Published in final edited form as:

Meas Phys Educ Exerc Sci. 2018 ; 22(4): 356–363. doi:10.1080/1091367X.2018.1476357.

Methods and Impact of Correcting Heart Rate Data Errors during Exercise in Older Adults

Kathleen K Mangione, PhD, PT¹, Kathy Macropol, PhD², Yanxia Jia, PhD², Michael Tevald, PhD, PT¹, Shane Harris, DPT¹, Edward Wolff, PhD², Rebecca Craik, PhD, PT³

¹Arcadia University, Department of Physical Therapy

²Arcadia University, Department of Computer Science and Mathematics

³Arcadia University, College of Health Sciences

Abstract

Heart rate (HR) by time curves could be useful as a measure of treatment fidelity. The purposes were to describe the frequency of common recording irregularities (e.g. errors) observed during exercise, validate a process to correct those errors, and determine whether there is a clinically meaningful benefit to data correction. 1895 exercise sessions from 69 older adults with hip fracture were inspected for errors in the placement of event markers (duration of exercise) and signal artifacts. Marker errors occurred in 33% of the participants, and corrected in 324 sessions. Signal artifacts occurred in 96% of the participants, and eliminated in 702 sessions. Computer code was 85% accurate in detecting signal artifacts, compared to 97% for investigators. HR was significantly higher after correction, but the average change was only 0.69 ± 1.20 beats per minute. A HR monitor showing HR by time curves can be used to evaluate treatment fidelity during exercise.

Keywords

Measurement; Assessment; cardiovascular; heart rate monitors; exercise

INTRODUCTION

Literature that describes exercise interventions during clinical trials is historically brief. Studies frequently lack key features including dose or intensity, mode of delivery, essential processes and monitoring (Glasziou, Meats, Heneghan, & Shepperd, 2008; Hoffmann et al., 2014). The lack of adequate reporting does not allow for replication by other investigators or translation to clinical practice. The Template for Intervention Description and Replication (TIDieR) was published in 2014 with the goal of improving completeness of exercise reporting. The TIDieR guidelines request the inclusion of items that detail “when and how much,” such as the number of times the intervention was delivered, the schedule, and the duration, intensity or dose (Hoffmann et al., 2014). In our opinion, it is equally important to

include measures of treatment fidelity (TF). Monitoring TF assures *internal validity*, that the intervention was delivered as intended; and *external validity*, that the intervention can be replicated and applied in real world settings (Bellg et al., 2004; Borrelli, 2011).

Adherence to specific instructions for delivery of the intervention is difficult to assess. The gold standard to assess TF is audio or videotapes, evaluated according to criteria developed *a priori* (Borrelli, 2011). There are cost, burden, and obtrusiveness issues that make this gold standard difficult to enact in exercise research. Alternate measures such as interventionist checklists and participant questionnaires have been used but are subject to recall bias and show low correlations with direct observation (Borrelli, 2011). Objective evaluations of treatment content and dose are needed.

Heart rate (HR) is an objective measure commonly reported in exercise research studies. HR is linearly related to exercise intensity, although important caveats apply. At low intensity levels, however, HR can be reflective of sympathetic activity such as caffeine, emotional state, and physiological stress (Strath et al., 2013). HR is also known to show a disproportional increase with upper extremity compared to lower extremity movement. In some cases, a time lag is demonstrated between HR and the onset and cessation of activity (Strath et al., 2013). Despite these issues, HR is typically used for the purpose of exercise prescription such as establishing training zones to improve sport performance or to improve endurance in clinical populations (Kemi & Wisloff, 2010; Scanlan et al., 2018). Fitness levels or health can be informed by HR indices such as HR variability, HR acceleration and HR recovery (Borresen & Lambert, 2008). These measures require examining data from the heart rate by time curves. With the advent of wrist worn HR receivers and wireless chest transmitters, continuous HR measurement during exercise sessions is feasible and practical for a variety of settings, including the home.

Continuous monitoring of HR may be useful as a measure of treatment fidelity as was reported by Taylor in a sample of healthy adolescents; and by Resnick in a sample of adults post stroke (Resnick et al., 2011; Taylor, Weston, & Batterham, 2015). The ease of obtaining both average HR and HR by time curves with commercial equipment may be useful in populations for whom there is limited information, such as older adults. Since older adults are encouraged to participate in multi-modal exercise; e.g. 20 minute bouts of moderate intensity physical activity, strength training, balance and flexibility exercises (Nelson et al., 2007); monitoring HR may provide information related to intensity of various modes of exercise, safety or issues with adherence.

Commercially available monitors have the advantage of being inexpensive, but have disadvantages that may make them unsuitable for research purposes. The integrity of the HR signal is affected by the contact of the electrodes with the skin. At times, the shape of the thorax, dry skin, or excessive movement causes the electrode to lose temporary contact with the skin and produces artifacts in HR recordings. In addition, weak signal transmission can cause the recorded value to remain steady until signal strength increases with abnormal values recorded due to interference from other equipment in the area or from cardiac arrhythmias typical for many older adult patients.

It is not known if the frequency and effects of these HR recording errors outweigh the benefits of using a simple, low cost measure that provides information on TF and physiological output. The purposes of this study, therefore, are to develop and validate a process to describe the frequency of common errors observed using a commercially available HR monitor during multi-modal exercise training in older adults, develop computer code to correct those errors, and to determine whether there is a clinically meaningful benefit to data correction via computer programming.

METHODS:

Sample & Description of Intervention

We used a subset of existing data from 69 participants from a multi-site trial testing a multi-modal exercise intervention for older adults with hip fracture (Orwig et al., 2017). Briefly, 210 participants (age > 65 years) who fell and had surgical repair of a hip fracture were recruited from hospitals to participate in an exercise program usual rehabilitation was completed (approximately 2-3 months post fracture). Major exclusions included dementia, not living in the community prior to fracture, walking more than 300 meters in 6 minutes, and recent stroke or major cardiovascular event. Participants were randomized to either a “specific” intervention of lower extremity strength training and endurance training or to a “non-specific” intervention of seated movements encompassing all joints of the body and lower extremity sensory level transcutaneous electrical nerve stimulations (TENS).

Each session was conducted in the participant’s home. For both intervention groups, phase 1 included measurement of vital signs to ensure safety to proceed with intervention and questions regarding adverse events since the prior visit (approximately 5-10 min). Phases 2 and 3 were the multi-modal active components, lasting approximately 20 minutes for each phase (40 minutes total). Phase 4 included a period of “cool down” and recovery vital signs. In the “specific” treatment group, phase 2 exercises included lower extremity strength training of the gluteal muscles, quadriceps, hamstrings, and gastrocnemius muscles performed in supine and standing positions. Phase 3 exercises included endurance training (indoor and/or outdoor walking were performed most frequently). For the “non-specific” treatment group, phase 2 exercises included seated activities involving all joints and planes of movement, while phase 3 was sensory level TENS while seated to bilateral gluteal, quadriceps, and gastrocnemius muscles.

Participants in the study wore Polar E600 monitors™ (Polar Electro Inc., Lake Success, NY) for each exercise session. The Polar brand monitors have among the highest reported validity with electrocardiography in comparison with other commercially available monitors (Terbizan DJ, 2002; Wang et al., 2017). The E600 monitor consists of a transmitter worn on the chest and a wrist receiver that records and stores the data. The receiver collected data continuously and recorded average HR every 15 seconds. Each exercise session lasted approximately 60 minutes and was comprised of 4 distinct phases. The phases of the intervention were indicated by event markers manually triggered by the interventionist.

Data Reduction and Definition of Errors

The data for this study consisted of the HR by time curves (HRxTC) for each session. (Figure 1) The average heart rates for active components were of interest to assess TF. Errors in the HRxTC were categorized as *marker errors* or *artifacts*. Marker errors consisted of missing or excessive event markers, which were used to divide the phases of the exercise session. Artifacts consisted of non-physiologic deflections in the HR curve, i.e., *signal dropping* to zero, *spikes* in HR greater than 5 SD of the mean in the absence of symptoms or distress indicated in the interventionist's notes from the session, or prolonged *flat lines* with no variation in HR. The frequency of both types of errors was determined and summarized using descriptive statistics.

Development of Coder for Error Identification and Correction

To develop computer code to identify marker errors, the data were first inspected visually for the number and placement of the markers, indicating the number and duration of the exercise phases. The average duration of the active phases for all sessions was calculated (26.1 ± 8.3 minutes). These observations were then used to determine the rules used in the computer code to identify and correct the marker errors (see Figure 2 for pseudocode). For example, HRxTCs with 3 markers (indicating 4 phases) and a duration of the active phase within 2 standard deviations of the mean were considered correct, while records with zero or 1 event marker, or a duration of the active component phases that was greater than or less than 2 SD from the mean duration were excluded. For records with incorrect numbers of markers (2 or more than 3 markers), code was developed to estimate marker placement.

To develop code to identify artifacts, the 35 randomly selected HRxTC were first inspected visually by the lead author (KM- considered the reference standard) and three exercise researchers not familiar with the study. The curves were selected from the pool of possible curves ($n=1295$ curves) by blindly selecting an individual participant and session. The lead author reviewed the selections to ensure that there were curves with and without artifacts. Consistency among the reviewers was evaluated by comparing the frequency of agreement on the presence of an artifact (true positives), frequency of agreement on the absence of artifacts (true negatives), and the frequency of disagreements (false positives and false negatives) between each of the reviewers and the reference standard. Sensitivity, specificity, and accuracy values were calculated for the identification of artifacts in each of the pairwise comparisons. The observations of the four reviewers were used to develop rules used in the computer code to identify and eliminate artifacts (see Figure 2 for pseudocode). If $> 25\%$ of the data points in the active component of the session were removed due to the presence of artifacts, that record was excluded. The validity of the rules to identify marker errors was established using a 3-fold cross validation technique (Saki M.J., 2014)

The final computer code implemented in Java was applied to all HRxTC (see Figure 3). To further validate the computer code's identification and correction of errors, the results of the computer analysis were compared to that of an independent researcher in a random sample of 38 records using the same process described above for comparing the reviewers to the reference standard.

Determining Impact of Error Correction

In our subset, data from two of three sites were used. To determine if the data from the two sites could be combined, the distributions of heart rates for the sites were compared using Kolmogorov-Smirnov tests. To determine the impact of the computer analysis on estimates of heart rate, we used paired t-tests to compare the average HR per subject (for the active phases) before cleaning artifacts and after cleaning artifacts to determine statistical differences and clinical meaningfulness. The Pearson correlation coefficient was used to determine if there was a relationship between HR and the number of artifacts present since the exercise groups had different types and intensities of exercise. We used IBM SPSS (24). Significance was determined at $\alpha = 0.05$.

RESULTS

The sample consisted of 1350 exercise sessions (48 participants) at site 1 and 545 sessions (21 participants) at site 2 (1895 sessions total). The distributions of the participants' HR means and standard deviations did not significantly differ between the two sites ($p=0.44$ and $p=0.59$ respectively) and the sites were combined into one data set for all subsequent analyses. There were 69 participants total, 34 of whom received the "specific" intervention and 35 received the "non-specific" intervention.

Marker errors were common in the sample with excessive event markers (> 3) being more common than missing event markers. (Figure 3) In cases in which there were no event markers or a single event marker, the data were excluded from further analyses (marker fail). For the cases in which there were 2 event markers, the third marker was imputed using computer code rules developed using data from the data set that had the 3 expected markers at the expected times. The computer "trained" on the data with 3 event markers and then "learned" where to impute the marker based on phase duration and expected pattern of activity. The program placed the marker within 2 minutes of where the user had originally marked it, which yielded an 81% accuracy rate. By applying the computer code to sessions with only 2 markers, we added an additional 73 sessions to the "marker pass" pool. For sessions that had extra event markers which occurred in quick succession of one another (< 1 minute), the second marker was deleted. Using this technique, we added 251 sessions to the "marker pass" pool. The largest group were sessions with 3 event markers; 12% of these sessions ($n= 176$) had phases that were further than 2 SD from the mean lap duration and were not subjected to additional analysis ("Marker Fail"). After using the computer code to correct the errors, approximately 26% of the total curves were excluded ("Marker Fail"), while the remaining 84% of the original sample was further analyzed for artifacts ("Marker Pass").

Artifacts occurred in 66 participants (96%) and across 39% of the sessions. We examined the type and distribution of artifacts. Drop in signal was the most frequent error and occurred between 1 and 10 times in 27.5% of the sessions, and greater than 10 times across 4.7% of the sessions. Spikes and flat lines occurred in 13.9% of the sessions. When examined by participant, 11.5% of the participants had no drops across sessions, 62% of the participants had between 1 and 50 drops in the signal across all the sessions, and the remaining participants experienced more than 51 (and up to > 300) drops across sessions. Spikes

occurred in 71% of the participants, while flat lines were the least frequent and occurred in 38% of the participants.

Compared to the reference standard, the accuracy of the 3 outside reviewers was 97%. Accuracy was 85.8% between the independent researcher and the computer generated rules. The process of cross validation in the sample did not identify gaps in the rules and the code was accepted.

The change in average HR for the active phases of intervention was 0.67 ± 1.20 beats per minute across the 69 participants who had artifacts corrected with the computer code. A paired t-test revealed that this change was statistically significant ($p < 0.001$). The average HR during the 2 active components was 80.1 ± 5 bpm. The correlation between heart rate and number of artifacts removed was not significant ($r = -0.14$; $p = 0.26$).

DISCUSSION

We identified two types of errors when examining the HR by time curves collected with a commercially available HR monitor in a large sample of older adults performing various modes of exercise. There were “marker errors” which influenced the duration of exercise phases and “signal artifacts” which influenced the average HR recorded for an exercise phase. These errors occurred in the majority of the sample, but correcting the errors, although statistically significant, produced an average change in HR of only 0.67 beats per minute. We have shown that the HR by time curves reported by the Polar™ monitor can be used with confidence as a measure of the delivery component of treatment fidelity. We also showed that relatively simple computer coding can be used to correct marker errors and thus increase the number of sessions included in analyses.

The participants in our study included a large sample of older adults recovering from a recent hip fracture. We examined heart rate responses for several modes of exercise, in transitions to different body positions (sitting, standing, and lying), and during both arm and leg activity. There are few studies that report HR during multimodal exercise in older adults. Campbell et al examined the cardiovascular responses to assess intensity of a single session of shallow water exercise in a sample of 11 women (mean age 67 yrs). They reported HR ranges from 92 to 124 bpm depending on the activity during the exercise bout (Campbell, D’Acquisto, D’Acquisto, & Cline, 2003). Puggaard et al investigated the impact of an 8 month exercise program on various indices of fitness and walking speed in a sample of healthy 85 year old women. They noted that average intensity during the training sessions was approximately 69% of maximum heart rate or approximately 87 beats per minute (Puggaard, Larsen, Stovring, & Jeune, 2000). The magnitude of HRs we recorded varied with the mode of exercise, with the extremities used, and with body position, but were as we would expect with a sample of deconditioned, older adults in their 80s. These values are obviously much less than in younger, or more fit individuals. None of the studies we found examined heart rate by time curves to determine if recording errors existed.

Marker errors occurred in 33% of the curves while artifacts appeared in 39% of the curves. The amount of data that could be misinterpreted or lost is large. Computer programing

allowed us to add 324 curves to the marker pass pool and to remove curves that had significant amounts of marker error or artifacts. However, while the correction for marker error in identifying the phases increased the number of sessions for analysis, the application of computer coding to remove the artifacts was not essential in our sample. The average change in HR per participant before and after correction was less than one beat per minute. The impact of artifact correction could be larger in smaller samples or during shorter bouts of recording. We would recommend that individual users of these types of monitors examine the heart rate by time curves to see if artifacts occur, and then determine if correction is warranted based on the frequency of errors. For use in research trials, we found that the HR monitor was a useful tool to estimate average heart rate during the active components of the intervention.

The cost and effort of monitoring HR during clinical exercise trials needs to be balanced with accuracy of the data. Wang et al recently reported that in a comparison of a Polar chest strap monitor to four wrist-worn optically based monitors (including the Fitbit and the Apple watch), the Polar device had the highest accuracy (Wang et al., 2017). Accurate heart rate measures are very important for older adults, especially for those with cardiac conditions for which HR needs to be monitored to keep the rates from exceeding safe values (Strath et al., 2013). Some older adults may use HR monitors to learn how to “pace themselves” to successfully and safely complete a 10 or 20 minute bout of continuous exercise. Resnick and colleagues reported that during supervised treadmill walking, only 48% of adults post stroke were able to achieve the goal of walking for 35 minutes at 60-70% of maximal heart rate (Resnick et al., 2011). Exercise staff or physical therapists working with clinical groups or frail, community-dwelling older adults would benefit from using heart rate monitors so that they can more accurately gauge and modify the intensity during the workout.

Choosing an accurate, but still cost-effective device would appear to benefit the persons engaged in exercise. For clinical research with older adults, the Polar device and use of a chest strap appears warranted since we showed that the recording errors did not impact the average heart rate values. As technology continues to advance, future research will need to test iterations of wrist worn devices or wearable devices in older adults and in less able-bodied samples who perform various modes of exercise.

The strengths of this study include the large sample of heart rate curves, the use of the monitor in four very different modes of exercise intervention, and the 2 step validation process of for correcting user errors associated with event markers placement and identifying artifacts. There are several limitations of the study. We used only one brand of commercially available HR monitors and our results may not generalize to all HR monitors. Our sample was comprised of non-fit older adults whose HRs were relatively low and the results may not apply to those who are younger or more fit with much higher heart rates. Data were obtained from only 2 of the 3 sites of the trial due to logistical issues, but the distributions in HR and standard deviations at 2 of the sites were not statistically different, which lessens our concern about site differences. Another limitation is that part of the research team was informed of the HR findings on a regular basis as part of monitoring the trial and, therefore, may have introduced bias into the coding procedures. We believe our

methods of validating the processes minimize this influence. We also recognize that event markers are not a necessary feature for all studies.

CONCLUSION

Exercise trials involving older adults can use the values reported by heart rate monitors as a means of monitoring the delivery of the exercise program. The artifacts that occur are not of sufficient frequency to make a meaningful impact on reported values.

Acknowledgements:

We acknowledge Jay Magaziner, PhD, Principal Investigator of the Community Ambulation Project (CAP) and the CAP clinical sites for support and data acquisition.

Funding Source: This research was supported by a grant from the National Institute on Aging: R01 AG035009-01A1

REFERENCES

- Bellg AJ, Borrelli B, Resnick B, Hecht J, Minicucci DS, Ory M, ... Treatment Fidelity Workgroup of the, N. I. H. B. C. C. (2004). Enhancing treatment fidelity in health behavior change studies: best practices and recommendations from the NIH Behavior Change Consortium. *Health Psychology*, 23(5), 443–451. doi:10.1037/0278-6133.23.5.443 [PubMed: 15367063]
- Borrelli B (2011). The Assessment, Monitoring, and Enhancement of Treatment Fidelity In Public Health Clinical Trials. *Journal of Public Health and Dentistry*, 71(s1), S52–S63. doi:10.1111/j.1752-7325.2011.00233.x
- Borresen J, & Lambert MI (2008). Autonomic control of heart rate during and after exercise : measurements and implications for monitoring training status. *Sports Medicine*, 38(8), 633–646. [PubMed: 18620464]
- Campbell JA, D'Acquisto LJ, D'Acquisto DM, & Cline MG (2003). Metabolic and cardiovascular response to shallow water exercise in young and older women. *Medicine and Science in Sports and Exercise*, 35(4), 675–681. doi:10.1249/01.MSS.0000058359.87713.99 [PubMed: 12673153]
- Glasziou P, Meats E, Heneghan C, & Shepperd S (2008). What is missing from descriptions of treatment in trials and reviews? *British Medical Journal*, 336(7659), 1472–1474. doi:10.1136/bmj.39590.732037.47 [PubMed: 18583680]
- Hoffmann TC, Glasziou PP, Boutron I, Milne R, Perera R, Moher D, ... Michie S (2014). Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *British Medical Journal*, 348, g1687. doi:10.1136/bmj.g1687 [PubMed: 24609605]
- Kemi OJ, & Wisloff U (2010). High-intensity aerobic exercise training improves the heart in health and disease. *Journal of Cardiopulmonary Rehabilitation and Prevention*, 30(1), 2–11. doi:10.1097/HCR.0b013e3181c56b89 [PubMed: 20040880]
- Nelson ME, Rejeski WJ, Blair SN, Duncan PW, Judge JO, King AC, ... American Heart A. (2007). Physical activity and public health in older adults: recommendation from the American College of Sports Medicine and the American Heart Association. *Circulation*, 116(9), 1094–1105. doi:10.1161/CIRCULATIONAHA.107.185650 [PubMed: 17671236]
- Orwig D, Mangione KK, Baumgarten M, Terrin M, Fortinsky R, Kenny AM, ... Magaziner J (2017). Improving community ambulation after hip fracture: protocol for a randomised, controlled trial. *Journal of Physiotherapy*, 63(1), 45–46. doi:10.1016/j.jphys.2016.10.001 [PubMed: 27964962]
- Puggaard L, Larsen JB, Stovring H, & Jeune B (2000). Maximal oxygen uptake, muscle strength and walking speed in 85-year-old women: effects of increased physical activity. *Aging (Milano)*, 12(3), 180–189. [PubMed: 10965376]
- Resnick B, Michael K, Shaughnessy M, Nahm ES, Sorkin JD, & Macko R (2011). Exercise intervention research in stroke: optimizing outcomes through treatment fidelity. *Topics in Stroke Rehabilitation*, 18 Suppl 1, 611–619. doi:10.1310/tsr18s01-611 [PubMed: 22120030]

- Saki MJ, Meira W (2014). *Data Mining and Analysis: Fundamental Concepts and Algorithms*: Cambridge University Press.
- Scanlan AT, Fox JL, Poole JL, Conte D, Milanovi Z, Lastella M, & Dalbo VJ (2018). A comparison of traditional and modified Summated-Heart-Rate-Zones models to measure internal training load in basketball players. *Measurement in Physical Education and Exercise Science*, 1–7. doi: 10.1080/1091367X.2018.1445089
- Strath SJ, Kaminsky LA, Ainsworth BE, Ekelund U, Freedson PS, Gary RA, ... Council (2013). Guide to the assessment of physical activity: Clinical and research applications: a scientific statement from the American Heart Association. *Circulation*, 128(20), 2259–2279. doi:10.1161/01.cir.0000435708.67487.da [PubMed: 24126387]
- Taylor KL, Weston M, & Batterham AM (2015). Evaluating intervention fidelity: an example from a high-intensity interval training study. *PLoS One*, 10(4), e0125166. doi:10.1371/journal.pone.0125166 [PubMed: 25902066]
- Terbizan DJ D. B, Albano C (2002). Validity of Seven Commercially Available Heart Rate Monitors. *Measurement in Physical Education and Exercise Science*, 6(4), 243–347.
- Wang R, Blackburn G, Desai M, Phelan D, Gillinov L, Houghtaling P, & Gillinov M (2017). Accuracy of Wrist-Worn Heart Rate Monitors. *JAMA Cardiology*, 2(1), 104–106. doi:10.1001/jamacardio.2016.3340 [PubMed: 27732703]

Figure 1:

Heart rate by time curve

A single exercise session is depicted in this graphic. Minutes of exercise are on the x axis and heart rate (HR) in beats per minute are on the y axis. In the first seven minutes of exercise, there are examples of *drops* (HR goes to zero), and *flat lines* (HR stays fixed at the same value for several 15 second episodes). There is a *spike* (HR rises to > 5 SD of the mean HR) that occurs at approximately minute 40. The short black lines on the x axis (apparent after minute 32 and after minute 54) are event markers triggered manually by the interventionist.

```

DATA CLEANING PROCEDURE
Input: All HR Sessions
Output: Set of Cleaned & Usable HR Sessions

For each session:
  if 0 or 1 markers:
    mark as unusable
  if 2 markers:
    use MARKER ESTIMATION PROCEDURE
  if 3 markers:
    if active session length not within 2SD of mean:
      mark as unusable
  if >3 markers:
    use MARKER ELIMINATION PROCEDURE

For each remaining session:
  eliminate HR drop values (HR of 0)
  eliminate HR spike values (HR  $\pm$  5SD from mean HR in session)
  eliminate HR flatline values (Where HR values either:
  Stay fixed for > 1.5 minutes
  Or HR stays fixed for  $\geq$  4 minutes, with a max of 6 non-
  fixed values interspersed between the fixed values.)

  if > 75% of HR values still remain in this session:
    Add session to list of Cleaned & Usable Sessions

MARKER ESTIMATION PROCEDURE
Input: One HR Session containing 2 markers, H
Set of similar HR Sessions with 3 good markers, S
Output: HR Session either marked as unusable, or a third additional
marker (at a computer-estimated location) is added

if 2-marker Session contained an active phase with cool-down:
  Find mean slope for HR drop during cool-down in set S
  if Session H has cool-down with HR drop slope  $\geq$  mean slope:
    Add new 3rd marker to Session H at point before HR begins
    dropping during cool-down
    if active session length not within 2SD of mean:
      mark as unusable
else:
  Find mean duration between 3rd marker & session end in set S
  Add new 3rd marker with above duration to Session H
  if active session length not within 2SD of mean:
    mark as unusable

MARKER ELIMINATION PROCEDURE
Input: One HR Session containing > 3 markers
Output: HR Session either eliminated, or with extra markers deleted (to
now have 3 markers).

For each session:
  if any two markers occur within < 1 minute of each other:
    eliminate the second marker
  if session does not now have 3 markers:
    mark as unusable
  else (session has 3 markers):
    if active session length > 2SD of mean session length:
      mark as unusable

```

Figure 2:

Pseudocode to identify and correct user errors and signal artifacts

This figure contains a general representation of the computer program's functions written in Java for data cleaning procedures, marker estimation procedures, and marker elimination procedures.

Figure 3:

Schematic of analytic processes

Heart rate by time curves were first analyzed by marker placement, which indicated the number and duration of the exercise phases for each session (1a. Marker placement screening). Curves where the marker placement could not be corrected by the computer code were discarded (Marker Fail). All other curves were analyzed for artifacts (2. Artifact Screening), including flat-lines (no change in HR), drops (HR values of zero), and spikes (non-physiologically high HR values in the absence of symptoms). If greater than or equal to 25% of the data points were removed due to artifacts, the curve was discarded. All other curves, including those with no artifacts and those in which the artifacts were corrected, were retained for further analysis.