Supplemental Material

Data S1

Supplemental Methods

Multimodal hardware design

Overall, the key upgrades from the hardware used in [19] include the addition of a flexible connector, two main sensing printed circuit boards as opposed to three, eventually the use of gel-electrode ECG, a separate photoplethysmogram sensor board with newer discretized photodiodes and light-emitting-diodes, and a foam-based spring backing mechanism for improved photoplethysmogram sensing. However, the photoplethysmogram signals were not explored in this work and therefore—to prevent detracting from the focus of this work—the specific details of that hardware will not be expanded on further. The sample rate of the ECG was 1kHz and the SCG either 500Hz or 2kHz depending on the prototype version. Specifically, in the newer version, shown in Figure 2a, the sample rate of the SCG was increased to 2 kHz to provide a bandwidth of 500 Hz; the SCG sampling frequency was adjusted to capture higher frequency sounds that may eventually be utilized to monitor patients with heart murmurs—a subsection of the CHD population at a greater risk of decline. Unfortunately, as with a proof-of-concept study, the hardware required few—mostly device housing—modifications at different stages of the study before reaching the current prototype pictured in Figure 2a. Most importantly, the earlier version of the hardware utilized in this study featured the use of a dry electrode ECG, using stainless steel tape, for which the device was pressed against the chest of the patient to acquire the biosignals, while the later version used standard infant AgCl gel electrodes (Kendall HP69, Medtronic PLC, Dublin, Ireland) to adhere to the chest, eliminating the need for an extra contact force. To help mitigate any issues from differences in contact pressure with the dry electrode

version, in addition to having the same group of few clinicians collect all data, only segments of the signals where the dry electrode acquired ECG—which is susceptible to variations in contact pressure due to changes in skin-electrode-impedance[32]—had a consistent amplitude were analyzed. Devices with both versions of the ECG featured a firmware modification which leveraged the lead-on detect feature of the ECG chip and would toggle a light-emitting diode facing the clinician between red and green for when ECG lead was detected as off or on, respectively. This also removed the possibility of accidentally applying an excessive amount of pressure without knowing whether a signal was being acquired.

In future work, though high-fidelity wearable measurements were acquired and only few minutes of data collection were necessary for this study, the wearable biosensor still needs further miniaturization to be used in future longitudinal studies in a pediatric population. However, given the considerably smaller footprint of the internal essential sensing elements, the hardware could readily be miniaturized and exploit the advent of flexible electronics which can offer a low-profile, less obtrusive solution for even greater convenience when performing longitudinal monitoring[33].

Signal processing and feature extraction

All signal processing and feature extraction was carried out in MATLAB 2018a (MathWorks, Natick, Massachusetts, USA) and entirely automated. A high-frequency SCG signal more closely related to the phonocardiogram was extracted for this analysis. The phonocardiogram, typically acquired from digital stethoscopes, is a wide bandwidth, high-frequency acoustic signal that captures heart sounds (i.e., S_1 and S_2) and obtains information of valve closures when placed at specific auscultation sites. Although, the phonocardiogram should be acquired using a wide-bandwidth, piezoelectric accelerometer (i.e., a contact microphone)

rather than the capacitive, direct current micro-electro-mechanical systems accelerometer used herein, the sampling rate of the accelerometer was increased to provide this bandwidth. First, the R-peaks of the ECG—marking ventricular depolarization—were found using Pan-Tompkins's algorithm and used to determine the wearable HR. Then the SCG and high-frequency SCG signals were segmented into different heartbeats using and beginning with the detected R-peaks of the ECG. Due to the large differences in HR in this dataset, all of the heartbeats were zero-padded to a fixed length of 1300 ms, based on the slowest HR in the dataset. Next, the SCG and high-frequency SCG heartbeats were ensemble averaged using 30 heartbeat windows with 50% overlap—to reduce zero-mean noise, remove respiratory induced variability, and improve the consistency of amplitude features—before selecting the highest SNR beat—calculated using the algorithm in[34]. First the envelope of the high-frequency SCG was computed which provided the profiles for the conventional heart sounds S_1 and S_2. The algorithm for detecting the aortic opening point on the max SNR SCG beat was the same as that used in[20], where the aortic opening was detected by finding the nearest zero-crossing after the peak of the high-frequency SCG envelope between 0 and 150 ms; the aortic closing point was determined by finding the most consistent peak of the high-frequency SCG itself between 250 ms to the end of the beat. The aortic opening point resembles the PEP with the difference between that and the aortic closing point being the VET. Two other reciprocal features, PEP/VET and VET/PEP—systolic marker robust to differences in HR–are the quotient of the PEP and VET. Two interpretable systolic amplitude features were calculated as the RMS amplitude of the SCG during the PEP and during the VET. In total 9 systolic features were extracted from the wearable signals. Note that the HR from the CMR was added as a feature, due to both the inability to acquire continuous measurements with the wearable patch during the CMR—because of magnetic interference and injury— and due to expected high

accuracy in HR estimation when using wearable ECG during baseline measurements, as a closer measure of the HR during the reference measurement.

Leveraging surrounding physiological information can contextualize and improve the estimation accuracy of wearable measurements. The SV measurement from the CMR, is computed from a composition of several images which are obtained a relatively slow sampling rate. Therefore, due to respiratory induced variability in SV readings—stemming from changes in venous return, preload, and HR—clinicians typically ask patients to hold their breath. However, as imaginable, for younger children this is obviously not possible. Instead, multiple scans are taken, are the resulting images are averaged before computing SV from the averaged image. Similarly, when using wearable measurements to accurately estimate SV compared to CMR readings should also factor in respiratory variability by averaging over a larger timespan—such as the 30 heartbeats employed in this analysis.

Machine learning

All machine learning and cross validation was performed in Python 3.0 using scikit-learn ridge regression and grid search packages, respectively. Despite a considerable sample size with respect to other SCG literature—especially given the diversity of demographics and diagnoses in such a diseased population—due to a small number of overall datapoints for a machine learning problem, multi-variate ridge regression was chosen as a less complex linear, and more interpretable, model to estimate SV[35]. Ridge regression is similar to multiple-linear regression but with a regularization penalty—commonly referred to as lambda—that penalizes the model to prevent overfitting to the training data, thereby hopefully improving model generalizability [35].

10-fold cross validation was chosen as a commonly regarded robust method for optimizing hyperparameters and given that each subject had only one datapoint there would be

no overlap of subject-specific data in each fold. The completely randomized, held-out test set was determined by utilizing a true random number generator (RANDOM.ORG, Dublin, Ireland). To approximately balance and have a representative number of the number of single-ventricle patients in the training and testing set based on their size, originally during the data collection we again randomly split them into groups of four and three, respectively. However, after the final data was collected, a last single-ventricle patient was added to the training set to achieve a perfect 80%-20% split, hence a slight imbalance.

When selecting features for biomedical machine learning problem with a small dataset size there is a greater importance placed on not only selecting a few features that can explain a lot of the variance but also ones that can be clinically understood. Therefore, the original feature set of predictor variables consisted of 14 features that were chosen based on those with strong overlap between commonly used SCG features in existing literature and those that are simple and intuitive to cardiologists. Using forward feature selection on the training set, we decreased the feature set from 14 down to nine features based on the simple linear regression coefficient of determination between ECG and SCG features and SV. The Pearson's correlation coefficient values between these final nine features and SV, for the training set, are shown in Table S1.

We compared different ridge regression models trained on unique feature sets in our work. Specifically, we sought to compare models leveraging different combinations of ECG, SCG, and demographic (i.e., age and body surface area) features. Both training and testing set features were normalized based on the training set mean and standard deviation. The hyperparameter lambda for the highest performing model, which combined ECG and SCG features, came out to the maximum regularization penalty of 1.0.

Table S1. Training Set Correlation Coefficients Between Physiological Features and Stroke Volume.

Physiological Feature	Training Set Pearson's *r*
HR_{CMR}	-0.65
$HR_{Wearable}$	-0.63
VET	0.39
PEP	-0.02
PEP/VET	0.01
VET/PEP	-0.06
AC	0.45
RMS_{PEP}	-0.20
RMS_{VET}	-0.21

HR indicates heart rate; CMR, cardiac magnetic resonance imaging; VET, ventricular ejection time; PEP, pre-ejection period; AC, timing of aortic valve closure; RMS, root-mean-square power.

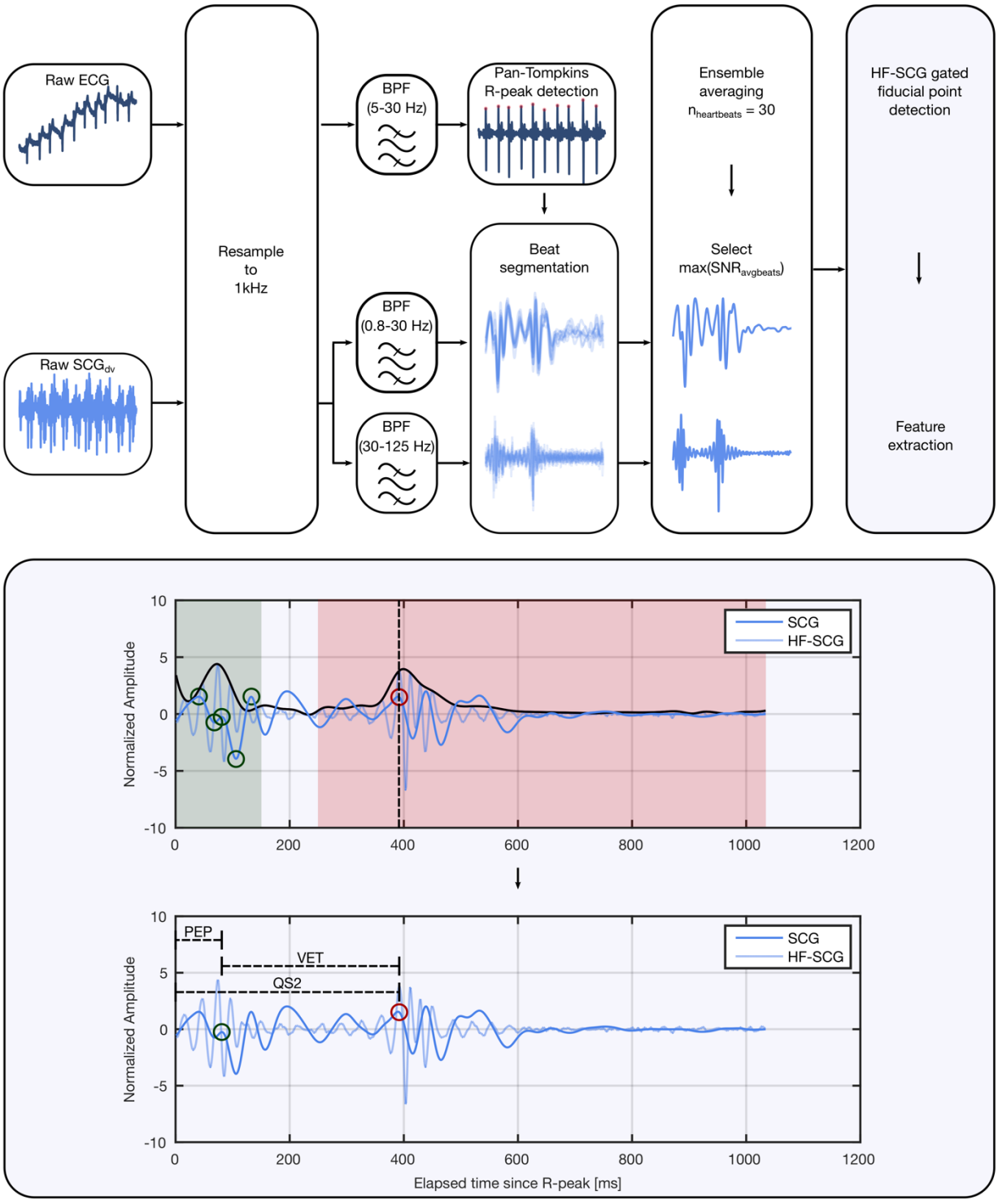

Figure S1. Signal processing pipeline. Block diagram of signal processing overview showing interpolation of electrocardiogram (ECG) and seismocardiogram (SCG) signals acquired from the wearable before bandpass filtering, R-peak detection, heartbeat windowing, and signal

quality assessment using the signal-to-noise ratio (SNR). Illustration of the custom high-frequency SCG (HF-SCG)—indicative of valve closures—assisted feature selection algorithm, helping to locate key fiducial points such as the aortic valve opening (AO) and aortic valve closure (AC) on the SCG—used to compute the pre-ejection period (PEP), ventricular ejection time (VET), and the AC. Additionally, the search radius for the AO (green) and AC (red) algorithm as well as their candidate points are shown.

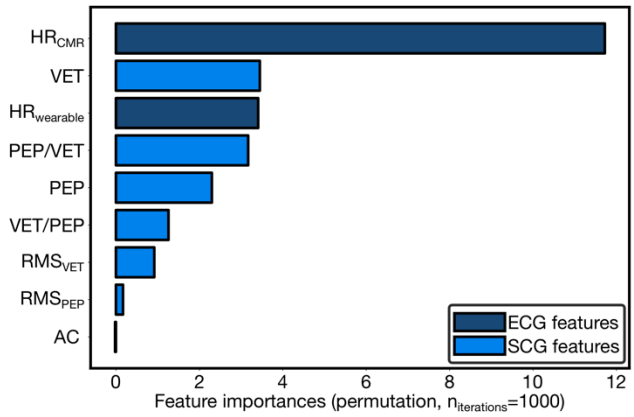

Figure S2. Permutation feature importances for stroke volume (SV) estimation model. Permutation feature importances for wearable system with features randomly shuffled 1000 times, ranked in order from top to bottom, and color-coded by wearable sensing modality—electrocardiogram (ECG) and seismocardiogram (SCG) signals.