. (2020). A standardized validity assessment protocol for physiological signals from wearable technology: Methodological underpinnings and an application to the E4 biosensor. *Behavior Research Methods*, 52(2), 607–629.
<https://doi.org/10.3758/s13428-019-01263-9>
- Verkuil, B., Brosschot, J. F., Tollenaar, M. S., Lane, R. D., & Thayer, J. F. (2016). Prolonged Non-metabolic Heart Rate Variability Reduction as a Physiological Marker of Psychological Stress in Daily Life. *Annals of Behavioral Medicine*, 50(5), 704–714.
<https://doi.org/10.1007/s12160-016-9795-7>
- Vijayan, V., Connolly, J. P., Condell, J., McKelvey, N., & Gardiner, P. (2021). Review of Wearable Devices and Data Collection Considerations for Connected Health. *Sensors*, 21(16), 5589.
<https://doi.org/10.3390/s21165589>

- 1 Von Haaren, B., Ottenbacher, J., Muenz, J., Neumann, R., Boes, K., & Ebner-Priemer, U. (2016). Does
 2 a 20-week aerobic exercise training programme increase our capabilities to buffer real-life
 3 stressors? A randomized, controlled trial using ambulatory assessment. *European Journal of*
 4 *Applied Physiology*, 116(2), 383–394. <https://doi.org/10.1007/s00421-015-3284-8>
- 5 Vrijkotte, T. G., van Doornen, L. J., & de Geus, E. J. (2000). Effects of work stress on ambulatory blood
 6 pressure, heart rate, and heart rate variability. *Hypertension (Dallas, Tex.: 1979)*, 35(4), 880–
 7 886. <https://doi.org/10.1161/01.hyp.35.4.880>
- 8 Ward, A. M., Takahashi, O., Stevens, R., & Heneghan, C. (2012). Home measurement of blood
 9 pressure and cardiovascular disease: Systematic review and meta-analysis of prospective
 10 studies. *Journal of Hypertension*, 30(3), 449–456.
 11 <https://doi.org/10.1097/HJH.0b013e32834e4aed>
- 12 Wilhelm, F. H., Roth, W. T., & Sackner, M. A. (2003). The LifeShirt: An Advanced System for
 13 Ambulatory Measurement of Respiratory and Cardiac Function. *Behavior Modification*,
 14 27(5), 671–691. <https://doi.org/10.1177/0145445503256321>
- 15 Wormwood, J. B., Khan, Z., Siegel, E., Lynn, S. K., Dy, J., Feldman Barrett, L., & Quigley, K. S. (2019).
 16 Physiological Indices of Challenge and Threat: A Data-Driven Investigation of Autonomic
 17 Nervous System Reactivity During an Active Coping Stressor Task. *Psychophysiology*, 56(12),
 18 e13454. <https://doi.org/10.1111/psyp.13454>