# Recommendations for determining the validity of consumer wearable heart rate devices: expert statement and checklist of the INTERLIVE Network

Jan M Mühlen [1], Julie Stang,[2] Esben Lykke Skovgaard,[3] Pedro B Judice [4,5]
Pablo Molina-Garcia [6] William Johnston [7,8] Luís B Sardinha [9]
Francisco B Ortega [6,10] Brian Caulfield [7,8] Wilhelm Bloch,[1] Sulin Cheng,[1,11]
Ulf Ekelund [2] Jan Christian Brønd [3] Anders Grøntved,[3] Moritz Schumann [1,11]

**OPEN ACCESS**

## ABSTRACT

Assessing vital signs such as heart rate (HR) by wearable devices in a lifestyle-related environment provides widespread opportunities for public health related research and applications. Commonly, consumer wearable devices assessing HR are based on photoplethysmography (PPG), where HR is determined by absorption and reflection of emitted light by the blood. However, methodological differences and shortcomings in the validation process hamper the comparability of the validity of various wearable devices assessing HR. Towards Intelligent Health and Well-Being: Network of Physical Activity Assessment (INTERLIVE) is a joint European initiative of six universities and one industrial partner. The consortium was founded in 2019 and strives towards developing best-practice recommendations for evaluating the validity of consumer wearables and smartphones. This expert statement presents a best-practice validation protocol for consumer wearables assessing HR by PPG. The recommendations were developed through the following multi-stage process: (1) a systematic literature review based on the Preferred Reporting Items for Systematic Reviews and Meta-Analyses, (2) an unstructured review of the wider literature pertaining to factors that may introduce bias during the validation of these devices and (3) evidence-informed expert opinions of the INTERLIVE Network. A total of 44 articles were deemed eligible and retrieved through our systematic literature review. Based on these studies, a wider literature review and our evidence-informed expert opinions, we propose a validation framework with standardised recommendations using six domains: considerations for the target population, criterion measure, index measure, testing conditions, data processing and the statistical analysis. As such, this paper presents recommendations to standardise the validity testing and reporting of PPG-based HR wearables used by consumers. Moreover, checklists are provided to guide the validation protocol development and reporting. This will ensure that manufacturers, consumers, healthcare providers and researchers use wearables safely and to its full potential.

## INTRODUCTION

Heart rate (HR) is defined as the number of heart beats per minute (bpm) and can be determined from the time interval between two successive cardiac cycles initiated by action potentials in the sinoatrial node.[1] While resting HR is a key vital sign and a well-established predictor of all-cause and cardiovascular mortality in the general population,[2] other features of HR such as the response to exercise and HR variability (HRV) are indicators of general health status, including fitness as well as both physiological and mental stress.[3–5] Furthermore, HR assessment during exercise training is an important tool for monitoring training load in elite athletes and recreational exercisers.[6 7]

Traditionally, HR is derived from electrocardiography (ECG) recordings through either multiple-lead channels or simple chest-straps, consisting of two electrodes. Thus, HR assessment has traditionally been limited to medical conditions, laboratory testing or training monitoring and was not suitable for long-term assessment during daily living. However, recently a wealth of wearables that assess HR by photoplethysmography (PPG) have entered the consumer market. This allows not only for continuous fitness monitoring, but also facilitates screening for incident disease and continuous monitoring of disease progression and complications (eg, detection of atrial fibrillation and stroke prevention, coronary artery disease or sleep apnoea).[8–13]

PPG is an optical technique that is based on the absorption and reflection of emitted light by the blood, where the systolic variations in blood volume modulate the amount of transmitted or reflected light.[14] However, considerable differences in the validity of HR assessed by PPG-based devices are observed,[15] which are likely related to difficulties in mathematical peak detection and a higher sensitivity to motion artefacts.[16] This, in turn, may have severe consequences for long-term adherence to regular exercise,[17] but also for risk stratification if the device is used in a clinical setting.[18]

Unfortunately, the validation quality of wearables remains often unknown to the consumer due to non-transparent standards for testing and reporting. The validity assessment of consumer wearables is most optimally performed by independent institutions, but the number of new devices introduced by a continuously rising number of device manufacturers makes it almost impossible for scientific institutions to keep up with recent developments. Moreover, the discontinuation of a device or implementation of important changes to a device firmware/software might invalidate previous

work.[19] Therefore, it is important to develop a common framework for the optimal validity evaluation of consumer wearables measuring HR by PPG, to be used by both manufacturers and research institutions in order to provide quality assurance of available devices.

In 2018, the Consumer Technology Association published a preliminary framework for evaluating and reporting the validity for measuring HR with consumer wearables, including considerations for testing protocols but also individual characteristics, such as skin tone, body mass index (BMI), sex and age.[20] However, recommendations for long-term monitoring of HR during free-living conditions are lacking in these guidelines and the scientific evidence for the suggested guidelines has not been presented. In addition, in a recently published review article factors that may affect the accuracy of wrist-worn HR wearables were critically discussed and initial considerations for performing validity testing of these devices provided.[21] However, the published work mainly targets scientific evaluations of these devices and specific guidelines that allow for an immediate transfer into practice have not been presented.

Therefore, the present expert statement aims to expand on previously published work by proposing a set of guidelines targeting both manufacturers and scientific institutions, to ensure the rigorous and transparent validation and accuracy reporting of PPG-based consumer wearable HR devices, while at the same time being feasible to carry out. Furthermore, the statement aims to propose a best-practice framework of rigorousness in evaluating criterion validity and provide recommendations for future development of evaluating the validity of wearable HR monitors used by consumers. The work presented is based on a systematic literature search as well as an unstructured review of the wider literature pertaining to factors that may introduce bias during the validation of these devices and evidence informed expert opinions of the INTElligent Health and Well-being: NetwoRk of PhysicaL ActIVity AssEssment (INTERLIVE). As a result, we provide a comprehensive summary of variables that require consideration when developing evaluation protocols (online supplemental table 1) and suggest practical checklists for validation protocol designing (table 1) and transparent data reporting (table 2).

## EXPERT STATEMENT PROCESS
### The INTERLIVE Network
INTERLIVE is a joint initiative of the University of Lisbon (Portugal), the German Sport University (Germany), University of Southern Denmark (Denmark), Norwegian School of Sport Sciences (Norway), University College Dublin (Ireland), University of Granada (Spain) and Huawei Technologies Finland. The consortium was founded in 2019 and strives towards developing best-practice protocols for evaluating the validity of consumer wearables. Moreover, we are aiming to increase awareness of the advantages and limitations of different validation protocols and to introduce novel health-related metrics, fostering a wide-spread use of physical activity indicators. As one of the initial key aims of the group, the consortium aimed to develop best-practice validation protocols for consumer wearable HR monitoring (part A) and wearable and smartphone devices for step-counting (part B, presented in a separate publication).

**Table 1** Checklist of items that need to be considered when planning validity protocols for consumer heart rate wearables

| | |
|---|---|
| **Target population** | |
| BMI | O |
| Body height | O |
| Skin tone | O |
| Sex | O |
| Sample size calculation via pilot study | O |
| **Criterion measure** | |
| Chest strap or ECG | O |
| Placement according to manufacturer's instructions | O |
| **Index device** | |
| Placement according to manufacturer's instructions | O |
| **Pretest preparations** | |
| Standardised meal replacement | O |
| Control caffeine intake | O |
| Medical screening | O |
| Exclude participants with medication affecting cardiovascular function | O |
| Control for previous intense physical activity | O |
| **Testing: laboratory conditions** | |
| Minimum of 3 walking intensities | O |
| Minimum of 2 running intensities | O |
| Minimum of 3 biking intensities | O |
| Steady-state (2–5 min) | O |
| HR kinetics (transitions and recovery) | O |
| Validity level | |
| 1. Graded ergometer test with a wide range of exercise intensities reported as % of $HR_{max}$ (or $VO_{2max}$) including rest and recovery | O |
| 2. Graded ergometer test with a wide range of exercise intensities in absolute values (ie, speed/incline, W/rpm) including rest and recovery | O |
| 3. Graded ergometer test with a moderate range of exercise intensities as % of $HR_{max}$ (or $VO_{2max}$) including rest and recovery | O |
| 4. Graded ergometer test with a moderate range of exercise intensities reported in absolute values (ie, speed/incline, W/rpm) including rest and recovery | O |
| 5. Graded ergometer test with a low range of exercise intensities reported as % of $HR_{max}$ (or $VO_{2max}$) including rest and recovery | O |
| 6. Graded ergometer test with a low range of exercise intensities reported in absolute values (ie, speed/incline, W/rpm) including rest and recovery | O |
| **Testing: semifree-living (sport-specific) conditions** | |
| Intermittent activities (ie, soccer, basketball, etc) | |
| 1. Inherent environmental conditions (eg, standard playing field, etc) | O |
| 2. Inherent no of players included | O |
| 3. Inherent duration with a minimum of 15–20 min | O |
| Continuous activities (running, walking, biking, swimming) | |
| 1. Minimum of three intensities (40 %, 60 %, 80% of $HR_{max}$) | O |
| 2. Inherent duration with a minimum of 2 min of each intensity | O |
| Activities with domestic behaviour (doing laundry, gardening, home construction) | |
| 1. Minimum of 15–20 min | O |
| **Testing: free-living conditions** | |
| Subject's wear index and criterion device for a minimum of 24 hours | O |
| Exclude subject's not presenting HR data above 40% of $HR_{max}$ | O |
| Exclude recordings missing more than 5% of the data (index or criterion device) | O |
| **Processing** | |
| Criterion measure processing | |
| 1. Apply an automated method for filtering ectopic beats and motion artefacts | O |
| Index measure processing | |
| 1. No post processing of the end-user data is allowed | O |
| | Continued |

#### Table 1   Continued

| | |
|---|---|
| 2. Resampling into a window of 5 s is allowed | O |
| Epochs for analysis/window size | |
| 1. Sample criterion measure with same epoch as available with the index measure | O |
| 2. Window size should be 5 s or shorter | O |
| Index and criterion synchronisation | |
| 1. Automated method for synchronisation (cross correlation or similar) | O |
| **Statistical analysis** | |
| Standard Bland-Altman LoA analysis for steady-state conditions | O |
| Repeated measure LoA analysis for non-steady state conditions (multiple paired observations of HR epochs per individual) | O |
| Evaluate within-device precision by comparing the within-person variability in average HR over 5 s windows separately for steady-state activity of at least 2 min duration | O |

BMI, body mass index; HR, heart rate; $HR_{max}$, maximal heart rate; LoA, limits of agreement; rpm, repetitions per minute; $VO2_{max}$, maximal oxygen uptake; W, Watts.

## Expert validation protocol development

### Expert validation process

An initial meeting was held in Cascais, Portugal on 15 November 2019. At this meeting, it was agreed that the optimal process for developing the best-practice validation protocol should begin with extracting key elements of the validation protocols previously used in the scientific literature. This information was then used as the foundation for discussions on the optimal and feasible protocol for conducting the validity assessment that describes the accuracy end-users can expect if the wearable is used in the designated or similar setting. The consortium formed two working groups: (1) HR monitoring (JMM, ELS, JS, SC, WB, JCB, UE, AG and MS), (2) step-counting (WJ, PMG, PBJ, BC, FBO and LBS). The working groups subsequently defined multiple systematic literature search strategies, prior to sharing them with the wider consortium. A second consortium meeting was held virtually on 10 March 2020 to finalise the search strategies, including the selection of the minimum a priori required criterion measure(s). Thereafter, the systematic search was performed and a framework was developed for extracting data of the validation process, including data on target population, criterion and index device, testing conditions, data processing and statistical analysis. In parallel, an unstructured review of the wider literature was conducted to include valid studies on factors that may affect the accuracy on consumer wearables not identified by our defined search strategies. Following that, the data extraction was performed and multiple workgroup meetings were held to discuss each aspect of the validation protocols used in the individual studies. Based on the data synthesised during the systematic literature review, the a priori knowledge relating to research grade device validation[22–25] and the evidence informed expert opinion of the INTERLIVE members, a set of key domains for the best-practice recommendations were proposed. The synthesised data were then reviewed with respect to these domains, and expert validation protocols for wearable HR monitors (part A) and wearable and smartphone step-counters (part B) and were iteratively developed by the working groups and subsequently shared with the entire consortium. At a virtual meeting held on 17 June 2020, the revised drafts were discussed and the two protocols were aligned to ensure harmonisation of the statements. The revised drafts were then edited for consistency and reviewed by the wider consortium prior to circulation for final approval.

#### Table 2   Minimum required reporting sheet for standardized and transparent data sharing

| | Description | Reporting |
|---|---|---|
| **Target population** | | |
| Sampling method | Random, convenient, and so on | |
| Distribution of sex | ♂=n/♀=n | |
| Body height | Mean±SD and range (cm) | |
| BMI | Mean±SD and range (kg/m²) | |
| Skin tone | Distribution of Fitzpatrick scale | |
| Sample size | Number of subjects | |
| **Criterion measure** | | |
| Chest strap or ECG (RR intervals) | Model and brand; chest strap: agreement with respect to bpm | |
| Placement | Manufacturer's instructions and actual placement | |
| **Index device** | | |
| Placement | Manufacturer's instructions and actual placement | |
| **Pre-test preparation** | | |
| Standardised meal replacement | Type of replacement and duration of control (hours prior to testing) | |
| Caffeine intake | Duration of control (hours prior to testing) | |
| Medical screening | Type of medical screening | |
| Exclusion of participants | Exclusions due to specific medication affecting cardiovascular function | |
| Intense physical activity | Duration of control (hours prior to testing) | |
| **Testing protocol** | | |
| Type of protocol | Laboratory, semi-free living/sport-specific, free-living | |
| Contextual factors | Indoors, outdoors | |
| Type of activity | Cycling, treadmill walking/running, swimming, other sports/activities | |
| Duration | Minutes, hours | |
| Exercise intensity | Relative to aerobic capacity (%$HR_{max}$, $VO_{2max}$, RM) Or Absolute values (ie, speed/ incline, W/ rpm) | |
| Steady-state (2–5 min) | Mean HR (ie, 5–30 s intervals) | |
| HR kinetics (transitions and recovery) | ΔHR | |
| **Processing** | | |
| Criterion measure processing | Method used for error correction and data smoothing | |
| Index measure processing | Method used for resampling (is used) | |
| Epochs for analysis/window size | In seconds | |
| Index and criterion synchronisation | Method used (cross-correlations or similar methods) | |
| **Statistical analysis** (report separately for each exercise intensity and/or activity) | | |
| N of paired observations | Paired HR (amount) | |
| Data availability | Data availability (%) | |
| Index device, mean HR | Mean±SD | |
| Criterion device, mean HR | Mean±SD | |
| Mean difference | Mean±SD and SE | |
| Mean absolute error | Bpm | |
| MAPE | % | |
| Standard LoA | Mean difference or mean relative difference and LoA including 95% CIs (separately for each steady-state intensity and/or activity) | |
| Repeated measure LoA analysis | Mean difference or mean relative difference and LoA including 95% CIs (separately for each non-steady-state activities) | |

**Table 2**   Continued

| | Description | Reporting |
|---|---|---|
| Within–device precision for steady-state activities | The 95% prediction interval and ICC (report separately for each steady-state intensity and/or activity) | |
| **Other** | | |
| Deviations in the validation process | | |

BMI, body mass index; bpm, beats per minute; ECG, electrocardiogram; HR, heart rate; HRmax, maximal heart rate; ICC, intra class correlation coefficient; LoA, limits of agreement; MAPE, mean absolute percentage error; RM, repetition maximum; rpm, repetitions per minute; SE, standard error; VO2max, maximal oxygen uptake; W, Watt.

### Systematic review process

The primary aim of our initial systematic literature review was to determine which methods and protocols are currently used in the scientific literature to validate HR with consumer-based wearables. Importantly, we did not aim to review the results from studies examining the validity of wearable consumer devices to assess HR. The search was conducted with respect to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) and registered with the international database of prospectively registered systematic reviews in health and social care (PROSPERO ID: CRD42020177667). Three-domain search terms were used to identify journal articles published in the electronic databases PubMed, Embase and Web of Science. More specifically, these search terms were defined as the control device, the outcome as well as the study design (online supplemental table 2).

Only English language publications in human populations with no restriction to publication year were included. Relevant articles had to be published prior to 18 March 2020. The inclusion criteria were defined as Population, Intervention, Comparison, Outcome and Study design .[26] No restrictions were made with regards to population (ie, healthy, patients, children, etc) and interventions (ie, protocols used). Protocols were classified as (1) laboratory settings (ie, well-controlled conditions, including isolated tasks such as walking, running or cycling on a treadmill or a stationary cycle ergometer), (2) semifree-living settings (semi-controlled conditions, including 'simulated' activities of daily living for the purpose of replicating 'free-living' conditions) or (3) free-living settings (long-term monitoring of daily living without restrictions of the completed tasks). As a comparison, a criterion measure using a gold standard (ie, assessment of the time elapsed between two successive R-waves [RR intervals] of the signal of sequence of the Q, R and S complex [QRS]) was required. Furthermore, only studies that assessed HR by a PPG-based consumer wearable as the primary outcome measure were included. However, no restrictions applied to the human-wearable interface of the index devices (eg, light wavelength or measurement site). The detailed search string can be found in online supplemental table 2.

Screening and data extraction were performed independently by three members of the consortium, using Covidence software (Veritas Health Innovation). The search process entailed saving the online search, removing duplicates as well as consequently screening titles, abstracts and eligible full texts. A minimum of two identical votes was required for eligibility judgement. In case of a lack of consensus, the third member of the team was consulted. Data extraction was performed according to specific criteria that are outlined in detail in online supplemental tables 3–6.

### CURRENT STATE OF KNOWLEDGE

The presented current state of knowledge is based on the studies that were identified by the systematic literature search as well as supplemental technological studies and our evidence informed expert opinions. Our systematic review led to a total of 1894 hits. Automatically removing duplicates and ineligible records led to 108 full texts for further assessment. Overall, 66 studies were primarily excluded for methodological reasons (ie, in terms of outcome, study design and comparator). Finally, a total of 43 articles were deemed eligible and retrieved. Additionally, one study was manually added by screening other resources, leading to a total number of 44 studies remaining for data extraction. The PRISMA flow chart of the systematic review process and the reasons for exclusions are presented in the supplements (online supplemental figure 1). The following section provides a short summary on the key considerations that appear to be important when testing the validity of PPG-based consumer wearables. Data gathered are presented in six key domains (figure 1), that were deemed relevant for validity testing (target population, criterion measure, device placement, testing conditions, data processing and statistical analysis). The most important aspects of validation protocols are also summarised in the supplements, table 1. The consortium acknowledges that the presented list of domains reflects the current state of knowledge but may not be considered exhaustive.

### Target population

Selecting the target population for the validity assessment appears to be a key factor that determines the significance of the findings obtained. Although PPG-based wearables could theoretically be validated in numerous populations that differ considerably in demographics, ethnicity, anthropometrics and activity level, we advise that the evaluation reflects the device performance in the hands of the intended user. However, even by assessing the validity in a sample that is homogeneous in one domain (eg, recreationally active young men), it is likely that other domains may not be controlled for simultaneously (eg, skin tone). Therefore, we suggest that the target populations generally reflect a heterogeneous sample, allowing for possible subgroup analysis. Homogeneous samples, on the other hand, may be included if the intention is to test the validity of the wearable for a very specific group (eg, athletes of a specific sport).

In addition to the aforementioned considerations, other factors may require attention. For example, the pathology of some heart-related diseases may affect the outline of the QRS complex and potentially provide poor identification of the R wave.[27] However, the number of heart-related conditions and their potential implications on the QRS complex impose numerous challenges with the validity assessment. These challenges of including patients with heart-related conditions must be appropriately addressed to ensure the accuracy of the HR measurements. In this context, assessment of atrial fibrillation has been targeted by some wearable devices[28] and it seems possible that the future will bring more devices that can address specific heart-related conditions that can be detected with a high degree of confidence.[29] In addition, abnormalities in blood pressure may affect the PPG signal.[30] Consequently, in many studies included in our review, participants with known systolic or diastolic blood pressure abnormalities were excluded[31 32] or at least reported.[33–38]

Other considerations concern the use of medication and dietary supplements that may affect HR recordings and should be considered when designing validation protocols. Interestingly,
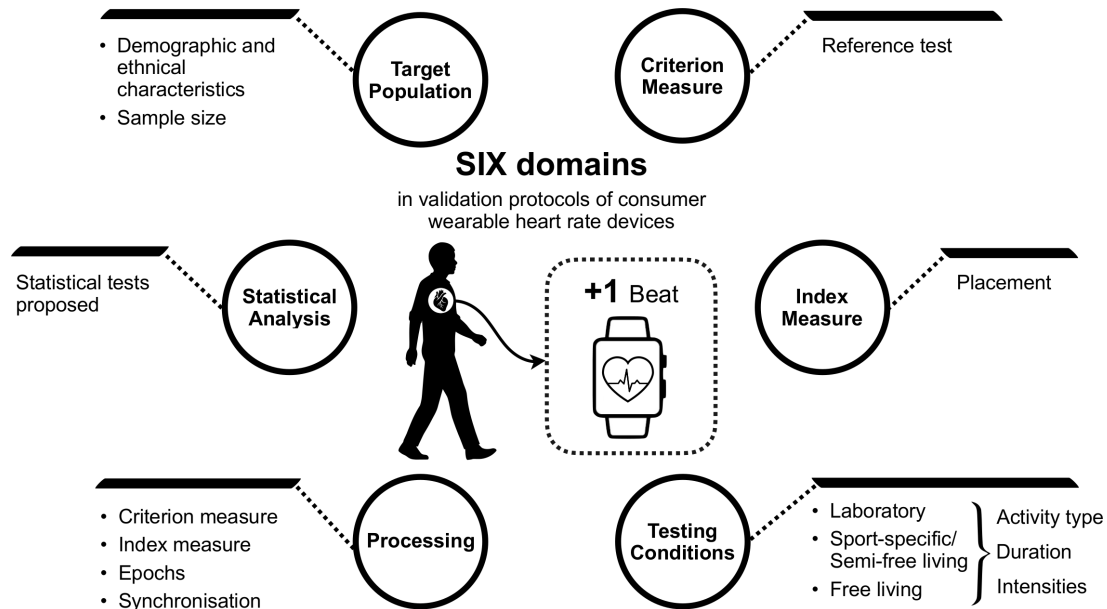
**Figure 1** The six domains identified as important factors to be considered during validity testing of wearable devices assessing heart rate by photoplethysmography.

none of the identified studies assessed the accuracy of wearable HR monitors in patients with cardiovascular conditions. These patients present a variety of potential challenges to monitor's accuracy, including hypertension, peripheral arterial disease, venous insufficiency, obesity, atrial fibrillation and use of medications that affect HR, vascular tone and volume status (eg, beta-blockers, ACE inhibitors, calcium channel blockers, and diuretics).[33] Furthermore, factors such as inked and damaged skin (eg, tattoos, scars etc.) may potentially affect the PPG signal and, thus, participants exhibiting either of these were excluded in few studies identified by our systematic review.[33 39–41] While we also recommend that exclusion will likely help to overcome potential errors originating from these factors, in this statement we have focused on the validity assessment of HR measurements in the general population. Principally, healthy samples are recommended for the general device validation. However, if a device is specifically designed for a special population, this needs to be reflected in the target population.

The following factors require consideration when designing an appropriate validation protocol.

### Sample size considerations
The sample size should be defined a priori. If an a priori specified level of 'in agreement' (ie, the difference of paired measurements falls within a specified interval) is considered, the sample size should be calculated based on an expected mean absolute difference, the expected standard deviation (SD) of the differences, and a predefined clinical maximum allowed difference needed to obtain a power of 80% or 90% to assess agreement between two methods of measurement with a sufficient precision.[42] It is advised to conduct a pilot study to obtain the mean and SD of differences between the wearable consumer device and the criterion measure to make these prior sample size calculations. If no a priori specified level of 'in agreement' is considered, for homogeneous samples we recommend a minimum of 45 participants as a rule of thumb.[43] This number is also in line with the average number of participants included in the studies identified by our systematic review. In any case, the variability

of relevant participant characteristics in the sample should be considered and for heterogeneous groups, a larger sample size might be necessary.

### Age
Ageing has previously been associated with increases in arterial stiffness, resulting in changes in the propagation of the pulse to the periphery, thereby affecting pulse timing and shape characteristics.[44] However, only three studies identified by our systematic literature review performed a statistical analysis for age and device error (total range 21–73 years), but did not find age to affect the error in the prediction of HR measurements.[41 45 46] In fact, this finding was confirmed by a very recent study validating the wearable fitness trackers Xiaomi Mi Band 2 and Garmin Vivosmart HR+.[47] Also, in this study, similar mean percentage errors for young (20–26 years) participants and seniors (>65 years) were observed. Thus, deteriorating effects on vascular function with increasing age may not be reflected in HR assessed by PPG but more research is needed to clearly assess these effects. However, if the validity of a wearable device is not needed for one specific target group (eg, children) we suggest testing in a heterogeneous sample.

### Sex
Sex is associated with cardiovascular function, affecting resting HR and arterial blood pressure.[48] Consequently, sex differences might also be reflected in PPG-based HR due to possible differences in device positioning and skin characteristics.[45] For example, differences seem to exist in the thickness and echo intensity of skin between males and females.[49] However, three studies identified by our systematic literature search did not find HR validity assessed via PPG to be affected by sex,[32 39 41] while others[45 46] found larger measurement errors in men as compared with women. A recent article clearly indicated that factors such as pulse arrival time, pulse transit time, systolic pulse transit time and the ratio of areas under the PPG waveform are affected by sex, with men showing a larger effect on the PPG signal.[48] Considering these findings, it appears that sex likely affects

wearable device validity and needs to be accounted for when evaluating PPG-based devices.

## Body height

Body height was previously identified as a contributor to larger pulse transit time due to longer distances between the heart and the periphery (ie, the measurement site). Indeed, the latency time between the QRS complex and the PPG peak was found to be larger in taller subjects and was significantly associated with changes in pulse transit time both on the fingers and toes but not on the ears.[44] However, in this study, no associations were found between body height and HR and these findings correspond well with a separate regression analysis,[46] showing body height not to interact with PPG-based HR validity (range: men 159.1–190.0 cm, women 154.4–184.2 cm). In fact, the latter study was the only group that accounted for body height when analysing their findings. However, whether possible delays in pulse transit time were accounted for in the algorithm of the devices tested in the other studies identified by our systematic search remains unknown due to restricted access to algorithms. Thus, we recommend that body height should be considered as a possible factor that may affect HR assessed by PPG. This might be accounted for by including heterogeneous samples of participants (eg, children, adolescents and adults).

## Body mass index

In two studies identified by our systematic review, possible associations between BMI and measurement error were assessed but it was concluded that BMI does not affect PPG-based HR accuracy.[50] However, both studies used a rather homogeneous sample (ie, BMI of 20–27 kg/m² and 19–33 kg/m², respectively). In contrast, in two studies using larger ranges for BMI (ie, 17.2–39.3 kg/m² and 17.1–45.0 kg/m², respectively) a higher BMI was statistically associated with larger error rates across multiple devices.[40 46] Moreover, a study also found that BMI was correlated with wrist circumference, which may in turn affect the PPG signal.[36] Thus, previous findings provide potential indications for BMI to affect HR measurements assessed by PPG.

## Skin tone

The interaction of light with biological tissue can be quite complex and may involve scattering, absorption and/or reflection.[51] For example, it was previously shown that darker skin pigmentation may attenuate the permeability of light wavelengths shorter than 650 nm.[37] The importance of skin tone is underlined by our systematic literature review, showing that skin tone or at least ethnicity was considered as a confounder in 20 studies.[33 35–37 39–41 46 50 52–62] Among these studies, a higher device error was found in participants across several devices or types of activities with darker skin tones assessed by the Fitzpatrick scale,[37 39 46 56 58] while in other studies this was not observed.[33 50 63] Thus, skin tone appears to affect the accuracy of HR readings based on PPG and should also be considered during validity testing.

## Criterion measure

The current gold-standard reference method for assessing HR is ECG with 12-lead, which is standard in clinical practice when a full ECG tracing of the cardiac cycle is desired.[64] The largest and most distinct feature in the QRS complex is the R wave, that represents the early ventricular depolarisation and is commonly used for identifying single cardiac cycles. ECG measurements are conducted using dry or wet surface Ag/AgCl electrodes.

Wet electrodes include a conductive gel used to decrease the electrode-skin impedance and, thus, increase the signal to noise ratio. However, the conductive gel tends to dry out with time, which potentially affects the data quality.[65] Dry electrodes are an alternative to wet electrodes although not commonly applied with long duration ECG measurements.

Chest strap HR monitors are electronically similar to dry electrodes and some commercially available devices provide beat to beat (RR) intervals for HRV analysis.[66] Chest strap HR monitors are specifically designed to be used with sports participation at various intensities. However, measuring RR intervals during strenuous exercise activities is challenging due to the substantial movement of the torso and occasionally high force impacts which generate motion artefacts in the ECG signal.[67] The validity of estimating RR intervals with commercially available chest strap devices has been investigated in various studies[68–72] and several devices provide RR intervals that demonstrate good to perfect agreement with ECG both in resting conditions and exercise. In online supplemental table 7, we have summarised several validated chest strap devices for measuring RR intervals. The HRV task force guidelines (and the update from 2015) suggest that an independent evaluation of commercially available devices is needed to ensure the validity of the RR interval with HRV analysis.[73 74] However, a required level of agreement for a device to be valid is not specified in these guidelines. Consequently, we suggest that any commercial device (chest strap or ECG measured using either dry or wet electrodes) providing RR intervals, which has been independently validated and demonstrates an excellent agreement with respect to bpm (ie, >95%, see online supplemental table 7), can be used as an appropriate criterion measure for evaluating wearable technologies providing HR to the end-users.

## Index measure

In addition to the target population, potential sources of bias originating from the placement of the index device need to be considered. We recommend wearing the index device according to manufacturer's instructions, which should result in standardisation. Nevertheless, the following sections provide a short overview on the most robust factors related to the device placement that may affect the validity of PPG-based HR readings.

## Motion artefacts

Previous studies have indicated that consumer wearables are reasonably accurate at resting and moderate steady-state intensities, while the accuracy is typically lower in activities inducing fluctuations in HR.[75] For example, the study by Müller *et al* showed relative higher errors in two activity trackers in a free-living condition compared with a laboratory-based cycling protocol.[62] It is likely that these differences in accuracy are attributed to motion artefacts, which are typically caused by displacement of the PPG sensor over the skin, changes in skin deformation, blood flow dynamics and/or ambient temperature.[76 77] This, in turn, may well manifest as missing or false beats, resulting in invalid HR calculations.[78–80] Even though it is likely that motion artefacts are apparent in every dynamic protocol, only five studies identified by our systematic search specifically reported signal noise originating from movement[37 41 58 81 82] (online supplemental table 6). Thus, protocols used for validity testing of wearables are recommended to include heterogeneous activities or in case the device is intended to be used in a sport or activity-specific setting, conditions similar to the intended setting (ie, providing a common level of movement) should be tested.

## Contact pressure

It was previously shown that the waveform of the PPG signal may be affected by the contact force between the sensor and the measurement site and that the waveform of the obtained PPG signal differs depending on the PPG probe contact.[51] The authors further stated that the most accurate PPG signal may be obtained under conditions of transmural pressure, defined as the pressure difference between the inside and outside of blood vessels (ie, the pressure across the wall of the blood vessel). Interestingly, none of the studies identified by our systematic search reported that contact pressure was measured. Future studies should assess whether the validity of HR readings indeed differs between different contact pressures and whether this is related to wearing comfort (ie, ecological validity). Thus, as for now it is recommended to wear the device according to manufacturer instructions during validity testing and to ensure a constant contact pressure, especially in the context of long-term HR monitoring (ie, by repositioning the device periodically).

## Ambient light

Light sensitive diodes may also be affected by ambient light. While this has been discussed in few studies identified by our systematic review,[33 37 41 81 83] the magnitude of this effect remains unknown at this stage. In this context, light interferences may be reduced by shading of the interface area site and by electronic filtering (eg, light modulation filtering).[84] Consequently, future studies should address ambient light as a potential source of bias in PPG measurements. Irrespective of this, we believe that potential irritations caused by the ambient conditions may be minimised by correct positioning of the device, as was previously also stated in a topical review.[84]

## Ambient temperature

PPG signal quality may also be influenced by the temperature originating both from the environment and changes in skin temperature. While eight studies included in our systematic review reported a controlled laboratory temperature,[34 37 50 53 62 82 85 86] in one study the underestimation of the index device was partially explained by low ambient temperatures of the laboratory assessment (18°C–20°C) compared with that of free-living conditions (30°C–32°C).[62] In addition, it was shown that the error in HR readings obtained from infrared but not green light appeared to be higher in cold (10°C) compared with hot (45°C) conditions.[87] Thus, it seems plausible that ambient temperature may affect the PPG signal quality and should be standardised during laboratory validations and considered as a potential source of bias in free-living conditions.

## Testing conditions: laboratory, semifree-living and free-living

Factors that affect the choice of protocol for examining the validity of wearable HR devices, and the validity of the device itself, include types of activities, the intensity of these activities and for how long and frequent these activities are performed. In general, agreement of a device compared with a criterion method during a specific type of activity with a specific intensity is only valid for these conditions. Thus, validity testing programmes of wearable devices may vary in length and complexity and should reflect the intended use of the device. Laboratory based protocols including steady-state activities of varying intensities may be appropriate when examining the basic validity of a device against a criterion, whereas free-living protocols are required when the device is intended to be used in everyday life including sleep.

Furthermore, we recommend to standardise the pretest preparation with a standardised meal replacement to avoid gastric complications during high exercise intensities. Moreover, caffeine should be avoided 12 hours and intense physical activity 48 hours prior to testing. In addition, we recommend a medical screening and to exclude participants using regular medication that affects cardiovascular function (eg, beta-blockers).

## Types of activities

Lab based protocols in the studies identified by our systematic search usually included treadmill locomotion[31 32 38–40 53 57 82 88 89] or a combination of treadmill locomotion and ergometer cycling.[33 41 46 52 90–92] In addition, few studies have included activities of daily living (eg, folding laundry and sweeping)[12 45 54 93 94] or resistance exercise[12 50 58 83 94 95] (see online supplemental table 4 for additional information).

HR data measured by consumer wearables were most accurate when assessing locomotor activities that are characterised by repetitive movements (eg, cycling, walking or running) in laboratory settings,[32 40 41 46 52 58] and were less accurate where the movements were inherently more complex, such as resistance exercise and activities of daily living.[50 58] For example, the accuracy of the wearable device was substantially higher during aerobic exercise (92%) as compared with resistance exercises (35%).[50] Similarly, HR measured by non-wrist worn devices (ie, worn in the ear) were relatively accurate during upper and lower body resistance exercises, whereas wrist-worn devices showed poor accuracy. Activities implying upper body movements induced a higher rate of errors and HR drop-outs than endurance exercises (ie, running, walking or cycling) in free-living conditions.[56] As this was not further assessed in the study, it was assumed that this imprecision was due to motion artefacts from the arm and chest movements, as reported during laboratory or semifree-living protocols.[33 41 50 58 95]

Upper body movements cause greater variability in error[41 50 58 91] for wrist-worn devices, probably caused by motion artefacts and variable contact between skin and device, due to variable pressure/contact induced by muscle contractions and changes in blood flow. During upper body work and work involving the arms, muscle and ligament tension in the wrist may interfere with HR detection from capillary blood flow.[50] Thus, devices that rely on HR detection through the skin may be inaccurate if speed or intensity is increased and during activities where skin contact is lost or if an isometric contraction is necessary to perform the activity. Therefore, the specific activities being examined must be clearly considered so the validity of the measurement device in question is aligned with the appropriate/actual use of the device. This is likely an inherited and significant limitation of PPG in general.

## Duration and repetitions

The duration of the laboratory protocols performed in the studies identified by our systematic review varied substantially from three to 80 min, with the longest duration observed in semifree-living protocols comprising multiple activities (online supplemental table 4). The length of free-living protocols varied between 2 and 24 hours of continuous monitoring (online supplemental table 4).

Assessing accuracy of HR measurements during steady-state exercise is relevant as consumers often use these devices to monitor HR during continuous exercise sessions or to monitor exercise load and energy expenditure. Steady-state is reached when the HR plateaus during continuous exercise at a submaximal

intensity level, and reflects the balance obtained when the cardiac output is sufficient to transport the oxygen needed to meet the energy cost of the work performed.[96] This usually occurs within the first 2 min of exercise, depending on the change in intensity and fitness level of the participant.[96] However, since HR tends to exhibit a lag compared with the external work performed or the corresponding oxygen cost, protocols should strive for a combination of steady-state activities and those with shorter duration and varying intensities. Indeed, some previous studies have reported lower accuracy when the activity is intermittent with swift changes in exercise intensity (ie, changes in speed of running) or changes in activity that cause changes in wrist movements for PPG wrist-worn devices.[50 58] Since PPG sensors estimate HR by measuring changes in blood flow, the limited blood flow to the wrist at the initiation of exercise might lower the confidence of the predictive algorithms to accurately measure HR (ie, measured heart beats are discarded until the algorithm is confident that it is recording a physiologically plausible value).[97] This must be considered during measurements of HR during activity with rapid changes in intensity and non-steady-state conditions (less than 3 min in duration).

### Exercise intensities

Accurate HR readings throughout a wide range of intensities from rest to near maximal is a prerequisite for any consumer device. The laboratory studies reviewed predominantly included multiple intensities ranging from light to very vigorous in their protocols, whereas the measure of intensity varied (eg, speed, watts, metabolic equivalents, % of maximal aerobic capacity) (online supplemental table 4). Semifree-living protocols also included various intensities (online supplemental table 4), whereas the intensity and variability in intensity during free-living is population specific (eg, athletes vs elderly) and cannot be controlled. The accuracy obtained during a free-living protocol also depends on the duration of the measurement and the variability in activities performed (see above). Relatively high measurement errors (10.1%) were observed in a study evaluating the accuracy of a wrist-worn device during a sedentary and light intensity semifree-living protocol,[93] which may be attributable to the freely selection of activities during the testing period. However, this format theoretically mimics everyday activity more effectively than traditional structured activities. It may, therefore, reflect more realistic estimates of validity than a laboratory protocol and may also provide new evidence of how effective the PPG technology is when used in consumer devices.

The studies identified by our systematic review clearly indicated that the accuracy of PPG devices is intensity dependent,[58 83 91 94] with increased accuracy during lower intensity exercise and at rest as compared with vigorous intensity exercise, such as running.[52 57 58 62 83 89 95 98] Conversely, the opposite has also been reported.[32 91] For example, Stahl et al observed the highest accuracy (3.06%) during the highest speed tested (9.6 km/hour on the treadmill). One possible explanation is that with increased intensity perfusion is improved, which could decrease the error rate. Consequently, exercise intensities clearly have a profound effect on accuracy of HR readings and should be considered when designing validity protocols in laboratory, semifree-living and free-living conditions.

### Processing of index and criterion data

Data processing and reporting is an integral part of validity testing and should be performed with caution. The following items provide recommendations that should be considered in terms of a best practice in the validation process.

### Index and criterion synchronisation

Evaluating the validity of consumer wearables measuring HR requires the comparison of two or more time series, which require an equal sampling interval and accurate temporal alignment. The sampling interval of the criterion measure and the wearable devices is most likely not the same and this can be addressed by either interpolation or simple resampling (averaging) of one of the time series. All studies included in the systematic review used resampling to ensure the equal sampling interval. However, out of the 44 included studies only 14 studies[12 34 36 37 50 56 58 62 81 83 90 91 99 100] described the synchronisation process and in three studies[12 56 91] an automated method was performed, whereas in the remaining 11 studies[34 36 37 50 58 59 62 81 83 90 99 100] a manual timestamp correction or visual method was used (online supplemental table 6). The study by Sartor et al[12] was the only one included both an automated and visual inspection. Manual correction using time stamps or visual inspection is an option, but the process is time-consuming and potentially error prone. Several methods are available for the automated synchronisation of two independent time series[101 102] and we recommend this approach. The performance of different methods currently available has not been investigated and new methods are continuously being developed. This makes it difficult to propose one single method for the optimal temporal alignment. We recommend that studies use a method that is publicly documented and has been benchmarked with reference to a data set that has been manually edited or generated synthetically.

### Data smoothing

Different sources of error may affect the criterion assessment of the RR interval from recordings during both sedentary activities and strenuous exercise. Some errors are related to placement or handling of the device, which can be minimised by the correct application as proposed by the manufacturer. However, ectopic beats (ie, the heart either skips or adds an extra beat) and motion artefacts are errors that are inherent with both chest strap and electrode ECG devices and must be addressed to provide an accurate RR interval with the criterion measure. Only ten studies identified by our systematic literature review described a method to reduce spurious and incorrect HR data[12 34 36 37 56 59 81 82 90 91] (online supplemental table 6). Seven of these studies used an automated method (software), and three studies used a manual method but did not describe this in detail. In the HRV Task Force guideline, it is suggested that manual editing of the RR interval is required for the optimal identification and handling of ectopic beats and motion artefacts.[73 74] Manual editing of long duration recordings is time consuming and requires expert training. However, since the proposal of the HRV guidelines (and the update in 2015) several new methods have been evaluated for the automatic identification and handling of ectopic beats and motion artefacts.[74 103 104] Some of these new methods demonstrate good validity with the assessment of instantaneous HR from RR intervals and should be considered for the validity testing. As with the temporal alignment, currently there is no study available that compares the performance of all the different methods and, therefore, no single method is suggested for the optimal handling of ectopic beats and motion artefacts. We recommend that studies use a method that is publicly documented and has been benchmarked with reference to a data set that has been manually edited or generated synthetically.

## Statistical analysis

Since HR is a continuously scaled parameter, the analysis of accuracy should primarily be based on estimation of mean difference or mean relative difference and Bland-Altman limits of agreement (LoA) analysis.[105] The calculated estimates of mean difference and the LoA for the mean difference should always be accompanied with 95% confidence intervals (CIs). The acceptable accuracy expressed as mean difference (bpm) or percentage difference between the criterion measure and the index device may vary and needs to be evaluated individually considering the factors described above. The LoA for the absolute or relative difference are expected to contain 95% of paired differences for each measurement point by the two methods. However, the estimated LoA only apply to the specific study sample and because of sampling error, a new study sample from the same target population might provide different limits. Thus, if accuracy should be compared between different devices (ie, different models and/or manufacturers), it is important to provide the CIs of the LoA and the mean differences. Furthermore, for steady-state activities (in lab and semifree-living conditions) we also recommend that the LoA analysis should be based on both individually averaged mean differences of pairs of HR epochs across the activity duration, and in a repeated measure LoA analysis (multiple paired observations of HR epochs per individual). We acknowledge that validity testing may be performed in order to provide accuracy levels to consumers but also in order to further improve readings of a given device. While not necessarily informative for consumers, in research-related validity testing also proportional or fixed bias may need to be considered. If there is evidence of proportional bias, this suggests that the index device does not agree equally with the criterion through the range of measurements. In this situation, researchers could also use least-products regression as part of the Bland-Altman analysis, as reported by Ludbrook.[106] In case of violations of these assumptions, evaluators could attempt to log-transform index and criterion data or use a non-parametric approach, as described by Bland and Altman.[107]

A correlation coefficient could also be estimated (ie, Pearson's r or concordance correlation coefficient) to provide an additional measure of the relationship between the index and criterion measures, however, the limitations of these measures should be acknowledged as described previously[105] and repeated observations per individual should be taken into account if applicable.

Because HR is obtained in a time series in the wearable consumer device and the criterion measure, the mean difference and the LOAs should be estimated while taking into account multiple observations per individual.[107] We recommend that evaluators check and report on the assumptions for estimating mean difference and LoA. Accordingly, the paired differences in HR from the wearable consumer device and the criterion measure should have an approximately normal distribution, constant variance of the differences between the two, and no proportional bias.[107]

As an additional secondary measure of accuracy, we also recommend reporting the proportion of the evaluated epochs (eg, the exact RR time interval or the averaged HR over a segment of time) of the wearable consumer device that were within the predefined maximum allowed difference and a range of differences of greater and less than the predefined allowed difference.[108] For example, the number of evaluated epochs within ±20 bpm, ±15 bpm, ±10 bpm, ±5 bpm and ±2 bpm. Finally, because some consumer devices may remove data points, for example due to motion artefacts, we recommend reporting the proportion of such missing epochs (total time duration of recorded but missing epochs) of the total epochs recorded. Descriptive data on the study sample, number of paired observations, mean and SD of the HR obtained from the consumer device and the criterion, the mean differences (with SD and standard error), and the mean absolute error and mean absolute percentage error should also be reported.

The within-device precision (ie, reliability) should also be reported based on the data obtained for steady-state activities with a minimum duration of 2 min. To limit the possibility of true biological variation in HR within participants, the within-device precision should be evaluated using the average HR over five seconds separately in each steady-state activity (during rest and exercise) of at least 2 min duration. Furthermore, we suggest that 95% prediction intervals and intraclass correlations with 95% CIs should be calculated to estimate within-device precision according to recommendations.[108]

For a detailed comparison of the statistical analysis in criterion and index devices used in the studies reviewed, please refer to online supplemental table 5).

## RECOMMENDED VALIDATION PROTOCOL

Studies aiming to determine the validity of a consumer wearable should be designed to evaluate the device against an accurate criterion measure in a relevant study sample and in conditions that reflect the expected use of the device. Furthermore, the evaluation should be sufficiently described, and the data should be appropriately processed, analysed and reported. Considering the domains presented above, it appears that validation protocols should be carefully designed in order to account at least for the most robust sources of bias. Based on the current state of knowledge, figure 1 provides a graphical matrix of factors that need to be considered when validating PPG-based devices against a gold standard criterion measure. Detailed recommendations and guidelines are provided in online supplemental table 1. In addition, in table 1, we provide a checklist that is intended for planning of the validity protocols. Furthermore, in table 2, a more comprehensive protocol reporting sheet can be found and is intended to be used by both research institutions and manufacturers in order to facilitate standardised and transparent data sharing

## DISCUSSION AND FUTURE DIRECTIONS

This expert statement of the INTERLIVE Network aimed to provide recommendations and guidelines for the validity testing of consumer wearables assessing HR by PPG. In this context, considerations for the test preparation, sampling of participants, testing protocols, and activities as well as data handling, analysis and reporting were critically discussed. Based on a systematic literature review as well as our evidence-informed expert opinion, we have suggested a framework for validity testing of PPG-based devices measuring HR.

A rigorous evaluation of validity should be the mutual interest of manufacturers, shareholders, scientific institutions and consumers in order to judge whether a wearable device for assessment of HR is useful and performs with satisfactory accuracy. At present, the decision on whether a validation of a PPG-based wearable complies with medical certifications lies with the manufacturer, inevitably leading to a large heterogeneity in validation protocols. However, new regulations have been put in force on 25 May 2020, requiring all wearables (including devices assessing HR based on PPG) after a transition period of 3 years to follow regulations for medical devices, such as the US

Food and Drug Administration or the CE Marking in Europe.[109] Importantly these regulations are to be adhered to even if the devices are not intended to be used for medical evaluation, risk stratification or patient treatment. Consequently, the present expert statement should be understood as an attempt to foster standards in validations of PPG-based consumer wearables.

The urgent need for standardised validity testing is underlined by the wealth of different protocols identified by our systematic literature review. Interestingly, even though a variety of potential sources of bias were acknowledged in most of these studies, only few have attempted to account for methodological shortcomings in the data analysis. Based on the existing literature, solid evidence exists for artefacts originating from sex, BMI, body height as well factors related to the placement of the device, such as motion artefacts (originating both from the movement itself but also from possible shifts of the device on the skin) and skin tone. Conversely, currently little is known on the effects of cardiovascular diseases and their medical treatment (eg, beta blockers, ACE inhibitors, calcium channel blockers or diuretics), as well as skin damage (eg, scars and tattoos), ambient light and temperature or contact pressure on the measurement of HR by PPG. It is likely that these factors may have profound effects on the validity testing[51] but accounting for this remains challenging. Future studies should focus on the potential sources of bias that stem from both technological as well as population-based characteristics, in order further refine validation protocols.

When considering the proposed recommendations and guidelines, one has to bear in mind that our approach included an initial systematic literature review in order to assess which protocols have previously been used in the scientific literature for validity assessments of PPG-based HR monitors. Thus, we aimed at summarising all studies that have validated numerous index devices against a gold-standard criterion measure and extracted the specifics of these protocols. Consequently, assessing the study quality by means of a risk of bias assessment was not deemed useful as this analysis could only be used to evaluate the quality in respect to the particular outcome of each study (ie, the level of agreement of a certain device compared with a criterion) but it does not provide information on the quality of the validation protocols. Therefore, the potential sources of bias that were indirectly addressed in these studies were aligned with our evidence-informed expert opinions and provided the base for the presented framework.

It is obvious that a best-practice protocol for standardised validations will need to consider interests of both the scientific community and/or customers as well as that of the manufacturers. In that regard, the required investments that has to be made in hardware and software engineering might not be substantial as this is already in place with most manufacturers, but the validity assessment might require employment of additional educated staff and resources for the actual evaluations. Moreover, the resources required to conduct the validity evaluation seem to increase proportionally with the extent of the optimal evaluation. Considering the number of different devices commonly available with many manufacturers, it clearly suggests that feasibility and simplicity is important for proposing a validity evaluation that will be adopted by manufacturers. Consequently, we believe that the present expert statement provides a reasonable base for validity testing by incorporating high scientific standards.

Considering the wealth of new wearables entering the consumer market without prior proof of validity, providing consumers with wearables demonstrating excellent validity seems to be a great opportunity for new companies to conquer a substantial market share. Furthermore, accurate devices will also increase the usability of PPG for diagnostics and therapeutic monitoring. Considering the worldwide increasing prevalence of cardiovascular diseases,[110–112] accuracy will likely become a key criteria for PPG-based wearables. Indeed, prototypes of wrist-worn devices exist that can sense radial artery pulsation and use the data to estimate central aortic pressure.[113] It is likely that wearable devices will soon be capable of extrapolating blood pressure[114] or even blood glucose concentrations through optical sensors,[115] thus, underlining the importance of rigorousness and transparency in evaluating criterion validity. As such, we hope that the provided recommendations and checklists will be deemed useful by both researchers and manufacturers in order to further foster standardised validity testing.

## CONCLUSIONS

This expert statement provides an evidence-informed best-practice protocol for the validation of consumer wearables assessing HR by PPG. Our initial systematic literature review underlined a high degree of heterogeneity between previously published methods, with many studies failing to address key sources of bias. Thus, the INTERLIVE Consortium recommends that the proposed validation protocol could be used when considering the validation of any PPG-based consumer wearable assessing HR, in order to overcome the methodological shortcomings highlighted in this statement. Adherence to this validation standard will help ensure a transparent methodological and reporting consistency and facilitate comparison between consumer devices. This will ensure that manufacturers, consumers, healthcare providers and researchers can use this technology safely and to its full potential.

**Author affiliations**
[1]Institute of Cardiovascular Research and Sports Medicine, Department of Molecular and Cellular Sports Medicine, German Sport University Cologne, Cologne, Germany
[2]Department of Sports Medicine, Norwegian School of Sport Sciences, Oslo, Norway
[3]Department of Sports Science and Clinical Biomechanics, Research Unit for Exercise Epidemiology, Centre of Research in Childhood Health, University of Southern Denmark, Odense, Denmark
[4]Exercise and Health Laboratory, CIPER, Faculdade de Motricidade Humana, Universidade de Lisboa, Lisboa, Portugal
[5]CIDEFES - Centro de Investigação em Desporto, Educação Física e Exercício e Saúde, Universidade Lusófona, Lisboa, Portugal
[6]PROFITH "PROmoting FITness and Health through physical activity" Research Group, Sport and Health University Research Institute (iMUDS), Department of Physical and Sports Education, Faculty of Sport Sciences, University of Granada, Granada, Spain
[7]SFI Insight Centre for Data Analytics, University College Dublin, Dublin, Ireland
[8]School of Public Health, Physiotherapy and Sports Science, University College Dublin, Dublin, Ireland
[9]Exercise and Health Laboratory, CIPER, Faculdade de Motricidade Humana, Universidade de Lisboa, Lisboa, Cruz-Quebrada Dafundo, Portugal
[10]Department of Biosciences and Nutrition, Karolinska Institute, Stockholm, Sweden
[11]Exercise Translational Medicine Centre, the Key Laboratory of Systems Biomedicine, Ministry of Education, and Exercise, Health and Technology Centre, Department of Physical Education, Shanghai Jiao Tong University, Shanghai, China

**Twitter** William Johnston @Will_Johns10, Ulf Ekelund @Ulf_Ekelund and Moritz Schumann @moritz_schumann

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**ORCID iDs**
Jan M Mühlen http://orcid.org/0000-0001-9983-6403
Pedro B Judice http://orcid.org/0000-0003-2791-258X
Pablo Molina-Garcia http://orcid.org/0000-0001-6888-0997
William Johnston http://orcid.org/0000-0003-0525-6577
Luís B Sardinha http://orcid.org/0000-0002-6230-6027
Francisco B Ortega http://orcid.org/0000-0003-2001-1121
Brian Caulfield http://orcid.org/0000-0003-0290-9587
Ulf Ekelund http://orcid.org/0000-0003-2115-9267
Jan Christian Brønd http://orcid.org/0000-0001-6718-3022
Moritz Schumann http://orcid.org/0000-0001-9605-3489

## REFERENCES

1. Keith A, Flack M. The Form and Nature of the Muscular Connections between the Primary Divisions of the Vertebrate Heart. *J Anat Physiol* 1907;41:172–89.
2. Fox K, Borer JS, Camm AJ, *et al*. Resting heart rate in cardiovascular disease. *J Am Coll Cardiol* 2007;50:823–30.
3. Åstrand P-O, Ryhming I. A nomogram for calculation of aerobic capacity (physical fitness) from pulse rate during sub-maximal work. *J Appl Physiol* 1954;7:218–21.
4. Kim H-G, Cheon E-J, Bai D-S, *et al*. Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature. *Psychiatry Investig* 2018;15:235–45.
5. Ross R, Blair SN, Arena R, *et al*. Importance of Assessing Cardiorespiratory Fitness in Clinical Practice: A Case for Fitness as a Clinical Vital Sign: A Scientific Statement From the American Heart Association. *Circulation* 2016;134:e653–99.
6. Foster C, Rodriguez-Marroyo JA, de Koning JJ. Monitoring Training Loads: The Past, the Present, and the Future. *Int J Sports Physiol Perform* 2017;12:S22–8.
7. Buchheit M. Monitoring training status with HR measures: do all roads lead to Rome? *Front Physiol* 2014;5:73.
8. Cheung CC, Krahn AD, Andrade JG. The Emerging Role of Wearable Technologies in Detection of Arrhythmia. *Can J Cardiol* 2018;34:1083–7.
9. Pereira T, Tran N, Gadhoumi K, *et al*. Photoplethysmography based atrial fibrillation detection: a review. *NPJ Digit Med* 2020;3:3. doi:10.1038/s41746-019-0207-9
10. Piwek L, Ellis DA, Andrews S, *et al*. The Rise of Consumer Health Wearables: Promises and Barriers. *PLoS Med* 2016;13:e1001953.
11. Raja JM, Elsakr C, Roman S, *et al*. Apple Watch, Wearables, and heart rhythm: where do we stand? *Ann Transl Med* 2019;7:417.
12. Sartor F, Gelissen J, van Dinther R, *et al*. Wrist-worn optical and chest strap heart rate comparison in a heterogeneous sample of healthy individuals and in coronary artery disease patients. *BMC Sports Sci Med Rehabil* 2018;10:10.
13. Turakhia MP, Desai M, Hedlin H, *et al*. Rationale and design of a large-scale, app-based study to identify cardiac arrhythmias using a smartwatch: The Apple Heart Study. *Am Heart J* 2019;207:66–75.
14. Fortino GG, R.; Galzarano S. *Wearable Computing: From Modeling to Implementation of Wearable Systems Based on Body Sensor Networks*. Wiley, 2018.
15. Bunn JA, Navalta JW, Fountaine CJ, *et al*. Current State of Commercial Wearable Technology in Physical Activity Monitoring 2015-2017. *Int J Exerc Sci* 2018;11:503–15.
16. Sartor F, Papini G, Cox LGE, *et al*. Methodological Shortcomings of Wrist-Worn Heart Rate Monitors Validations. *J Med Internet Res* 2018;20:e10108.
17. Hardey MM. On the body of the consumer: performance-seeking with wearables and health and fitness apps. *Sociol Health Illn* 2019;41:991–1004.
18. Nauman J, Janszky I, Vatten LJ, *et al*. Temporal changes in resting heart rate and deaths from ischemic heart disease. *JAMA* 2011;306:2579–87.
19. John D, Sasaki J, Hickey A, *et al*. ActiGraph™ activity monitors: "the firmware effect". *Med Sci Sports Exerc* 2014;46:834–9.
20. Consumer Technology Association. *ANSI/CTA Standard. Physical Activity Monitoring for Heart Rate. ANSI/CTA-2065*, 2018.
21. Nelson BW, Low CA, Jacobson N, *et al*. Guidelines for wrist-worn consumer wearable assessment of heart rate in biobehavioral research. *NPJ Digit Med* 2020;3:90.
22. Bassett DR, Rowlands A, Trost SG. Calibration and validation of wearable monitors. *Med Sci Sports Exerc* 2012;44:S32–8.
23. Keadle SK, Lyden KA, Strath SJ, *et al*. A Framework to Evaluate Devices That Assess Physical Behavior. *Exerc Sport Sci Rev* 2019;47:206–14.
24. Freedson P, Bowles HR, Troiano R, *et al*. Assessment of physical activity using wearable monitors: recommendations for monitor calibration and use in the field. *Med Sci Sports Exerc* 2012;44:S1–4.
25. Welk GJ, McClain J, Ainsworth BE. Protocols for evaluating equivalency of accelerometry-based activity monitors. *Med Sci Sports Exerc* 2012;44:S39–49.
26. Huang X, Lin J, Demner-Fushman D. Evaluation of PICO as a knowledge representation for clinical questions. *AMIA Annu Symp Proc* 2006;2006:359–63.
27. Chatterjee S, Changawala N. Fragmented QRS complex: a novel marker of cardiovascular disease. *Clin Cardiol* 2010;33:68–71.
28. Inui T, Kohno H, Kawasaki Y, *et al*. Use of a Smart Watch for Early Detection of Paroxysmal Atrial Fibrillation: Validation Study. *JMIR Cardio* 2020;4:e14857.
29. Drexler M, Elsner C, Gabelmann V, *et al*. Apple Watch detecting coronary ischaemia during chest pain episodes or an apple a day may keep myocardial infarction away. *Eur Heart J* 2020;41:2224.
30. Hsiu H, Hsu CL, Wu TL. Effects of different contacting pressure on the transfer function between finger photoplethysmographic and radial blood pressure waveforms. *Proc Inst Mech Eng H* 2011;225:575–83.
31. Thomson EA, Nuss K, Comstock A, *et al*. Heart rate measures from the Apple Watch, Fitbit Charge HR 2, and electrocardiogram across different exercise intensities. *J Sports Sci* 2019;37:1411–9.
32. Stahl SE, An H-S, Dinkel DM, *et al*. How accurate are the wrist-based heart rate monitors during walking and running activities? Are they accurate enough? *BMJ Open Sport Exerc Med* 2016;2:e000106.
33. Etiwy M, Akhrass Z, Gillinov L, *et al*. Accuracy of wearable heart rate monitors in cardiac rehabilitation. *Cardiovasc Diagn Ther* 2019;9:262–71.
34. Georgiou KE, Dimov RK, Boyanov NB, *et al*. Feasibility of a New Wearable Device to Estimate Acute Stress in Novices During High-fidelity Surgical Simulation. *Folia Med (Plovdiv)* 2019;61:49–60.
35. Hahnen C, Freeman CG, Haldar N, *et al*. Accuracy of Vital Signs Measurements by a Smartwatch and a Portable Health Device: Validation Study. *JMIR Mhealth Uhealth* 2020;8:e16811.
36. Menghini L, Gianfranchi E, Cellini N, *et al*. Stressing the accuracy: Wrist-worn wearable sensor validation over different conditions. *Psychophysiology* 2019;56:e13441.
37. Dur O, Rhoades C, Ng MS, *et al*. Design Rationale and Performance Evaluation of the Wavelet Health Wristband: Benchtop Validation of a Wrist-Worn Physiological Signal Recorder. *JMIR Mhealth Uhealth* 2018;6:e11040.
38. Coca A, Roberge RJ, Williams WJ, *et al*. Physiological monitoring in firefighter ensembles: wearable plethysmographic sensor vest versus standard equipment. *J Occup Environ Hyg* 2009;7:109–14.
39. Pasadyn SR, Soudan M, Gillinov M, *et al*. Accuracy of commercially available heart rate monitors in athletes: a prospective study. *Cardiovasc Diagn Ther* 2019;9:379–85.
40. Dooley EE, Golaszewski NM, Bartholomew JB. Estimating Accuracy at Exercise Intensities: A Comparative Study of Self-Monitoring Heart Rate and Physical Activity Wearable Devices. *JMIR Mhealth Uhealth* 2017;5:e34.
41. Gillinov S, Etiwy M, Wang R, *et al*. Variable Accuracy of Wearable Heart Rate Monitors during Aerobic Exercise. *Med Sci Sports Exerc* 2017;49:1697–703.
42. MJ L, Zhong WH, Liu YX, *et al*. Sample Size for Assessing Agreement between Two Methods of Measurement by Bland-Altman Method. *Int J Biostat* 2016;12.
43. Carstensen B. *Comparing Clinical Measurement Methods: A Practical Guide*. Wiley, 2010.
44. Allen J, Murray A. Age-related changes in peripheral pulse timing characteristics at the ears, fingers and toes. *J Hum Hypertens* 2002;16:711–7.
45. O'Driscoll R, Turicchi J, Hopkins M, *et al*. The validity of two widely used commercial and research-grade activity monitors, during resting, household and activity behaviours. *Health Technol* 2020;10:637–48.
46. Shcherbina A, Mattsson C, Waggott D, *et al*. Accuracy in Wrist-Worn, Sensor-Based measurements of heart rate and energy expenditure in a diverse cohort. *J Pers Med* 2017;7:3.
47. Chow H-W, Yang C-C. Accuracy of Optical Heart Rate Sensing Technology in Wearable Fitness Trackers for Young and Older Adults: Validation and Comparison Study. *JMIR Mhealth Uhealth* 2020;8:e14707.

48 Dehghanojamahalleh S, Kaya M. Sex-Related Differences in Photoplethysmography Signals Measured From Finger and Toe. *IEEE J Transl Eng Health Med* 2019;7:1–7.

49 Firooz A, Rajabi-Estarabadi A, Zartab H, *et al*. The influence of gender and age on the thickness and echo-density of skin. *Skin Res Technol* 2017;23:13–20.

50 Horton JF, Stergiou P, Fung TS, *et al*. Comparison of Polar M600 Optical Heart Rate and ECG Heart Rate during Exercise. *Med Sci Sports Exerc* 2017;49:2600–7.

51 Tamura T, Chen W. *Seamless Healthcare Monitoring: Advancements in Wearable, Attachable and Invisible Devices*. Springer International Publishing, 2018.

52 Wallen MP, Gomersall SR, Keating SE, *et al*. Accuracy of Heart Rate Watches: Implications for Weight Management. *PLoS One* 2016;11:e0154420.

53 Claes J, Buys R, Avila A, *et al*. Validity of heart rate measurements by the Garmin Forerunner 225 at different walking intensities. *J Med Eng Technol* 2017;41:480–5.

54 Hendrikx J, Ruijs LS, Cox LG, *et al*. Clinical Evaluation of the Measurement Performance of the Philips Health Watch: A Within-Person Comparative Study. *JMIR Mhealth Uhealth* 2017;5:e10.

55 Nelson BW, Allen NB. Accuracy of Consumer Wearable Heart Rate Measurement During an Ecologically Valid 24-Hour Period: Intraindividual Validation Study. *JMIR Mhealth Uhealth* 2019;7:e10828.

56 Hermand E, Cassirame J, Ennequin G, *et al*. Validation of a Photoplethysmographic Heart Rate Monitor: Polar OH1. *Int J Sports Med* 2019;40:462–7.

57 Khushhal A, Nichols S, Evans W, *et al*. Validity and Reliability of the Apple Watch for Measuring Heart Rate During Exercise. *Sports Med Int Open* 2017;1:E206–11.

58 Spierer DK, Rosen Z, Litman LL, *et al*. Validation of photoplethysmography as a method to detect heart rate during rest and exercise. *J Med Eng Technol* 2015;39:264–71.

59 Konstantinou P, Trigeorgi A, Georgiou C, *et al*. Comparing apples and oranges or different types of citrus fruits? Using wearable versus stationary devices to analyze psychophysiological data. *Psychophysiology* 2020;57:e13551.

60 Brazendale K, Decker L, Hunt ET, *et al*. Validity and Wearability of Consumer-based fitness trackers in free-living children. *Int J Exerc Sci* 2019;12:471–82.

61 Pope ZC, Lee JE, Zeng N, *et al*. Validation of Four Smartwatches in Energy Expenditure and Heart Rate Assessment During Exergaming. *Games Health J* 2019;8:205–12.

62 Müller AM, Wang NX, Yao J, *et al*. Heart Rate Measures From Wrist-Worn Activity Trackers in a Laboratory and Free-Living Setting: Validation Study. *JMIR Mhealth Uhealth* 2019;7:e14120.

63 Sañudo B, De Hoyo M, Muñoz-López A, *et al*. Pilot Study Assessing the Influence of Skin Type on the Heart Rate Measurements Obtained by Photoplethysmography with the Apple Watch. *J Med Syst* 2019;43:195.

64 Bayès de Luna A. *Clinical Electrocardiography: A Textbook*. 4th edn. Wiley-Blackwell, 2012.

65 Satija U, Ramkumar B, Manikandan MS. A Review of Signal Processing Techniques for Electrocardiogram Signal Quality Assessment. *IEEE Rev Biomed Eng* 2018;11:36–52.

66 Giles D, Draper N, Neil W. Validity of the Polar V800 heart rate monitor to measure RR intervals at rest. *Eur J Appl Physiol* 2016;116:563–71.

67 Tobon DP, Jayaraman S, Falk TH. Spectro-Temporal Electrocardiogram Analysis for Noise-Robust Heart Rate and Heart Rate Variability Measurement. *IEEE J Transl Eng Health Med* 2017;5:1–11.

68 Kingsley M, Lewis MJ, Marson RE. Comparison of Polar 810s and an ambulatory ECG system for RR interval measurement during progressive exercise. *Int J Sports Med* 2005;26:39–44.

69 Nunan D, Donovan G, Jakovljevic DG, *et al*. Validity and reliability of short-term heart-rate variability from the Polar S810. *Med Sci Sports Exerc* 2009;41:243–50.

70 Caminal P, Sola F, Gomis P, *et al*. Validity of the Polar V800 monitor for measuring heart rate variability in mountain running route conditions. *Eur J Appl Physiol* 2018;118:669–77.

71 Weippert M, Kumar M, Kreuzfeld S, *et al*. Comparison of three mobile devices for measuring R-R intervals and heart rate variability: Polar S810i, Suunto t6 and an ambulatory ECG system. *Eur J Appl Physiol* 2010;109:779–86.

72 Wallén MB, Hasson D, Theorell T, *et al*. Possibilities and limitations of the Polar RS800 in measuring heart rate variability at rest. *Eur J Appl Physiol* 2012;112:1153–65.

73 force T. Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. Task force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. *Eur Heart J* 1996;17:354–81.

74 Sassi R, Cerutti S, Lombardi F, *et al*. Advances in heart rate variability signal analysis: joint position statement by the e-Cardiology ESC Working Group and the European Heart Rhythm Association co-endorsed by the Asia Pacific Heart Rhythm Society. *Europace* 2015;17:1341–53.

75 Bent B, Goldstein BA, Kibbe WA, *et al*. Investigating sources of inaccuracy in wearable optical heart rate sensors. *NPJ Digit Med* 2020;3:18.

76 Castaneda D, Esparza A, Ghamari M, *et al*. A review on wearable photoplethysmography sensors and their potential future applications in health care. *Int J Biosens Bioelectron* 2018;4:195–202.

77 Lemay M, Bertschi M, Sola J, *et al*. Chapter 2.3 - Application of Optical Heart Rate Monitoring. In: Sazonov E, Neuman MR, eds. *Wearable Sensors*. Oxford: Academic Press, 2014: 105–29.

78 Zhang Y, Song S, Vullings R, *et al*. Motion Artifact Reduction for Wrist-Worn Photoplethysmograph Sensors Based on Different Wavelengths. *Sensors (Basel)* 2019;19:673.

79 Tăuțan A-M, Young A, Wentink E, *et al*. Characterization and reduction of motion artifacts in photoplethysmographic signals from a wrist-worn device. *Annu Int Conf IEEE Eng Med Biol Soc* 2015;2015:6146-9.

80 Tamura T, Maeda Y, Sekine M, *et al*. Wearable Photoplethysmographic Sensors—Past and Present. *Electronics* 2014;3:282–302.

81 Passler S, Müller N, Senner V. In-Ear Pulse Rate Measurement: A Valid Alternative to Heart Rate Derived from Electrocardiography? *Sensors (Basel)* 2019;19:3641.

82 Zheng Y, Leung B, Sy S, *et al*. A clip-free eyeglasses-based wearable monitoring device for measuring photoplethysmograhic signals. *Annu Int Conf IEEE Eng Med Biol Soc* 2012;2012:5022–5.

83 Jo E, Lewis K, Directo D, *et al*. Validation of Biofeedback Wearables for Photoplethysmographic Heart Rate Tracking. *J Sports Sci Med* 2016;15:540–7.

84 Allen J. Photoplethysmography and its application in clinical physiological measurement. *Physiol Meas* 2007;28:R1–39.

85 Falter M, Budts W, Goetschalckx K, *et al*. Accuracy of Apple Watch Measurements for Heart Rate and Energy Expenditure in Patients With Cardiovascular Disease: Cross-Sectional Study. *JMIR Mhealth Uhealth* 2019;7:e11889.

86 Parak J, Uuskoski M, Machek J, *et al*. Estimating Heart Rate, Energy Expenditure, and Physical Performance With a Wrist Photoplethysmographic Device During Running. *JMIR Mhealth Uhealth* 2017;5:e97.

87 Maeda Y, Sekine M, Tamura T. The advantages of wearable green reflected photoplethysmography. *J Med Syst* 2011;35:829–34.

88 Abt G, Bray J, Benson AC. The validity and inter-device variability of the Apple Watch™ for measuring maximal heart rate. *J Sports Sci* 2018;36:1447–52.

89 Cadmus-Bertram L, Gangnon R, Wirkus EJ, *et al*. The Accuracy of Heart Rate Monitoring by Some Wrist-Worn Activity Trackers. *Ann Intern Med* 2017;166:610–2.

90 Hettiarachchi IT, Hanoun S, Nahavandi D, *et al*. Validation of Polar OH1 optical heart rate sensor for moderate and high intensity physical activities. *PLoS One* 2019;14:e0217288.

91 Parak J, Korhonen I. Evaluation of wearable consumer heart rate monitors based on photopletysmography. *Annu Int Conf IEEE Eng Med Biol Soc* 2014;2014:3670–3.

92 Støve MP, Haucke E, Nymann ML, *et al*. Accuracy of the wearable activity tracker Garmin Forerunner 235 for the assessment of heart rate during rest and activity. *J Sports Sci* 2019;37:895–901.

93 Bai Y, Hibbing P, Mantis C, *et al*. Comparative evaluation of heart rate-based monitors: Apple Watch vs Fitbit Charge HR. *J Sports Sci* 2018;36:1734–41.

94 Reddy RK, Pooni R, Zaharieva DP, *et al*. Accuracy of Wrist-Worn Activity Monitors During Common Daily Physical Activities and Types of Structured Exercise: Evaluation Study. *JMIR Mhealth Uhealth* 2018;6:e10338.

95 Boudreaux BD, Hebert EP, Hollander DB, *et al*. Validity of Wearable Activity Monitors during Cycling and Resistance Exercise. *Med Sci Sports Exerc* 2018;50:624–33.

96 McArdle WKF, Katch V. *Exercise Physiology: Energy, Nutrition and Human Performance*. 8 edn. Lippincott Williams & Wilkins, 2014.

97 Johnson JM. Physical training and the control of skin blood flow. *Med Sci Sports Exerc* 1998;30:382–6.

98 Gorny AW, Liew SJ, Tan CS, *et al*. Fitbit charge HR wireless heart rate monitor: validation study conducted under free-living conditions. *JMIR Mhealth Uhealth* 2017;5:e157.

99 Kroll RR, Boyd JG, Maslove DM. Accuracy of a Wrist-Worn Wearable Device for Monitoring Heart Rates in Hospital Inpatients: A Prospective Observational Study. *J Med Internet Res* 2016;18:e253.

100 Pelizzo G, Guddo A, Puglisi A, *et al*. Accuracy of a Wrist-Worn Heart Rate Sensing Device during Elective Pediatric Surgical Procedures. *Children (Basel)* 2018;5:38.

101 Rhudy M. Time Alignment Techniques for Experimental Sensor Data. *IJCSES* 2014;5:1–14.

102 Coakley KJ, Hale P. Alignment of Noisy Signals. *IEEE Trans Instrum Meas* 2001;50:141–9.

103 Ghaleb FA, Kamat MB, Salleh M, *et al*. Two-stage motion artefact reduction algorithm for electrocardiogram using weighted adaptive noise cancelling and recursive Hampel filter. *PLoS One* 2018;13:e0207176.

104 Luo S, Johnston P. A review of electrocardiogram filtering. *J Electrocardiol* 2010;43:486–96.

105 Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986;1:307–10.

106 Ludbrook J. Statistical techniques for comparing measurers and methods of measurement: a critical review. *Clin Exp Pharmacol Physiol* 2002;29:527–36.

107 Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999;8:135–60.

108 Kottner J, Audigé L, Brorson S, *et al*. Guidelines for Reporting Reliability and Agreement Studies (GRRAS) were proposed. *J Clin Epidemiol* 2011;64:96–106.

109 Ravizza S, De Maria C, Di Pietro L, *et al*. Comprehensive Review on Current and Future Regulatory Requirements on Wearable Sensors in Preclinical and Clinical Testing. *Front Bioeng Biotechnol* 2019;7:313.

110 Shah AD, Langenberg C, Rapsomaniki E, *et al*. Type 2 diabetes and incidence of cardiovascular diseases: a cohort study in 1·9 million people. *Lancet Diabetes Endocrinol* 2015;3:105–13.

111 Andersson C, Vasan RS. Epidemiology of cardiovascular disease in young individuals. *Nat Rev Cardiol* 2018;15:230–40.

112 Bansilal S, Castellano JM, Fuster V. Global burden of CVD: focus on secondary prevention of cardiovascular disease. *Int J Cardiol* 2015;201(Suppl 1):S1–7.

113 Massoomi MR, Handberg EM. Increasing and Evolving Role of Smart Devices in Modern Medicine. *Eur Cardiol* 2019;14:181–6.

114 Elgendi M, Fletcher R, Liang Y, *et al*. The use of photoplethysmography for assessing hypertension. *NPJ Digit Med* 2019;2:60.

115 Rodin D, Kirby M, Sedogin N, *et al*. Comparative accuracy of optical sensor-based wearable system for non-invasive measurement of blood glucose concentration. *Clin Biochem* 2019;65:15–20.

116 Vandenberk T, Stans J, Mortelmans C, *et al*. Clinical validation of heart rate Apps: mixed-methods evaluation study. *JMIR Mhealth Uhealth* 2017;5:e129.

117 Wang Z, Fu S. Evaluation of a strapless heart rate monitor during simulated flight tasks. *J Occup Environ Hyg* 2016;13:185–92.

118 Chellakumar PJ, Brumfield A, Kunderu K, *et al*. Heart rate variability: comparison among devices with different temporal resolutions. *Physiol Meas* 2005;26:979–86.

119 Engström E, Ottosson E, Wohlfart B, *et al*. Comparison of heart rate measured by polar RS400 and ECG, validity and repeatability. *Adv Physiother* 2012;14:115–22.

120 Gilgen-Ammann R, Schweizer T, Wyss T. RR interval signal quality of a heart rate monitor and an ECG Holter at rest and during exercise. *Eur J Appl Physiol* 2019;119:1525–32.

121 Romagnoli M, Alis R, Guillen J, *et al*. A novel device based on smart textile to control heart's activity during exercise. *Australas Phys Eng Sci Med* 2014;37:377–84.