

Prospective Validation of Smartphone-based Heart Rate and Respiratory Rate Measurement Algorithms

Sean Bae

Google Health, Palo Alto

Silviu Borac

Google Health, Palo Alto

Yunus Emre

Google Health, Palo Alto

Jonathan Wang

Google Health, Palo Alto

Jiang Wu

Google Health, Palo Alto

Mehr Kashyap

Work done at Google Health via Advanced Clinical, Deerfield

Si-Hyuck Kang (✉ eandp303@snu.ac.kr)

Seoul National University Bundang Hospital <https://orcid.org/0000-0003-4755-6367>

Liwen Chen

Google Health, Palo Alto

Melissa Moran

Google (United States)

Julie Cannon

Work done at Google Health via Advanced Clinical, Deerfield

Eric Teasley

Google Health, Palo Alto

Allen Chai

Google Health, Palo Alto

Yun Liu

Google <https://orcid.org/0000-0003-4079-8275>

Neal Wadhwa

Google Research, Cambridge

Mike Krainin

Google Research, Cambridge

Michael Rubinstein

Google Research, Cambridge

Alejandra Maciel

Google Health, Palo Alto

Michael McConnell

Google Health, Palo Alto

Shwetak Patel

Google Health, Palo Alto

Greg Corrado

Google, Inc

James Taylor

Google Health, Palo Alto

Jiening Zhan

Google Health, Palo Alto

Ming Po

Ansible Health, Mountain View

Article

Keywords: Validation of Smartphone, Heart Rate, Respiratory Rate, Measurement Algorithms

Posted Date: June 22nd, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-620411/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at Communications Medicine on April 12th, 2022. See the published version at <https://doi.org/10.1038/s43856-022-00102-x>.

Abstract

Measuring vital signs plays a key role in both patient care and wellness, but can be challenging outside of medical settings due to the lack of specialized equipment. In this study, we prospectively evaluated smartphone camera-based techniques for measuring heart rate (HR) and respiratory rate (RR) for consumer wellness use. HR was measured by placing the finger over the rear-facing camera, while RR was measured via a video of the participants sitting still in front of the front-facing camera. In the HR study of 95 participants (with a protocol that included both measurements at rest and post exercise), the mean absolute percent error (MAPE) \pm standard deviation of the measurement was $1.6\% \pm 4.3\%$, which was significantly lower than the pre-specified goal of 5%. No significant differences in the MAPE were present across colorimeter-measured skin-tone subgroups: $1.8\% \pm 4.5\%$ for very light to intermediate, $1.3\% \pm 3.3\%$ for tan and brown, and $1.8\% \pm 4.9\%$ for dark. In the RR study of 50 participants, the mean absolute error (MAE) was 0.78 ± 0.61 breaths/min, which was significantly lower than the pre-specified goal of 3 breath/min. The MAE was low in both healthy participants (0.70 ± 0.67 breaths/min), and participants with chronic respiratory conditions (0.80 ± 0.60 breaths/min). These results validate the accuracy of our smartphone camera-based techniques to measure HR and RR across a range of pre-defined subgroups.

Introduction

Measurement of heart rate (HR) and respiratory rate (RR), two of the four cardinal vital signs—HR, RR, body temperature, and blood pressure—is the starting point of physical assessment for both health and wellness. However, taking these standard measurements via a physical examination becomes challenging in telehealth, remote care, and consumer wellness settings.¹⁻³ In particular, the recent COVID-19 pandemic has accelerated trends towards telehealth and remote triage, diagnosis, and monitoring.^{4,5} Although specialized devices are commercially available for consumers and have the potential to motivate healthy behaviors,⁶ their cost and relatively low adoption limit general usage.

On the other hand, with smartphone penetration exceeding 40% globally and 80% in the US,⁷ up to 3.8 billion individuals already have access to a myriad of sensors and hardware (video cameras with flash, accelerometers, gyroscope, etc.) that are changing the way people interact with each other and their environments. A combination of these same sensors together with novel computer algorithms can be used to measure vital signs via consumer-grade smartphones.⁸⁻¹² Indeed, several such mobile applications (“apps”) are available, some with hundreds of thousands of installs.¹³ However, these apps seldom undergo rigorous clinical validation for accuracy and generalizability to important population and patient subgroups.

In this work, we present and validate two algorithms that make use of smartphone cameras for vital sign measurements. The first algorithm leverages photoplethysmography (PPG) acquired using smartphone cameras for HR measurement. PPG signals are recorded by placing a finger over the camera lens, and the color changes captured in the video are used to determine the oscillation of blood volume after each

heart beat.¹⁴ In the second algorithm, we leverage upper-torso videos obtained via the front-facing smartphone camera to track the physical motion of breathing to measure RR. Herein, we describe both the details of the algorithms themselves and report on the performance of these two algorithms in prospective clinical validation studies. For the HR study, we sought to demonstrate reliable and consistent accuracy on diverse populations (in terms of objectively measured skin tones, ranging from very light to dark skin), whereas for the RR study, we aimed to demonstrate robust performance in subgroups with and without chronic respiratory conditions.

Results

We conducted two separate prospective studies to validate the performance of smartphone-based HR and RR measurements (Figure 1). The user interfaces of the two custom research apps are shown in Supplementary Figure 1. The HR algorithm used PPG signals measured from the study participants placing their finger over the rear camera, and the enrollment for the corresponding validation study was stratified to ensure diversity across objectively measured skin tones. The RR algorithm used video captures of the face and upper torso, and the enrollment for the corresponding validation study was stratified to capture participants with and without chronic respiratory conditions.

Heart rate measurement

A total of 101 participants were enrolled. Study eligibility criteria are described in Methods (Supplementary Table 1A). After excluding one participant who was found to have an exclusion criterion (pregnancy), there were 100 valid enrollees. Among these, 3 were withdrawn due to skin tone distribution requirements, and 2 were withdrawn during data collection due to difficulty in data collection (failure to follow the instruction on holding the phone and a health condition that prevents us from collecting reference HR). Thus, 95 participants completed data collection (Supplementary Figure 2A). The participants had a mean age of 41.8 years, 75% were female, and skin-tone subgroups were evenly distributed as planned: 33% were subgroup 1 (very light, light and intermediate), 34% were subgroup 2 (tan and brown), and 34% were subgroup 3 (dark) (Table 1). Skin tone measurement and categorization criteria are described in Methods and Supplementary Table 2.

From these participants, 379 total recordings were attempted. A valid HR was successfully obtained (see details on signal-to-noise ratio [SNR] in Methods) in 361 cases (95.3%). The success rate increased with retries up to 3 times: 316 measurements (83.4%) were successful on the first try, another 31 measurements (cumulative 91.6%) on the second try, and another 14 measurements (cumulative 95.3%) on the third try. The baseline characteristics of the 14 participants for whom HR values were not successfully reported by the study app for at least one measurement (due to low SNR) did not differ significantly from the remaining participants (Supplementary Table 3). In addition, a corresponding valid reference HR could not be obtained for 9 recordings from 4 participants due to technical error. The remaining 352 recordings with paired valid reference HR contributed to the final analysis (Supplementary

Figure 1A). The average reference HR was 79.8 ± 14.6 beats/min overall, 75.5 ± 11.2 beats/min at rest, and 92.9 ± 16.6 beats/min post-exercise (Supplementary Table 4).

Compared to the reference HR, the mean absolute percentage error (MAPE) of the overall study population was 1.63%. The MAPE values for all 4 data collection windows (including 3 at rest and 1 post-exercise measurements) were significantly lower than the pre-specified study target of 5% ($p < 0.001$ for all comparison). The MAPE showed a left-skewed distribution with a long tail (median, 1.14%; range, 0.0–50.6%). The MAPE by skin-tone subgroup was 1.77% for subgroup 1, 1.32% for subgroup 2, and 1.77% for subgroup 3. The MAPE values by skin-tone subgroups were all significantly $< 5\%$ for each data window (all P values < 0.001) (Figure 1 and Supplementary Table 4). We also found no statistically significant variation in MAPE across the three different lighting conditions.

Figure 2 shows the Bland-Altman plots for comparing the algorithm-estimated HR with the reference HR for the overall population and the three subgroups. Most observations (344/352, 97.8%) were within 5 beats/min. Supplementary Figure 3 shows the Bland-Altman plots for HR stratified by at-rest versus post-exercise.

Respiratory rate measurement

A total of 50 participants were enrolled in the RR study, including 10 healthy participants and 40 participants with chronic respiratory conditions, i.e., chronic obstructive pulmonary disease (COPD) or asthma (Supplementary Table 1B, and Supplementary Figure 2B). All of the 50 study participants contributed to the final analysis. Self-identified baseline characteristics are presented in Table 1. The mean age was 50 years old; 80% self-identified as White, 14% as Black or African American, and 46% as having Hispanic or Latino ethnicity. The average reference RR was 15.3 ± 3.7 breaths/min (Supplementary Table 5).

Both versions of the algorithm successfully estimated RR in all of the study subjects. The mean absolute error (MAE) in the overall study population was 0.84 ± 0.97 and 0.78 ± 0.61 breaths/min for algorithm versions A and B, respectively (Figure 1 and Supplementary Table 5), which were significantly lower than the pre-specified threshold of 3 breaths/min ($p < 0.001$ for both). Each subgroup also showed MAE values significantly lower than the threshold: algorithm version A, 0.60 ± 0.52 breaths/min ($p < 0.001$) for the healthy cohort and 0.90 ± 1.05 breaths/min ($p < 0.001$) for the cohort with chronic respiratory conditions; algorithm version B, 0.70 ± 0.67 breaths/min ($p < 0.001$) and 0.80 ± 0.60 breaths/min ($p < 0.001$), respectively. No statistically significant differences across age and race subgroups were seen (Supplementary Table 6).

Figure 3 shows the Bland-Altman plots for comparing the algorithm-estimated RR with the reference RR for the overall population and the two subgroups. All observations were within 2 breaths/min of the reference RR for algorithm version B, while one observation was >2 breaths/min of the reference RR for version A. The accuracy of the two algorithm versions did not differ significantly ($p = 0.70$).

Supplementary Figure 4 shows the results from a user experience survey (see Methods) of the study participants. More than 90% of responses were classified as positive, with participants reporting anticipated ease in setting up within a home environment, ease in following the instructions in the app, and comfort using the app to assess general wellness.

Discussion

We report the results of two prospective clinical studies validating the performance of smartphone algorithms to estimate HR and RR. Both algorithms showed high accuracy compared to the reference standard vital sign measurements, with the mean MAPE for HR < 2% and the mean MAE for RR < 1 breaths/min (both significantly below the pre-specified targets). In addition, the HR estimation was robust across the full range of skin tones, and the RR estimation generalized to participants with common chronic respiratory conditions, COPD and asthma.

Only a limited number of previous smartphone apps have undergone clinical evaluation for HR measurement.¹⁵ The accuracy of the HR algorithm in this study is especially notable. An MAE less than 5 beats/min or a MAPE less than 10% are standard accuracy thresholds for HR monitors.^{16,17} The MAE of 1.32 beats/min in HR is lower than that reported for smartphone apps (2.0 to 8.1 beats/min)¹⁵ and for contemporary wearable devices (4.4 to 10.2 beats/min at rest)¹⁸, albeit with several differences in study design and population. The MAPE of 1.63% is comparable to the performance of current wearable devices. Shcherbina et al. tested six wrist worn devices and reported a median error <5% for each across various activities and a median error of 2.0% for the best-performing device.¹⁹ Because skin tone can be a potential source of bias in medical devices,²⁰⁻²² and the accuracy of PPG-based HR estimation can be affected by melanin's light-absorbing property,^{18,23} we enrolled participants with diverse skin tones to validate the robustness of our HR estimation algorithm across skin tones.

The authors are unaware of any smartphone-based RR measurement apps that have undergone rigorous clinical validation. One previous study tested an algorithm with a similar concept to ours, but enrolled only 10 healthy subjects.²⁴ For consumer-grade RR monitoring devices, there is no well-accepted accuracy standard.²⁵ Our MAE of 0.78 breaths/min is comparable to that of professional healthcare devices, which have reported accuracy of $\pm 2-3$ breaths/min.²⁶⁻²⁹ One important strength of our approach is determining RR from direct measurement of respiratory motion, rather than trying to derive RR from variation of PPG-based interbeat interval, which has limitations. The MAE found in this study could be a helpful reference point for future studies. In this study, we tested two algorithm versions for RR estimation that differed only in the SNR threshold. Our results suggest that this parameter had little impact on the accuracy or error rates.

This work supports the use of consumer-grade smartphones for measuring HR and RR. One application of these measurements is in fitness and wellness for the general consumer user. Specifically, an elevated resting HR or slower heart rate recovery after exercise has been linked to lower physical fitness and higher

risk of all-cause mortality.^{30,31} Evidence suggests use of direct-to-consumer mobile health technologies may enhance positive lifestyle modification such as increased physical activity, more weight loss, and better diabetes control.^{32–34} Tracking one’s own health-related parameters over time by the general public can potentially increase motivation for a healthier lifestyle by providing an objective, quantifiable metric.⁶ Additionally, there exists strong evidence that regular physical activity is key to improving one’s health independent of age, sex, race, ethnicity, or current fitness level for maintaining cardiovascular health.³⁵ Monitoring one’s HR is also an easy and effective way to assess and adjust exercise intensity, or enable smartphone-based measurement of cardiorespiratory fitness.^{36–38}

With further clinical validation across broad populations, such smartphone-based measurement could also be useful in various settings, most notably telehealth where vital sign measurement is challenging due to the remote nature of the patient encounter.^{39,40} Though patients can in principle count their own HR or RR, this can be error prone due to factors such as biases that acute awareness of the self-examination can cause.^{41,42} Because the demand for remote triage, diagnosis and monitoring is burgeoning in the wake of the COVID-19 pandemic, there is increased attention being paid to accurate remote physical examination.^{3,20,21}

There are several limitations in this work. First, our quantitative results focused on specific study devices (Pixel 3 and 4) and quantitative data on generalization to other devices will be needed. The current studies were also conducted in a controlled setting with structured study protocols. Though the participants used these features without significant study staff assistance, their ease of use in a general population will need further study. Next, our reference HR comes from a clinical PPG device instead of an electrocardiogram. There may be infrequent instances of electromechanical dissociation (such as various heart blocks, ventricular tachycardia, etc), in which a pulse rate measured at the periphery may not be the same with the reference electrical “heart” rate.⁴³ Our HR algorithm is similarly subject to such errors because it relies on the pulsatile movement of blood in the fingertips. Further, awareness of self-measurement may affect users’ RR. A study demonstrated that people may have lower RR when they are aware that they are monitored by observers.⁴⁴ It is yet unclear how awareness would affect RR measurements using apps in the absence of human-to-human interaction. In addition, though the study enrollment was optimized for diversity, this impacted the sample size in each subpopulation and the number of covariates that can be analyzed. Larger real-world validation that also controls for additional factors such as body-mass index will be helpful. Also, these clinical validation studies aimed to evaluate the algorithms at a “steady state” and further work will be needed to investigate acute clinical scenarios such as elevated or depressed HR or RR in urgent medical situations. Last but not least, although both HR and RR estimation algorithms met the predefined goals, observations with high deviation from the reference values were still produced by the algorithms (albeit infrequently). Future work to reduce errors is needed.

In addition to the clinical validation studies reported in this work, these HR and RR algorithms are currently undergoing additional broad usability testing across users, different Android devices, and

different environments as we make the algorithms more widely accessible for consumers, beginning by incorporating into the Google Fit app.

Conclusion

We developed HR- and RR-measurement algorithms for smartphones, and conducted two clinical studies to validate their accuracy in various study populations. Both algorithms showed acceptable error ranges with a MAPE under 2% for HR and MAE under 1 breath/min for RR. These algorithms may prove useful in wellness settings such as fitness monitoring. Additional research is warranted before consideration for any future use in clinical settings, such as remote physical examination.

Methods

HR measurement

Algorithm description

Prior work in computer vision to extract heart rate from RGB (red-green-blue) video signals has leveraged manually extracted features in PPG signals from the finger for arrhythmia detection,⁴⁵ ballistocardiographic movements from fingertips,⁴⁶ red-channel PPG from fingertip videos,⁴⁷ and the relationship between RGB channels.⁴⁸

Our method estimates HR by optically measuring the PPG waveform from participants' fingertips and then extracting the dominant frequency. First, several rectangular regions of interest (ROI) were manually selected from the video frames (linear RGB at 15 frames per second and at a resolution of 640x480 pixels). The chosen ROIs were the full frame, the left half, the right half, the top half, and the bottom half of the frames. Since camera pixels are illuminated non-homogeneously, signal strength can have spatial variations across pixels.⁴⁹ Our method simultaneously analyzes different ROIs to identify one with the greatest SNR.

Pixels in each ROI were averaged per-channel to reduce the effects of sensor and quantization noise, similar to prior work.⁴⁸ The pulsatile blood volume changes were present as the AC components in these smoothed signals. We then weighted the three RGB waveforms to predict a single PPG waveform (after an empirical grid search across all 3 channels: RGB, we arrived at weights 0.67, 0.33, and 0, respectively) for each ROI.

The resulting PPG waveforms were bandpass filtered to remove low- and high-frequency noise unlikely to be valid HR. Filter cut-off frequencies corresponded to a low of 30 beats/min and high of 360 beats/min. Next, large amplitude changes in PPGs due to motion were suppressed by limiting maximum allowed changes in amplitudes as 3x of moving average values. Then, frequency domain representations of PPGs were computed using the Fast Fourier Transform (FFT), from which we identified the dominant

frequencies with maximum power. Because the PPG signals are periodic with multiple harmonics, the powers of the base frequencies were computed by summing the powers of their first, second, and third harmonics. SNRs were estimated for each ROI by computing the ratio between the power of the dominant frequency and the powers of non-dominant frequencies on a logarithmic scale. ROIs were filtered to only those with a SNR ≥ 0 dB, and the dominant frequency of the ROI with the highest SNR was reported. If no such ROI existed, no HR was reported.

Study design and participants

We performed a prospective observational clinical validation study to assess the accuracy of the study algorithm in estimating HR in individuals of diverse skin tones (Supplementary Figure 1A). Participants were enrolled at a clinical research site (Meridian, Savannah, GA) from October 2020 to December 2020. Study eligibility criteria were limited to excluding participants with significant tremor or inability to perform physical activity. The inclusion/exclusion criteria are detailed in Supplementary Table 1A. Study enrollment was stratified into 3 skin-tone subgroups (mapped to Fitzpatrick skin types;⁵⁰ see Supplementary Table 2) to ensure broad representation: (1) types 1–3 (very light, light and intermediate); (2) types 4–5 (tan and brown), and (3) type 6 (dark). Skin tone was objectively measured from the participants' cheek skin using a RM200QC Spectro colorimeter (X-Rite, Grand Rapids, MI). Evidence suggests that darker skin tone is frequently under-represented in medical datasets,⁵¹ and that medical devices using optical sensors may be less accurate in those individuals.^{20–22} Therefore, the darkest skin-tone subgroup was intentionally oversampled to ensure the algorithm's unbiased performance over various skin tones. Informed consent was obtained from all study participants in accordance with the tenets of the Declaration of Helsinki. The study protocol was approved by Advarra IRB (Columbia, MD; protocol no. Pro00046845). The clinical research site followed standard safety precautions for COVID-19 in accordance with the Centers for Disease Control and Prevention guidelines.

Data collection

Each participant underwent four 30-second data collection episodes with their index finger (of a hand of their choice) held directly over the study phone camera. Three of the 30-second episodes were collected at rest under various ambient brightness/lighting conditions: (1) with camera flash on and under normal ambient light, (2) with flash off and under normal ambient light, and (3) with flash off and under dim light. The fourth episode was collected post-exercise. In the original protocol, participants were instructed to ride a stationary bicycle for 30 seconds as strenuously as possible against light to medium resistance. After enrolling 37 participants, the exercise protocol was modified (with an IRB amendment) to achieve higher participant HR: participants were encouraged to achieve 75% of their maximal HR, which was calculated by subtracting the participant's age from 220 beats/min. Exercise was completed either when the goal HR was achieved or when the participant asked to stop. The data were collected with flash off and under normal ambient light. Lighting conditions were controlled using 2 overhead and 1 front light

emitting diode (LED) lights. The brightness level of the study environment was measured by a Lux meter (LT300 Light Meter, Extech, Nashua, NH) prior to each study. Measured brightness values were between 160 and 200 Lux for normal ambient light, and between 95 and 110 Lux for dim light.

The study was conducted using a mobile app deployed to a Pixel 3 smartphone running Android 10 (Google LLC, Mountain View, CA). HR estimation using the app was generally completed by the study participants following the in-app instructions, with the coordinators providing feedback on usage when needed. The reference HR was measured simultaneously during each data collection episode using a Masimo MightySat® (Masimo, Irvine, CA), which is US Food and Drug Administration-cleared for fingertip measurement of pulse rate.⁵² The measurements were conducted in accordance with the manufacturer's manual and taken at the end of each episode.

Statistical analysis

Each participant contributed up to 3 HR measurements at rest (with different lighting conditions), and up to 1 post-exercise. Measurements were paired observations: the algorithm-estimated HR and the reference HR from the pulse oximeter. For each algorithm measurement, up to three tries were allowed, and the number of tries required was recorded. The baseline characteristics of the participants in whom a valid measurement for HR could not be obtained were compared to those of participants for whom a valid measurement was obtained using Fisher's exact test. A paired measurement was dropped if either the algorithm estimation or reference measurement failed. The absolute error of each paired measurement was calculated as the absolute value of the difference between the algorithm-estimated and reference HR values. The MAE was the mean value of all absolute errors. Similarly, the absolute error from each paired measurement was divided by the reference value for that measurement and multiplied by 100 to produce the absolute percentage error. The MAPE was the mean value for all absolute percent error values. The standard deviation of MAPE was calculated; no adjustment for multiple observations was made since the effects of clustering were negligible.

The MAPE was the primary study outcome, as recommended by the current standards for HR monitoring devices.²⁵ We also computed the standard deviation and 95th percentiles. Sign tests were used to determine whether the absolute percentage errors were significantly <5%, both for the entire group of participants and the 3 skin-tone subgroups; data from individual data windows were analyzed separately. Bland-Altman plots were used to visualize the agreement between the estimated values and the reference measurements and assess for any proportional bias (trends in the error with increasing values).⁵³ The subgroup analysis across the three skin-tone subgroups was pre-specified.

Sample size calculation

HR data collection was planned for approximately 100 participants. Enrollment up to a maximum of 150 participants was allowed as we anticipated that some enrolled participants would be excluded prior to

contributing HR data because they failed to meet the required skin tone distribution or because they were not able to exercise. Requirements for participant enrollment termination included ≥ 60 paired HR measurements in the dark skin tone subgroup and $\geq 20\%$ of the post-exercise reference HR > 100 beats/min. The study hypothesis was that MAPE was $< 5\%$ in all of the 3 skin-tone subgroups. To estimate the sample size required for the study, we first conducted an IRB-approved feasibility study with a different set of 55 participants and similar measurements both at rest and post-exercise. In that study, the MAPE \pm standard deviation was $0.91\% \pm 3.68\%$. Assuming double the mean and SD (i.e., 1.82% and 7.36% , respectively), a minimum of two paired measurements per participant, a skin-tone subgroup of ~ 25 participants, and some dropout from incomplete data, the power to detect a MAPE $> 5\%$ was > 0.8 .

RR measurement

Algorithm description

Prior work in computer vision and sensors to extract RR from RGB video signals relied on changes in color intensities at specific anatomical points,^{54,55} tracking head motions,^{56,57} estimating optical flow along image gradients,⁵⁸ or factorizing the vertical motion matrix.⁵⁹

Our contactless method estimates RR by performing motion analysis in a ROI of the video stream that includes the base of the neck, shoulder line and upper torso of the participant. The main challenge was that variations in video due to respiratory motions are hard to distinguish from noise. We build on Eulerian, phase-based motion processing⁶⁰ that is particularly suited for analyzing subtle motions. In each video frame, the position at each pixel was represented by the phase of spatially localized sinusoids in multiple scales (frequencies). To aggregate the information across scales and to obtain an intuitive representation of motion, we then transformed the spatial phases into optical flow by linearly approximating the position implied by each phase coefficient and averaging across scales. Using the Halide high-performance image library⁶¹, we were able to speed up the phase and optical flow computation to achieve real-time processing (1-4 ms per frame on Pixel 3a and Pixel 4 mobile devices).

Ensembling was then used to improve the predictive performance. A spectral-spatial ensemble was built in the following way. The respiratory ROI, together with the four quadrants obtained by equally subdividing the ROI defined five regions over which the vertical component of the optical flow was averaged. This resulted in five respiratory waveforms. Next, frequency-domain representations for each of these respiratory waveforms were computed via FFT, from which power spectra were computed. The power spectra were then aggregated to obtain a final ensembled power spectrum. Bandpass-filtering was performed to remove low and high frequencies unlikely to represent valid RRs. Filter cut-off frequencies corresponded to a low of 6 breaths/min and a high of 60 breaths/min. The maximum power frequency and the corresponding SNR value were computed from the ensembled power spectrum. The waveform corresponding to the entire ROI is used for displaying the breathing pattern to the user in the mobile app.

Often there was insufficient periodicity in the respiratory waveform (e.g., the participant briefly held their breath or changed their respiratory rate within the time window used for analyzing the waveform). To increase the robustness of RR estimation, the algorithm falls back on a time domain estimation method based on counting zero crossings of the waveform corresponding to the entire ROI whenever the SNR obtained via the FFT-based method was lower than a certain threshold. We tested two versions of the algorithm, differing only in terms of this threshold: SNR < -6.0 dB (“version A”) and SNR < -4.0 dB (“version B”). The higher value for the threshold in version B invoked the time domain estimation method more often, which was hypothesized to improve accuracy by improving robustness to irregular breathing.

Study design and participants

We performed a prospective observational clinical validation study to assess the accuracy of the study algorithm in measuring the RR in healthy adults and patients with chronic respiratory conditions (Supplementary Figure 1B). Participants were enrolled at a clinical research site (Artemis, San Diego, CA) between June 2020 and July 2020. Chronic respiratory conditions included moderate or severe COPD and asthma that was not well-controlled based on specific study criteria (Supplementary Table 1B). Also, participants with significant tremor were excluded. Further details and criteria are presented in Supplementary Table 1B. Informed consent was obtained from all study participants in accordance with the tenets of the Declaration of Helsinki. The study protocol was approved by Aspire IRB (now WCG IRB, Puyallup, WA; protocol no. 20201594). The clinical research site followed standard safety precautions for COVID-19 in accordance with the Centers for Disease Control and Prevention guidelines.

Data collection

Each participant underwent 30 seconds of data collection using a Pixel 4 smartphone running Android 10 (Google LLC, Mountain View, CA). The two algorithm versions (A and B) were tested sequentially. The participants followed the study protocol via instructions from the study app, without intervention from the study staff. Participants were prompted to prop the study phone on a table using provided common household items, such that the upper body was centered in the video capture (Figure 1). There were no specific requirements on the type of clothing worn during the study or additional custom lighting equipment. The in-app instructions guided the participants to wait several minutes after any active movement and to stay comfortable and breathe normally during the measurements.

During the data collection, RR was manually counted and recorded by two research coordinators. The two observers counted the number of breaths independently and blinded to the algorithm-estimated results. The agreement between the two measurements was high (Pearson correlation coefficient: 0.962; mean difference: 0.48 ± 0.88 breaths/min; range, 0–4). The mean of the two human-measured RRs, rounded off to the nearest integer, was taken to be the reference RR.

Statistical analysis and sample size calculations

Each participant contributed a single pair of measurements for each algorithm version, and the MAE was used as the primary evaluation metric. The study hypothesis was that MAE would be < 3 breaths/min. One-sample t-tests were done to determine whether the MAE was statistically significantly < 3 breaths/min. A pre-specified subgroup analysis was also performed, stratified by history of chronic respiratory conditions. In addition, *post-hoc* subgroup analyses were performed for age and race/ethnicity subgroups. Bland-Altman plots were used to analyze further for any trends in errors. Differences between the two algorithm versions were compared using a paired t-test.

To estimate the sample size required for the study, we first conducted an IRB-approved feasibility study with 80 healthy adults. Based on that MAE \pm standard deviation (0.96 ± 0.72 breaths/min), a sample size of 50 participants was estimated to provide a power of > 0.99 to detect an MAE < 3. The power was also >0.99 for both the subgroup of 10 healthy participants and the subgroup of 40 with chronic respiratory conditions. If the MAE and standard deviation were doubled, the power would be >0.99, 0.71, and >0.99, respectively, for the full sample, healthy participants, and those with chronic respiratory conditions.

User experience survey

The participants were surveyed about their experience using the app. The questions covered their ease of setting up the phone at the desired angle to capture their face/torso; the clarity of the instructions; their comfort in using the app to assess their general wellness; their comfort in teaching someone else how to use the app; and their expected comfort in using the app several times a day (Supplementary Table 7).

Table 1. Baseline characteristics of the study participants.

	Heart rate (HR) study	Respiratory rate (RR) study
No. participants analyzed	95	50
No. recordings	352	50 (for each algorithm version*)
Age (mean \pm standard deviation)	41.8 \pm 15.0	50.0 \pm 16.0
Age groups		
< 40 years	41 (43%)	17 (34%)
40–59 years	39 (41%)	21 (42%)
\geq 60 years	15 (16%)	12 (24%)
No. female (%)	71 (75%)	26 (52%)
No. male (%)	24 (25%)	24 (48%)
Race/ethnicity: n, %		
White, non-Hispanic	25 (26%)	18 (36%)
White, Hispanic	0 (0%)	22 (44%)
Black, non-Hispanic	61 (64%)	6 (12%)
Black, Hispanic	0 (0%)	1 (2%)
Asian / pacific islander	7 (7%)	3 (6%)
Multiple races, non-Hispanic	1 (1%)	0 (0%)
Multiple races, Hispanic	1 (1%)	0 (0%)
Measured skin tone**: n (%)		
1 (Fitzpatrick types 1-3)	31 (33%)	N/A
2 (Fitzpatrick types 4-5)	32 (34%)	
3 (Fitzpatrick type 6)	32 (34%)	
Chronic respiratory conditions: n (%)		
None		10 (20%)
Asthma	N/A	33 (66%)
COPD		4 (8%)
Both		3 (6%)

*RR was measured twice, once for each one of two algorithm versions (see Methods).

**Measurements were done on the cheek using a Pantone RM200QC Spectro (see Methods).

Abbreviations: COPD=chronic obstructive pulmonary disease.

Declarations

Data availability

The data are not publicly available because they contain videos of the participants and public release was not part of informed consent.

Acknowledgements

This work would not have been possible without funding by Google Health and the contribution of many collaborators. We would like to thank Michael Righter and Shilpa Mydur for input on regulatory considerations, Xia Bellavia and Neil Smith for leading the mobile apps testing, Joe Nagle and Hyun Ji Bae for user interface / user experience (UI/UX) designs, Bhavna Daryani for data quality control in the heart rate validation study, Matt Shore for early stage product development, Rajroshan Sawhney for business development expertise, Benny Ayalew for help in data infrastructure, Justin Tansuwan and Rebecca Davies for visualization tools, Jassi Pannu, Tiffany Kang, and Jacinta Leyden for background research, and Robert Harle and Kapil Parakh for heart rate study protocol feedback. We are also grateful for the valuable manuscript and research feedback from Tiffany Guo, Jacqueline Shreibati, Ronnachai (Tiam) Jaroensri, Jameson Rogers, and Michael Howell. Lastly, we would like to thank Mark Malhotra, Bob MacDonald, Katherine Chou and (again) Shilpa Mydur and Michael Howell for their support of this project.

Competing interests

All authors have completed the ICMJE uniform disclosure form at www.icmje.org/coi_disclosure.pdf and declare: all authors are employees of Google LLC except M.Kashyap and J.C. who are paid consultant of Google via Advanced Clinical; Google LLC may commercialize this technology or have patent rights now or in the future related to this paper. All the authors except M.Kashyap, S.K., and J.C. own Alphabet stock. M.V.M. is a member of the American Heart Assoc Health Tech Advisory Group and Consumer Technology Assoc Health, Fitness, and Wellness standards committees. M.J.P. is a member of the board of El Camino Health, advisory committee of ONC Interoperability Task Force, and scientific advisory board of National Library of Medicine, NIH, the board of AcademyHealth, advisory board of Johns Hopkins University and BME Department, and a advisor, and program committee chair of IEEE EMBS Society.

Author contributions

Y.E. and S.Bae developed the HR algorithms; S.Borac, J.Wu., N.W., M.Krainin, and M.R., developed the RR algorithms; J.Wang. and J.Wu, developed the mobile apps with the user interface for the clinical studies; J.Wang. developed software infrastructure for algorithms research; S.Bae. and S.Borac. developed software infrastructure for the HR study. E.S.T. provided data collection support for early-stage research; J.A.T., M.J.P, M.Kashyap, and J.C. developed the clinical study protocols; A.M. provided operational support for the clinical validation studies and managed inter-institutional communications; J.A.T. performed statistical analysis for RR study; J.A.T., S.Bae, and E.S.T. performed statistical analysis for the HR study; L.C. implemented the RR study and managed inter-institutional communications; L.C., M.Kashyap, and A.C. implemented the HR study and developed protocol optimizations; M.M. managed contracting and logistics required to begin the clinical validation studies; M.M. and A.C. performed quality control for the HR study data. S.K. and Y.L. provided research feedback and wrote the manuscript with the assistance and feedback of all authors. M.V.M. provided feedback on the HR study protocol and contributed to interpretation of the data and revisions to the manuscript; J.Z. oversaw technical developments and contributed to algorithm development; M.J.P. was the principal investigator for the clinical studies and developed the roadmaps encompassing the studies; S.P. and G.S.C. provided strategic guidance and funding for the project. All authors reviewed and approved the submitted manuscript.

Code availability

The algorithms were released as a feature in the Google Fit app to support widespread consumer use as of March 2021. Users and researchers will be able to directly access the feature via the app, on supported phones.

References

1. Tuckson, R. V., Edmunds, M. & Hodgkins, M. L. Telehealth. *N. Engl. J. Med.* 377, 1585–1592 (2017).
2. Benziger, C. P., Huffman, M. D., Sweis, R. N. & Stone, N. J. The Telehealth Ten: A Guide for a Patient-Assisted Virtual Physical Examination. *Am. J. Med.* 134, 48–51 (2021).
3. Hyman, P. The Disappearance of the Primary Care Physical Examination-Losing Touch. *JAMA Intern. Med.* (2020) doi:10.1001/jamainternmed.2020.3546.
4. Hollander, J. E. & Carr, B. G. Virtually Perfect? Telemedicine for Covid-19. *N. Engl. J. Med.* 382, 1679–1681 (2020).
5. Cutler, D. M., Nikpay, S. & Huckman, R. S. The Business of Medicine in the Era of COVID-19. *JAMA* 323, 2003–2004 (2020).

6. Rowland, S. P., Fitzgerald, J. E., Holme, T., Powell, J. & McGregor, A. What is the clinical value of mHealth for patients? *NPJ Digit Med* 3, 4 (2020).
7. Contributors to Wikimedia projects. List of countries by smartphone penetration. https://en.wikipedia.org/wiki/List_of_countries_by_smartphone_penetration (2014).
8. Shao, D., Liu, C. & Tsow, F. Noncontact Physiological Measurement Using a Camera: A Technical Review and Future Directions. *ACS Sensors* (2020) doi:10.1021/acssensors.0c02042.
9. De Ridder, B., Van Rompaey, B., Kampen, J. K., Haine, S. & Dilles, T. Smartphone Apps Using Photoplethysmography for Heart Rate Monitoring: Meta-Analysis. *JMIR Cardio* 2, e4 (2018).
10. Sun, Y. & Thakor, N. Photoplethysmography Revisited: From Contact to Noncontact, From Point to Imaging. *IEEE Trans. Biomed. Eng.* 63, 463–477 (2016).
11. Allen, J. Photoplethysmography and its application in clinical physiological measurement. *Physiol. Meas.* 28, R1–39 (2007).
12. Poh, M.-Z., McDuff, D. J. & Picard, R. W. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Trans. Biomed. Eng.* 58, 7–11 (2011).
13. Koetsier, J. The Top 10 Health & Fitness Apps Of 2020 Have One Thing In Common (Mostly). *Forbes* <https://www.forbes.com/sites/johnkoetsier/2020/10/05/the-top-10-health-fitness-apps-of-2020-have-one-thing-in-common-mostly/> (2020).
14. Hertzman, A. B. The Blood Supply of Various Skin Areas as Estimated by the Photoelectric Plethysmograph. (1938).
15. Coppetti, T. et al. Accuracy of smartphone apps for heart rate measurement. *Eur. J. Prev. Cardiol.* 24, 1287–1293 (2017).
16. American National Standards Institute. & Association for the Advancement of Medical Instrumentation. Cardiac monitors, heart rate meters, and alarms. (Arlington, Va. : Association for the Advancement of Medical Instrumentation, ©2002., 2002).
17. CONSUMER TECHNOLOGY ASSOCIATION. Physical Activity Monitoring for Heart Rate (ANSI/CTA-2065). <https://shop.cta.tech/products/physical-activity-monitoring-for-heart-rate>.
18. Bent, B., Goldstein, B. A., Kibbe, W. A. & Dunn, J. P. Investigating sources of inaccuracy in wearable optical heart rate sensors. *NPJ Digit Med* 3, 18 (2020).
19. Shcherbina, A. et al. Accuracy in Wrist-Worn, Sensor-Based Measurements of Heart Rate and Energy Expenditure in a Diverse Cohort. *J Pers Med* 7, (2017).

20. Bickler, P. E., Feiner, J. R. & Severinghaus, J. W. Effects of skin pigmentation on pulse oximeter accuracy at low saturation. *Anesthesiology* 102, 715–719 (2005).
21. Sjoding, M. W., Dickson, R. P., Iwashyna, T. J., Gay, S. E. & Valley, T. S. Racial Bias in Pulse Oximetry Measurement. *N. Engl. J. Med.* 383, 2477–2478 (2020).
22. Ries, A. L., Prewitt, L. M. & Johnson, J. J. Skin color and ear oximetry. *Chest* 96, 287–290 (1989).
23. Fallow, B. A., Tarumi, T. & Tanaka, H. Influence of skin type and wavelength on light wave reflectance. *J. Clin. Monit. Comput.* 27, 313–317 (2013).
24. Nam, Y., Lee, J. & Chon, K. H. Respiratory rate estimation from the built-in cameras of smartphones and tablets. *Ann. Biomed. Eng.* 42, 885–898 (2014).
25. Consumer Technology Association: CTA Committees, Subcommittees and Working Groups. <https://standards.cta.tech/kwspub/home/Committees/>.
26. U.S. Food & Drug Administration. 510(k) Premarket Notification: Philips Biosensor BX100. <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K192875>.
27. U.S. Food and Drug Administration. The C100 Contactless Breathing Monitor: 510(k) Premarket Notification. <https://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfpmn/pmn.cfm?ID=K200445>.
28. MightySat™ Rx Fingertip Pulse Oximeter. https://techdocs.masimo.com/globalassets/techdocs/pdf/lab-10169a_master.pdf.
29. Li, T. et al. A pilot study of respiratory rate derived from a wearable biosensor compared with capnography in emergency department patients. *Open Access Emerg. Med.* 11, 103–108 (2019).
30. Jensen, M. T., Suadicani, P., Hein, H. O. & Gyntelberg, F. Elevated resting heart rate, physical fitness and all-cause mortality: a 16-year follow-up in the Copenhagen Male Study. *Heart* 99, 882–887 (2013).
31. Cole, C. R., Blackstone, E. H., Pashkow, F. J., Snader, C. E. & Lauer, M. S. Heart-rate recovery immediately after exercise as a predictor of mortality. *N. Engl. J. Med.* 341, 1351–1357 (1999).
32. Flores Mateo, G., Granado-Font, E., Ferré-Grau, C. & Montaña-Carreras, X. Mobile Phone Apps to Promote Weight Loss and Increase Physical Activity: A Systematic Review and Meta-Analysis. *J. Med. Internet Res.* 17, e253 (2015).
33. Wu, Y. et al. Mobile App-Based Interventions to Support Diabetes Self-Management: A Systematic Review of Randomized Controlled Trials to Identify Functions Associated with Glycemic Efficacy. *JMIR Mhealth Uhealth* 5, e35 (2017).
34. Cadmus-Bertram, L. A., Marcus, B. H., Patterson, R. E., Parker, B. A. & Morey, B. L. Randomized Trial of a Fitbit-Based Physical Activity Intervention for Women. *Am. J. Prev. Med.* 49, 414–418 (2015).

35. U.S. Department of Health and Human Services. Physical Activity Guidelines for Americans, 2nd edition. https://health.gov/sites/default/files/2019-09/Physical_Activity_Guidelines_2nd_edition.pdf.
36. Centers for Disease Control and Prevention. Target Heart Rate and Estimated Maximum Heart Rate. Centers for Disease Control and Prevention <https://www.cdc.gov/physicalactivity/basics/measuring/hearttrate.htm> (2020).
37. Arnett, D. K. et al. 2019 ACC/AHA Guideline on the Primary Prevention of Cardiovascular Disease: A Report of the American College of Cardiology/American Heart Association Task Force on Clinical Practice Guidelines. *Circulation* 140, e596–e646 (2019).
38. Webster, D. E. et al. Heart Snapshot: a broadly validated smartphone measure of VO₂max for collection of real world data. *bioRxiv* (2020) doi:10.1101/2020.07.02.185314.
39. Reilly, B. M. Physical examination in the care of medical inpatients: an observational study. *Lancet* 362, 1100–1105 (2003).
40. Verghese, A., Charlton, B., Kassirer, J. P., Ramsey, M. & Ioannidis, J. P. A. Inadequacies of Physical Examination as a Cause of Medical Errors and Adverse Events: A Collection of Vignettes. *Am. J. Med.* 128, 1322–4.e3 (2015).
41. Bickley, L. S. *Bates' Guide to Physical Examination and History Taking*. (Lippincott Raven, 2007).
42. Bickley, L. & Szilagyi, P. G. *Bates' Guide to Physical Examination and History-Taking*. (Lippincott Williams & Wilkins, 2012).
43. Nakajima, K., Tamura, T. & Miike, H. Monitoring of heart and respiratory rates by photoplethysmography using a digital filtering technique. *Med. Eng. Phys.* 18, 365–372 (1996).
44. Hill, A., Kelly, E., Horswill, M. S. & Watson, M. O. The effects of awareness and count duration on adult respiratory rate measurements: An experimental study. *J. Clin. Nurs.* 27, 546–554 (2018).
45. Chong, J. W., Esa, N., McManus, D. D. & Chon, K. H. Arrhythmia discrimination using a smart phone. *IEEE J Biomed Health Inform* 19, 815–824 (2015).
46. Zaman, R. et al. Novel Fingertip Image-Based Heart Rate Detection Methods for a Smartphone. *Sensors* 17, (2017).
47. Chatterjee, A. & Prinz, A. Image Analysis on Fingertip Video to Obtain PPG. *Biomedical and Pharmacology Journal* 11, 1811–1827 (2018).
48. Wang, W., den Brinker, A. C., Stuijk, S. & de Haan, G. Algorithmic Principles of Remote PPG. *IEEE Trans. Biomed. Eng.* 64, 1479–1491 (2017).

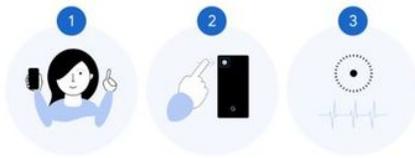
49. Karlen, W., Ansermino, J. M., Dumont, G. A. & Scheffer, C. Detection of the optimal region of interest for camera oximetry. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2013, 2263–2266 (2013).
50. Del Bino, S. & Bernerd, F. Variations in skin colour and the biological consequences of ultraviolet radiation exposure. *Br. J. Dermatol.* 169 Suppl 3, 33–40 (2013).
51. Kinyanjui, N. M. et al. Estimating Skin Tone and Effects on Classification Performance in Dermatology Datasets. (2019).
52. Bohnhorst, B., Peter, C. S. & Poets, C. F. Pulse oximeters' reliability in detecting hypoxemia and bradycardia: comparison between a conventional and two new generation oximeters. *Crit. Care Med.* 28, 1565–1568 (2000).
53. Altman, D. G. & Bland, J. M. Measurement in Medicine: The Analysis of Method Comparison Studies. *The Statistician* vol. 32 307 (1983).
54. Jorge, J. et al. Non-Contact Monitoring of Respiration in the Neonatal Intensive Care Unit. 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017) (2017) doi:10.1109/fg.2017.44.
55. Massaroni, C., Lopes, D. S., Lo Presti, D., Schena, E. & Silvestri, S. Contactless Monitoring of Breathing Patterns and Respiratory Rate at the Pit of the Neck: A Single Camera Approach. *Journal of Sensors* vol. 2018 1–13 (2018).
56. Al-Naji, A. & Chahl, J. Simultaneous Tracking of Cardiorespiratory Signals for Multiple Persons Using a Machine Vision System With Noise Artifact Removal. *IEEE J Transl Eng Health Med* 5, 1900510 (2017).
57. Yang, Q. et al. HealCam: Energy-efficient and privacy-preserving human vital cycles monitoring on camera-enabled smart devices. *Computer Networks* vol. 138 192–200 (2018).
58. Chatterjee, A., Prathosh, A. P. & Praveena, P. Real-time respiration rate measurement from thoracoabdominal movement with a consumer grade camera. *Conf. Proc. IEEE Eng. Med. Biol. Soc.* 2016, 2708–2711 (2016).
59. Janssen, R., Wang, W., Moço, A. & de Haan, G. Video-based respiration monitoring with automatic region of interest detection. *Physiol. Meas.* 37, 100–114 (2016).
60. Wadhwa, N., Rubinstein, M., Durand, F. & Freeman, W. T. Riesz pyramids for fast phase-based video magnification. 2014 IEEE International Conference on Computational Photography (ICCP) (2014) doi:10.1109/iccp.2014.6831820.
61. Ragan-Kelley, J. et al. Halide: a language and compiler for optimizing parallelism, locality, and recomputation in image processing pipelines. in *Proceedings of the 34th ACM SIGPLAN Conference on*

Figures

A. Smartphone-based heart rate (HR) measurement

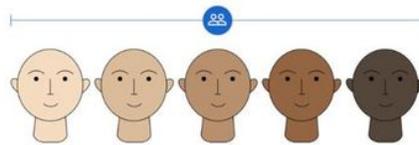
Setup

User places finger over smartphone camera



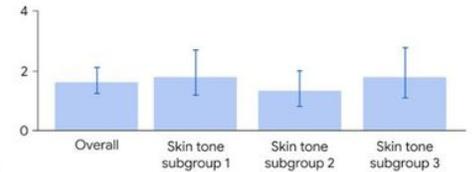
Study design

Different skin tones



MAPE (mean absolute percent error)

HR MAPE (%) with 95% confidence interval



B. Smartphone-based respiratory rate (RR) measurement

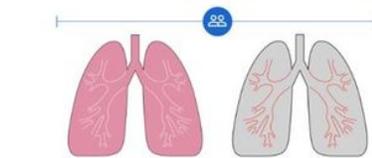
Setup

User faces a smartphone



Study design

With and without chronic respiratory conditions



MAE (mean absolute error)

RR MAE (breaths/min) with 95% confidence interval

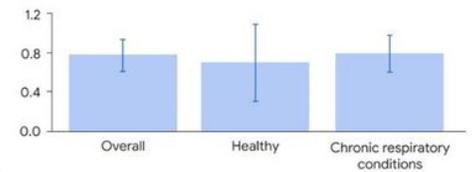


Figure 1

Smartphone-based monitoring of two key vital signs: (A) heart rate (HR) and (B) respiratory rate (RR). Left: the “setup” of how measurements are taken – with the finger over the rear-facing camera for HR, and using a video of the participant via the front-facing camera for RR. Middle: the study was designed to ensure generalization across skin tones for HR, and to participants with chronic respiratory conditions (chronic obstructive pulmonary disease and asthma) for RR. Skin-tone subgroup 1 corresponds to Fitzpatrick skin types 1–3 (very light, light and intermediate); subgroup 2 corresponds to types 4–5 (tan and brown); and subgroup 3 corresponds to type 6 (dark). Right: the main measurements were mean absolute percent error (MAPE) for HR, and mean absolute error (MAE) for RR.

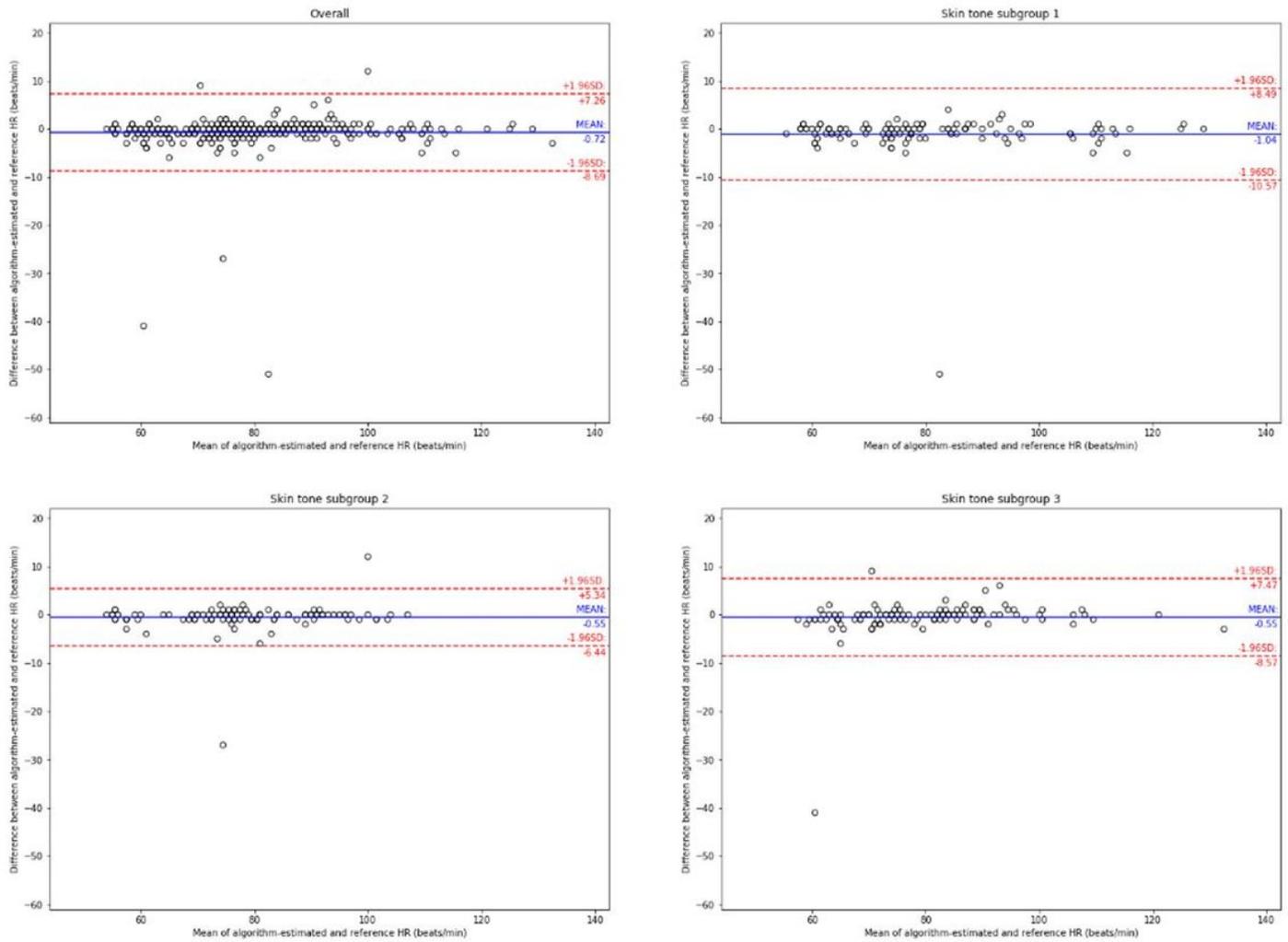


Figure 2

Bland-Altman plots for the heart rate (HR) study. Top to bottom and left to right: plots from the full study followed by subgroups based on skin type (see Figure 1). The reference HR was obtained from a pulse oximeter (see Methods). Dots represent individual participants; blue lines indicate the mean difference; red lines indicate the 95% limits of agreement (mean difference \pm 1.96 standard deviations).

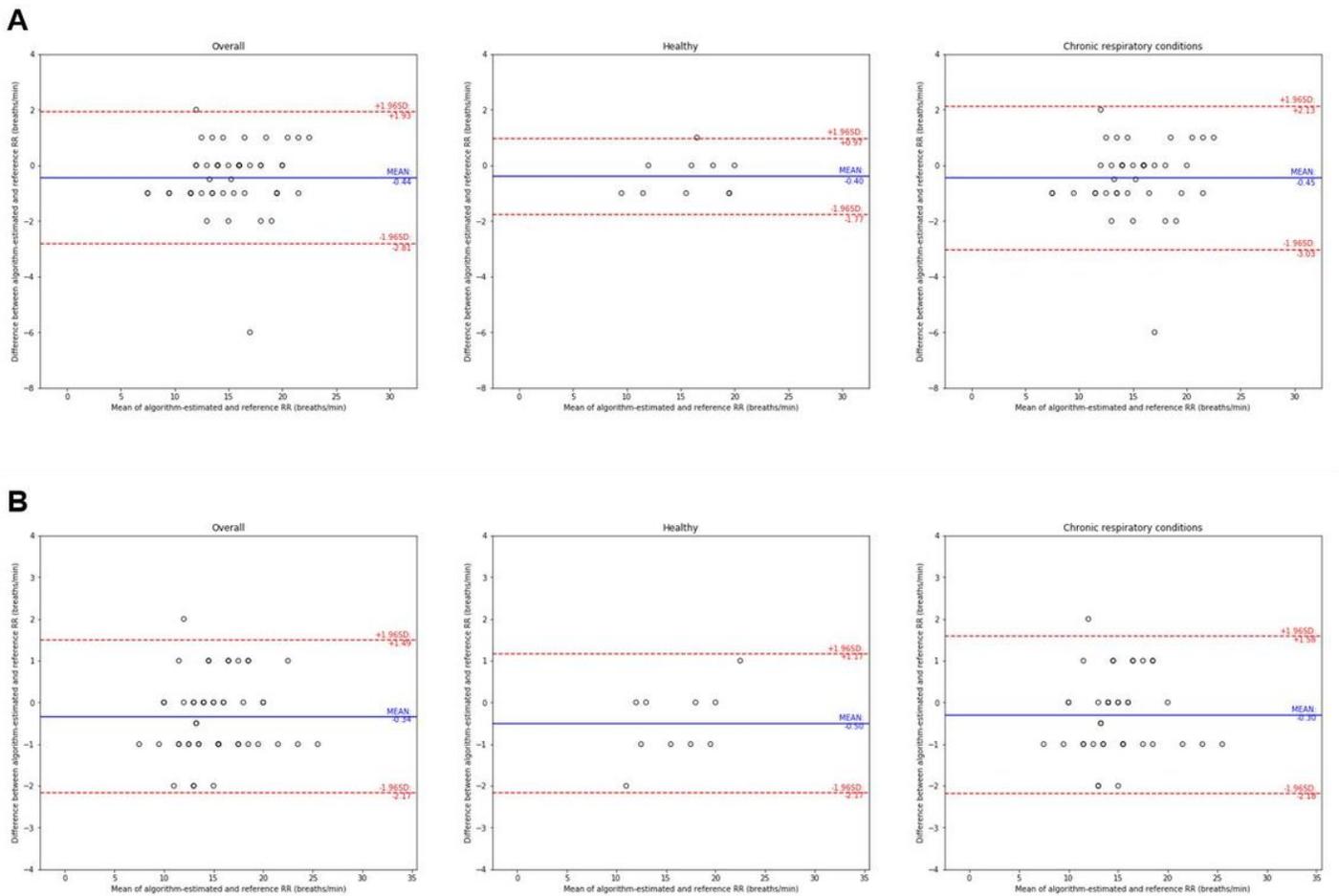


Figure 3

Bland-Altman plots for the respiratory rate (RR) study, for (A) algorithm version A and (B) algorithm version B. Left to right: plots from the full study followed by subgroups based on presence vs. absence of chronic respiratory conditions (see Figure 1). The reference RR was obtained by research coordinators manually counting breaths (see Methods). Dots represent individual participants; blue lines indicate the mean difference; red lines indicate the 95% limits of agreement (mean difference \pm 1.96 standard deviations).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [4Supp.docx](#)