Journal of Cardiac Failure 30 (2024) 703–716

JCF
Journal of Cardiac Failure

Review Article

Wearable Sensors to Monitor Physical Activity in Heart Failure Clinical Trials: State-of-the-Art Review

RUBEN BUENDIA, PhD[1] MARTIN KARPEFORS, PhD[1] FOLKE FOLKVALJON[2] ROBERT HUNTER[3] HENRIK SILLEN[4] LONG LUU, PhD[5] KIERAN DOCHERTY, MD[6] and MARTIN R COWIE, MD, PhD[7]

Gothenburg, Sweden; Luton, and Glasgow, UK; and Gaithersburg, and Boston, US

ABSTRACT

Background: Estimation of the effects that drugs or other interventions have on patients' symptoms and functions is crucial in heart failure trials. Traditional symptoms and functions clinical outcome assessments have important limitations. Actigraphy may help to overcome these limitations due to its objective nature and the potential for continuous recording of data. However, actigraphy is not currently accepted as clinically relevant by key stakeholders.

Methods and Results: In this state-of-the-art study, the key aspects to consider when implementing actigraphy in heart failure trials are discussed. They include which actigraphy-derived measures should be considered, how to build endpoints using them, how to measure and analyze them, and how to handle the patients' and sites' logistics of integrating devices into trials. A comprehensive recommendation based on the current evidence is provided.

Conclusion: Actigraphy is technically feasible in clinical trials involving heart failure, but successful implementation and use to demonstrate clinically important differences in physical functioning with drug or other interventions require careful consideration of many design choices. (*J Cardiac Fail 2024;30:703–716*)

Key Words: Actigraphy, wearables, clinical trials, heart failure.

Patients with heart failure (HF) experience a high symptom burden and have among the lowest physical activity measured across chronic diseases.[1] There are treatments that reduce morbidity and mortality in HF patients, but there remains an unmet need to meaningfully demonstrate improvement in symptom burden and degree of physical limitation that people with HF live with on a day-to-day basis. Moreover, the FDA has published a draft guidance in June 2019 stating that an improvement in symptoms or physical function can be a basis for approving drugs to treat HF, even in the absence of any significant beneficial effect on morbidity or mortality outcomes.[2] In any case, the effect of a treatment on symptoms or physical function is a fundamental part of its value for patients.

Despite the FDA guidance and the introduction of many new wearable technologies designed specifically to monitor mobility, the research community continues to struggle with the challenge of identifying the most

From the [1]Data Science, Late-Stage Development, Cardiovascular, Renal and Metabolic, BioPharmaceuticals R&D, AstraZeneca, Gothenburg, Sweden; [2]Patient Centered Science, BioPharmaceuticals Business, AstraZeneca, Gothenburg, Sweden; [3]Regulatory, Late-Stage Development, Cardiovascular, Renal and Metabolic, BioPharmaceuticals R&D, AstraZeneca, Luton, UK; [4]Digital Health R&D, AstraZeneca, Gothenburg, Sweden; [5]Digital Health R&D, AstraZeneca, Gaithersburg, MD, US; [6]British Heart Foundation Cardiovascular Research Centre, University of Glasgow, Glasgow, UK and [7]Late-Stage Development, Cardiovascular, Renal and Metabolic, BioPharmaceuticals R&D, AstraZeneca, Boston, MA, US.

Manuscript received November 15, 2023; revised manuscript received January 24, 2024; revised manuscript accepted January 30, 2024.

Reprint requests: Ruben Buendia, PhD. Data Science, Late-Stage Development, Cardiovascular, Renal and Metabolic, BioPharmaceuticals R&D, AstraZeneca, Pepparedsleden 1, 431 83 Mölndal, Sweden. Tel: +467 220 52328. E-mail: Ruben.Buendia@astrazeneca.com

See page 714 for disclosure information.

1071-9164/$ - see front matter

https://doi.org/10.1016/j.cardfail.2024.01.016

suitable approach for measuring physical function. The use of actigraphy, defined as a noninvasive technique for recording and analyzing activity (movement) using wearables,[3] in clinical trials has been limited by patient burden, unreliable accuracy and lack of standardization. The aim of this state-of-the-art article is to provide recommendations on how to fill the current knowledge gaps that prevent actigraphy outcomes from being clinically relevant endpoints acceptable to key stakeholders, including regulators and payers, in addition to those living with HF and their healthcare teams.

Constructing an Actigraphy-Based Endpoint

Currently, the most common measure of symptoms in trials related to heart failure (HF) is based on a subset of questions from the Kansas City Cardiomyopathy Questionnaire (KCCQ),[4] a patient-reported outcome (PRO) instrument. The KCCQ also includes questions regarding other aspects of health-related quality of life as it pertains to HF, including physical limitations. Analogously, the 6-minute walk test (6MWT) is a commonly used exercise test in HF trials, providing the distance walked in 6 minutes (6MWDistance) as a measure of submaximal exercise capacity. Despite KCCQ and 6MWT being well-established, they have important limitations.[5]

Actigraphy offers an objective measure of physical activity by means of wearable devices containing accelerometers. It has the potential to overcome the limitations of exercise tests being captured at fixed (and often a limited number of) timepoints and the subjective nature of PROs. Physical activity captured through actigraphy can be regarded as providing a complementary aspect of physical function that is distinct from that captured by patients' reports or exercise tests. Actigraphy can, thus, complement current clinical outcome assessments (COAs) to provide a more comprehensive assessment of functional status in those with HF. This concept of a multimodality assessment of patients' functional status is represented in Fig. 1, targeting what patients are capable of doing (exercise tests), what they perceive that they do (questionnaires), and what they actually do (actigraphy).

Developing novel endpoints for studies, including those based on actigraphy, requires careful consideration of various factors. These factors include the study's purpose and context, the characteristics of the target population, the sample-size determinations, the actigraphy device and software selection, the monitoring duration and frequency, the actigraphy outcome selection and definition, the statistical analysis approaches, and the data interpretation. Clinical-development programs addressing HF symptoms or physical function need to identify fit-for-purpose COAs to capture these concepts of interest.[6,7]

It is important to consider why a proposed measure represents an outcome that is clinically important for patients affected by a specific disease and how that impact will be quantified within the context of a clinical trial (ie, in an endpoint). These questions must then be answered in the specific context of HF. This is done through the accumulation of both qualitative evidence collected in interviews with patients and clinicians and quantitative evidence collected in experiments. With this objective, a conceptual model of physical functioning in patients with HF was generated by Niklasson and colleagues.[8] Mobility concepts that were most often recorded by patients as being affected by their HF were walking distance, walking speed, going up a steep incline, going up steps, standing for long periods of time, and carrying and lifting objects. Of these, walking distance was the mobility concept most strongly associated with physical and general limitations as well as with symptoms.

Several clinical trials in HF featuring drugs have included endpoints based on actigraphy measures[9–16]; Table 1 summarizes these trials in chronological order of publication. In addition to drug trials, in the WATCHFUL (Pedometer-Based Walking Intervention in Patients With Chronic Heart Failure With Reduced Ejection Fraction) trial,[17] a lifestyle walking intervention in patients with HFrEF, improved daily step count, as measured by a wearable device, by about 25%. This improvement was not matched by improvements in 6MWDistance, NT-proBNP levels or functional capacity scores.

Actigraphy: What Should Be Measured?

Walking Distance and Step Count

To date, walking-distance algorithms derived from wearables acceleration have been limited to remote versions

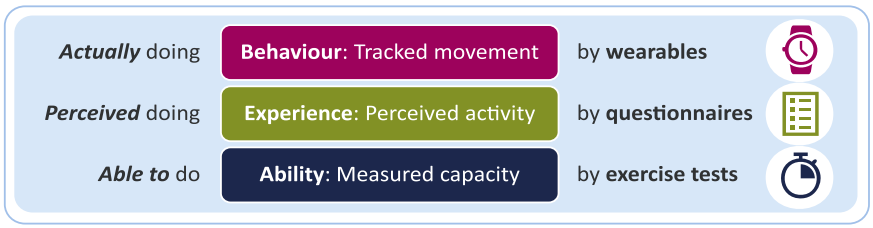

Fig. 1. Proposed model of patient functional status.

Table 1 Actigraphy in published heart failure drug clinical trials

Trial, population	Intervention and design	Actigraphy endpoints	Device and wearing instructions	Number of patients, adherence and missing data	Treatment effect on actigraphy endpoints	Observations
NEAT-HF, patients with HFpEF (Redfield et al., 2015)[9]	Isosorbide mononitrate vs placebo Crossover design with increasing dosage: baseline and washout: 2 weeks, 30 mg−1 week, 60 mg−1 week, 120 mg−2 weeks	The primary endpoint was change in daily activity level, quantified as the average daily accelerometer units (ADAU), (ie, similar to activity counts, during the 120-mg phase). Secondary endpoints included change in hours of activity per day (ie, from very light activity during the 120-mg phase, and change in ADAU during the 3 dose phases combined.	Belt outfitted with 2 kinetic activity monitors (Kersh Health) containing triaxial accelerometers (KXUD9-2050, Kionix) Patients were instructed to wear the device for 24 hours a day throughout the study's duration; to consider the data valid, minimum 10 hours per day and 3 days were required.	During the 120-mg phase, 101 patients for the first period and 91 for the second period Median of 16 complete days (interquartile range, 12−20) during the first period, and 14 complete days (interquartile range, 10−18) during the second period	ADAU decreased progressively and significantly with increased doses of isosorbide mononitrate but not placebo. ADAU was not significantly lower in the 120-mg phase; however, it was in the 3 phases combined. Hours active per day as assessed by actigraphy was significantly lower in the treatment group.	In an ancillary study of NEAT-HF,[81] a modest relation among actigraphy parameters, 6MWT, QoL scores, and Nt-ProBnP was found. Similarly, there was no relation among changes in these COAs. These results are in concordance with a recent study.[82]
INDIE-HF, patients with HFpEF (Borlaug et al., 2018)[10]	Inorganic mitrite vs placebo Same design as NEAT-HF	ADAU during the 120-mg phase, mimicking the primary endpoint in NEAT-HF	Same as NEAT-HF	98 patients for the first period and 93 for the second period All patients had valid data	No difference in ADAU	Very similar design to NEAT-HF
OUTSTEP-HF, patients with HFrEF (Piepoli et al., 2020)[11]	Sacubitril/valsartan vs enalapril Parallel-group design randomized 1:1	Mean difference in nonsedentary daytime	Assessment of PA was performed during the entire duration of the study (from week 2−week 12) using a wrist-worn accelerometery device (MotionWatch8, Camntech, Cambridge, UK) on the nondominant wrist; minimum 10 hours per day and 7 days for each period were required.	493 patients at baseline (14.6% missing data) Sacubitril/valsartan: 287 patients completed the study; 244 with actigraphy data at baseline and 202 at end visit Enalapril: 283 patients completed the study; 238 with actigraphy data at baseline and 193 at end visit	Mean nonsedentary time decreased by 27 min with sacubitril/valsartan and by 21 min with enalapril, which was not statistically significant Activity for sacubitril/valsartan increased at week 1, then sharply decreased; same for enalapril with a lower trend; until week 8, activity for sacubitril/valsartan was higher than baseline and until week 12 higher than enalapril	Baseline values of nonsedentary time were higher than expected. The subsequent decrease in activity might indicate a regression to usual levels.[66] This indicates potential reactivity.

(*Continued*)

Table 1. (*Continued*)

Trial, population	Intervention and design	Actigraphy endpoints	Device and wearing instructions	Number of patients, adherence and missing data	Treatment effect on actigraphy endpoints	Observations
EMPIRE-HF, patients with HFrEF (Jensen et al. 2020)[12]	Empagliflozin vs placebo Parallel-group design, randomized 1:1	Daily activity counts were a secondary endpoint.	Patients were instructed to wear an Actigraph wGT3X-BT (Actigraph, Pensacola, FL) around the waist continuously for 7 days.	Of 190 randomized patients, 166 had valid accelerometer data at both baseline and end-visit.	There was almost no difference in activity counts between baseline and 12 weeks in either of the treatment arms.	
AWAKE-HF, patients with HFrEF (Khandwalla et al. 2021)[13]	Sacubitril/valsartan vs enalapril Parallel-group design, randomized 1:1	Change in counts per minute at week 8 during the most active 30- and 6- min/day, total awake and total sleep time	Actigraphy data were collected continuously during the 18-week study using a Philips Actiwatch Spectrum (Philips Respironics, Boston, MA) worn on the nondominant wrist.	140 were randomized and 120 completed study treatment with valid actigraphy data Patients adhered to wearing the actigraphy wrist monitor throughout the study, with a mean time recorded off-wrist of $\leq$1% at each time point	Whereas there was no detectable difference from baseline or between treatment arms for any endpoint; point estimates indicated more counts per minute (cpm) for enalapril during awake time and periods of maximum activity, and less cpm during sleep	Divergence between actigraphy results and PROs There was an additional 8-week open-label phase during which all patients received sacubitril/valsartan.
CHIEF-HF, patients with HF regardless of EF (Spertus et al., 2022)[14]	Canagliflozin vs placebo Parallel-group design randomized 1:1	Difference between daily steps averaged through 2 weeks at baseline (weeks 1 and 2) and after 12 weeks (weeks 11 and 12)	Fitbit Versa 2 worn on the wrist continuously	Out of 448 patients randomized, 413 had valid accelerometer data at both baseline and end visit	No detectable difference from baseline or between treatment arms	Completely decentralized and virtual trial with no in-person visits The Canagliflozin cohort had >10% more steps than placebo at baseline.
METEORIC-HF, patients with HFrEF (Lewis et al., 2022)[15]	Omecamtiv mecarbil vs placebo Parallel-group design randomized 2:1	Mean daily activity counts measured over a 2-week period from baseline to weeks 18−20	The device was not specified but worn on the wrist for 10 hours during awake time for 2 weeks; minimum 7 days	Of 276 randomized patients, 257 had valid accelerometer data at both baseline and end visit	No detectable difference from baseline or between treatment arms	
DETERMINE trials (McMurray et al., 2023)[16]	Twin trials, DETERMINE-Reduced for HFrEF and DETERMINE-Preserved for HFpEF. Dapagliflozin vs placebo Parallel-group design randomized 1:1	Change from baseline to 16 weeks in time spent in light to vigorous physical activity (nonsedentary time)	Patients were instructed to wear the MoveMonitor (McRoberts) around their waist continuously during 7-day periods at 3 points during the trial; minimum 10 hours and 3 days	Wearable substudy in which 348 of 817 patients (313 and 504 for Reduced and Preserved studies, respectively) wore the device (pooled analysis); 319 patients had adequate baseline data, and 163 had valid accelerometer data at both baseline and end visit.	No detectable difference between treatment arms in DETERMINE-Reduced, DETERMINE-Preserved or the pooled analysis	Other accelerometer-based exploratory endpoints have not yet been reported. The COVID-19 pandemic curtailed the use of monitors in the Preserved study.

of established exercise tests such as the 6MWT.[18] This is because deriving walking distance with high precision during passive monitoring (ie, unobtrusive and continuous measurement of movement) is challenging. Accurate estimation of stride length is likely to be the main challenge. Kirk and colleagues[19] proposed a method to estimate walking speed from wearable data in which steps per second (cadence) were doubled and multiplied by stride length (stride was defined as 2 steps) to obtain walking distance per second. Unfortunately, the estimation of stride length yielded a large mean absolute percentage error (MAPE) of 25.3% for patients with HF.[20] Taking a different approach, Taoum and colleagues[21] trained linear mixed models (LMM), using wearable data from healthy participants, to estimate walking distance and speed. These models were tested in an independent cohort of patients with peripheral artery disease and impaired walking due to lower-limb symptoms. The MAPE for hip devices against Global Positioning System (GPS) distance ranged from 8.4% to 18.8%, depending on the accelerometer-derived parameter that served as input to the linear mixed models. Unfortunately, using the relatively inaccurate GPS-derived distance instead of more accurate measurements limits the validity of these results.

Further efforts to derive walking distance accurately are warranted, but step count might be an acceptable and simple proxy. A change in steps is often proportional to a change in distance.[22] Furthermore, steps might be more representative of an individual's effort than distance itself. For example, a person with a shorter stride length would need more steps to cover the same distance than someone with a longer stride. Analogously, walking on a challenging surface (eg, snow) may require more steps to cover the same distance. Step count is also familiar and easy to communicate, is recognized as meaningful by patients and clinicians and is commonly offered by commercially available activity monitors, albeit with varying levels of accuracy. Moreover, an increasing number of steps is associated with a progressively decreased risk of both cardiovascular disease[23] and all-cause death,[24] particularly for older adults (aged $\geq$ 60 years). For step count to be a suitable outcome in HF trials, however, acceptable accuracy needs to be demonstrated in this population of patients.

Step-count algorithms are often based on data from healthy individuals. However, patients with HF may have different walking patterns. In particular, they often walk at slow speeds.[25] To the best knowledge of the authors, the study by Femiano et al.[26] is the only study that evaluates the accuracy of step count specifically in patients with cardiovascular disease and uses a valid reference. In that study, raw acceleration was captured by the Axivity AX3 (Newcastle upon Tyne, UK) device worn on the wrist, preprocessed using the open software package GGIR,[27] then used as input to 2 different algorithms; walking speed was not controlled. One of the 2 algorithms used by Femiano et al., named Verisense (Cambridge, MA, US),[28] exhibited a MAPE of 4.1%,[26] which can reasonably be considered adequate for use in trials. However, the Verisense algorithm was shown to underestimate step counts at higher paces, thus, greatly increasing the error with increasing walking speed.[29] Rowlands and colleagues attempted to refine this algorithm,[29] but the overall MAPE remained around 10%−12% and, therefore, is not accurate enough to support endpoints in HF clinical trials.[29] Otherwise, Genovese et al.[30] designed ADAM (Adaptive Moment Estimation), an algorithm for step count in slow, very slow and intermittent walking (ie, frequent stops and starts, abrupt directional changes), using the Gear 2 (Samsung, Suwon, Republic of Korea) smartwatch.[30] This algorithm exhibited a MAPE between 1% and 5% in walking experiments, not including turning, and 5%−9% in walk-turn-walk at differing speeds. ADAM could be applied to any wrist-worn device that outputs raw acceleration. It may, however, require additional testing, because the initial study was evaluated using only 8 healthy adults in a battery of experiments not including free walking.

Other studies evaluating step-count algorithms at differing walking speeds concluded that accuracy decreases at lower speeds.[31−36] This is consistent with the conclusion of a recent review where Actigraph (Pensacola, FL) devices were found to be inaccurate for step count at speeds $\leq$ 3.2 km/hour.[37] Storm and colleagues[31] evaluated step-count accuracy for 7 activity monitors with build-in algorithms, worn simultaneously on various body locations and at differing walking speeds. Performance was worst at slow walking speeds in all devices. In that study, the MoveMonitor (McRoberts Technologies, Miami, FL) performed best at all speeds and during "free" outdoor walking, with a MAPE $<$ 2.0%.

The 3 wrist-worn or arm-worn devices tested by Storm and colleagues[31] exhibited low performance at low speeds. This is consistent with the low accuracy found for wrist-worn ActiGraph devices.[36−39] Otherwise, some Garmin (Olathe, Kansas), Fitbit (San Francisco, CA) and Apple Watch (Apple, Cupertino, CA) devices exhibited great accuracy at low and normal walking speeds.[32,36,40] Additionally, some wrist-worn devices (eg, Apple Watch) have the capability to measure heart rate accurately,[41] an important measure related to functional capacity and a measure that may be useful for clinicians and researchers.

In conclusion, step count can be used in HF trials and might be an important measure of patients' function. It is, nevertheless, critical to ensure that the device and the algorithm provide acceptable accuracy. In this regard, Table 2 summarizes the MAPE of devices and algorithms discussed in this section. Although useful, comparison among devices is limited by differences in experiment methodologies and populations in the studies.

Table 2 Step-count MAPE of various devices and algorithms

Wearable Device	Algorithm	Body Placement	Average MAPE at low walking speed (3.2 km/h unless other specified)	Average MAPE at normal walking speed (4.8 km/h unless other specified)	Observations
Samsung Gear 2 smartwatch [30]	ADAM	Wrist	1% MAPE for healthy adults walking slower than their normal speed	3% MAPE walking at variable speed centered at normal speed	5% MAPE for healthy adults walking much slower than their normal speed (experiments not including turning)
MoveMonitor [31]	In-built	Lower back	2% MAPE indoors and 0.5% outdoors for healthy adults walking slower than their normal speed	1.5% MAPE indoors and 0.5% outdoors walking at their normal speed	1% MAPE for fast indoor walking
Fitbit [31,36]	In-built	Hip	One, 2.6% MAPE indoors for healthy adults walking slower than their normal speed[31]. One, 0.8%[36], Zip, 2.7%[36]	One, 1.1% MAPE for healthy adults walking at their normal speed[31] One, 0.5%[36] Zip, 1.2%[36]	1% MAPE for fast indoor walking[31] In [36], on a treadmill at 6.4 km/h, MAPE was 0.4% for both One and Zip.
Garmin[32,36]	In-built	Wrist	Vivosmart, 1.0%*,[32] Vivofit 2, 1.8%[36] Vivofit 3, 2%[36] Vivoactiv HR, 2.3%[36] Vivoactiv 3, 2.5%[36]	0.3%[32] 2.6%[36] 1.9%[36] 3.8%[36] 1.4%[36]	Experiments on a treadmill In [32], average of 2 sessions, MAPE = 10.5% at 6.4 km/h In [36], at 6.4 km/h, MAPE was 17.1% for Vivofit 2, 1.5% for Vivofit 3, 17.7 for Vivoactiv HR, and 1.6 for Vivoactiv 3.
Fitbit[32,36]	In-built	Wrist	Charge HR, −0.8%*[32] Ionic, 4.6%[36]	1.9%[32] 4.2%[36]	In [32], average of 2 sessions, MAPE = 6.5% at 6.4 km/h In [36] MAPE = 3.2% at 6.4 km/h
Apple Watch[32,36,40]	In-built	Wrist	Sport, 1.7%*,[32] Series 1, 4.4%[36] Unspecified Model, −2.6%*,[40]	1.3%[32] 3.1%[36] −0.9%*,[40]	Experiments on a treadmill In [32], average of 2 sessions, MAPE = 0.3% at 6.4 km/h In [36] MAPE = 3.5% at 6.4 km/h In [40] MAPE = -1.6% at 6.4 km/h
Samsung Gear[32,36]	In-built	Wrist	S, 5.2%*,[32] Fit 2, 3.7%[36] Fit 2 Pro, 4.8%[36]	3.8%[32] 2.3% [36] 3.3% [36]	In [32], average of 2 sessions, MAPE = 0.3% at 6.4 km/h, In [36] MAPE was 2.7% for Fit 2 and 5.6% for Fit 2 Pro at 6.4 km/h
Axivity AX3[26]	Verisense[28]	Wrist			MAPE = 4.1% for CV patients with uncontrolled speed
Axivity AX3, GENEActiv Original and ActiGraph GT9X[29]	Verisense[28]	Wrist	9.6%, 9.4%, 11.2% At 4.1 km/h	10.2%, 8.7%, 10.1% At 4.6 km/h	MAPE highly increase with increased speed, >15% at 5.3 km/h and >70% for running at 10.3 km/h
Axivity AX3, GENEActiv Original, ActiGraph GT9X[29]	Revised Verisense[29]	Wrist	12.9%, 12%, 16.9%, At 4.1 km/h	7.1%, 6%, 9.4% At 4.6 km/h	Overall MAPE ∼ 10%−12% across 3 devices and different waking speeds

Devices and algorithms are ordered by increasing MAPE at low speed. Devices with in-built algorithms are in the table only if they were tested in at least 2 studies; nevertheless, different models of a similar device are noted together. Only wrist and waist/hip/lower back devices are considered. Devices with MAPE > 20 % were excluded.

*In [32] and[40] the error is similar to the MAPE, with a difference that can be either negative (underestimation) or positive (overestimation). MAPE, mean absolute percentage error.

Time Spent in Various Levels of Activity

Estimated nonsedentary time and time in moderate-to-vigorous physical activity are recognized as valuable by patients and clinicians; they may also be suitable for capturing the physical impacts of HF.[8] Indeed, international physical activity (PA) recommendations for patients with cardiac illnesses are based on moderate-to-vigorous physical activity (United Kingdom Chief Medical Officers, 2019; World Health Organization, 2010). Unfortunately, assessing various activity levels for various people carries a large intrinsic error. Cut-off values are used to enable the classification of performed activities into light, moderate or vigorous intensity of PA (LPA, MPA VPA) and are usually linked to more established units such as metabolic equivalent of task (MET).[42] A single MET is the oxygen consumed per kg of body weight per minute of an individual sitting quietly. A standard MET is set by convention at 3.5 mL of oxygen per kg per minute. However, the between-individual variability in this measure is a source of error. In a study by Dibben and colleagues,[43] of 22 patients with HF (age 71 $\pm$ 14 years and mostly in New York Heart Association class II), 18 undertook a range of daily living activities (including lying down, sitting, standing, and walking) while measuring PA via wrist- and hip-worn accelerometers and VO_2 via indirect calorimetry. On average, 1 standard MET corresponded to 1.3 METs for a patient with HF. This indicates a large bias for patients with HF compared to healthy individuals; a large variance for both groups was demonstrated by Schwendinger et al.[44] This variance remained when stratifying by age. Therefore, what is light or moderate activity for 1 person can correspond to moderate or vigorous activity for another person of the same age. Another source of error comes from the type of activity and the location of the device. For example, yard sweeping was measured as an average of 3.39 standard METs, comparable to walking at 4 km/hour (3.88 METs).[42] Wrist-worn acceleration during sweeping was slightly higher than acceleration during walking; otherwise, waist-worn acceleration during sweeping was 4.5 times lower than that for walking.[42] These sources of error can be minimized in patients with HF by considering only nonsedentary time. This way, a threshold that accurately distinguishes activities, such as household tasks or walking, from activities that require minimal physical effort, such as standing or sitting, are easier to identify. This is because the separation, in acceleration, between sedentary and nonsedentary activities is larger and more consistent across activities and device placements; this has been shown for patients with HF[43] and for healthy individuals.[44]

Intensity thresholds, which are currently most commonly used for wrist accelerometers, are those considered in the R package GGIR.[27] These thresholds are 100 and 400 milligravities (mg) for moderate and vigorous activity[45] and were derived from 30 healthy adults. The threshold for light activity is 30 mg; Rowland et al.[46] are referenced for this threshold; however, in that study 40 and 50 mg were considered superior so as to distinguish meaningful physical activity from being sedentary. Applying thresholds derived from healthy adults to patients with HF may lead to misclassification of PA levels. Thresholds proposed by Dibben et al.[43] for light and moderate intensities were approximately 40% lower than those derived from healthy adults for both moderate[45] and light activity[47] for wrist-worn devices; differences were much larger for hip-worn or waist-worn devices. Therefore, commonly used activity thresholds are unlikely to be accurate for patients with HF. A large between-individual variability for patients with HF was also demonstrated by Dibben et al.[43] Despite this variability, nonsedentary time can be estimated in patients with HF with relatively high accuracy by devices worn on the wrist or near the center of mass (waist, hip or lower back).[43]

Amount of Movement

Activity counts integrate acceleration over time in order to quantify movement. Using signal processing to derive activity counts from raw acceleration implies choices, which are different for different devices, and have an inherent error.[48] Supplementary Fig. 1 shows how the counts for the CentrePoint Insight Watch (Actigraph) are derived.[48] Activity counts were initially developed because raw acceleration could not be stored due to the limitation of on-board storage and battery capacity.[48] However, this no longer applies to modern devices. Therefore, activity counts can be substituted by average acceleration.

There is no norm, but most of the HF trials that have included actigraphy-based endpoints have used activity counts as an outcome variable.[9,10,12,13,15] These trials exhibited a disparity of endpoint design using activity counts as the building block.

Clinical interpretability of average acceleration is limited. Nevertheless, it has the advantage of simplicity, allowing for the comparison of amount of movement in just 1 number while preserving the granularity of the original signal. This also reduces the amount of estimation error.[49] Therefore, it would be suitable as a complementary building block in the main actigraphy-based endpoint or the main building block as a supportive endpoint.

Actigraphy Measures which Are Less Suitable as COAs

Reduced walking speed was reported as a mobility limitation experienced by people with HF,[8] but this is an elusive concept to capture. The MAPE of walking speed in HF was estimated as 15.5%,[19] which is too high for an outcome measure in HF trials. Otherwise, going up steps could be captured by human-activity recognition algorithms, but this activity is commonly infrequent in daily life (and determined by need and

housing structure for most individuals), so it is not suitable as an outcome measure to evaluate treatment effects outside of controlled settings.

The intensity gradient[50] is an actigraphy-derived measure to assess the distribution of movement intensity across time. It has been claimed that, together with average acceleration, it can fully describe the activity profile of an individual.[51,52] This might have potential value, but clinical interpretability of intensity gradient is limited, and its derivation is not straightforward, as in the case of average acceleration. Moreover, it has never been applied to patients with HF, and further research in the relevant population would be worthwhile.

Other potential measures are active or total energy expenditure (EE). Unfortunately, EE cannot be derived accurately from accelerometers. Large errors (ie, MAPE between 20% and 30%, depending on various factors), in EE estimation from accelerometry data were found for Fitbit, Apple Watch and Actigraph on the wrist and hip.[38,53] Moreover, a recent systematic review and meta-analysis that included 20 studies evaluating EE in a range of conditions (including 1 HF study),[54] concluded that the measurement of energy expenditure is too inaccurate for research purposes.

The strengths and limitations of the measures discussed in this section are summarized in Table 3.

How Should Activity Be Measured?

Once the actigraphy measures used in a trial are decided, the measurement set-up is vital. This section reviews the differing aspects needed to capture reliably and accurately the measures discussed in the previous section.

Selection of Wearable Device and Body Placement

Even if devices can be placed on legs, feet and other body locations, only wrists[11,13–15] and hips/waists/lower backs[9,10,12,16] have been used in large clinical trials. The wrist was the placement of choice in NHANES (National Health and Nutrition Examination Survey)[55] and UK Biobank,[56] the largest studies to date to use accelerometers. Wrist placement is, for most people, most comfortable

Table 3 Summary of actigraphy outcome variables ranked by suitability for clinical trials

Actigraphy Measure	Strengths	Limitations	Suitability as outcome variable
Step count	• Recognized as meaningful by patients and clinicians • Proportional to walking distance • Is familiar and easy to communicate • Can be measured with commercially available activity monitors	• Accuracy may vary and needs further research in HF population • Algorithms are often based on data from healthy individuals, which may not fully capture walking patterns specific to HF patients	Good and likely the most suitable
Nonsedentary time	• Recognized as meaningful by patients and clinicians • Is familiar and easy to communicate	• Whereas it might be estimated with relatively high accuracy, thresholds for HF were derived on only 22 patients	Good
Average acceleration	• It is a measure of movement • Low estimation error • Close to the raw accelerometer data	• Clinical interpretability may be limited	Potentially good
Walking speed	• Reported as a mobility limitation by HF patients	• Currently, estimation of walking speed is inaccurate in HF	Could be considered
Intensity gradient	• Assesses the distribution of movement intensity over time • Complements average acceleration as an activity profile measure	• Clinical interpretability is limited, and its derivation is not straightforward as in the case of average acceleration • To date, it has not been applied to HF patients	Could be considered
Moderate to vigorous PA	• Recognized as meaningful by patients and clinicians • Is familiar and easy to communicate	• Its estimation is currently inaccurate	Unlikely to be suitable
Energy Expenditure	• Recognized as meaningful by patients and clinicians • Is familiar and easy to communicate	• Its estimation is currently very inaccurate	Unlikely to be suitable

and is associated with higher wear-time compliance,[57] thus reducing the potential for bias due to nonwear or selective wear.[58] In addition, wrist devices can be used in clinical practice, not only in clinical trials. Nevertheless, the MoveMonitor (McRoberts), worn on the lower back, exhibited high patient acceptability in the Mobilize-D technical validation study.[59]

The Apple Watch is accurate for step count[32,40] and is well accepted by (and familiar to) the public; however, it includes substantial challenges. Chief among these are issues related to data privacy and lack of transparency in the algorithm. Other disadvantages are the need of calibration, the purchase cost, and that it has to be charged daily. Garmin or Fitbit devices (Table 2) might be alternatives. Another alternative for wrist placement is to use any device that offers acceleration (eg, Axivity AX3), which can then be put through an algorithm that meets a predefined set of criteria to estimate number of steps. The algorithm developed by Genovese et al.[30] is an option that would, however, require validation for HF. Another approach is to develop a new algorithm tailored specifically for patients with HF, which is 1 aim of the Mobilize-D consortium.[19] That endeavor includes recordings of wrist-worn acceleration from 20 healthy adults and 68 patients, including 12 with HF. It included 2.5 hours of free movement and controlled tasks, from a reference system (for steps) and acceleration data from a device worn on the lower back.

The estimation of nonsedentary time might be accurate enough for wrist- and hip-worn devices when using the cut-points proposed by Dibben et al.[43] These thresholds should, nevertheless, be confirmed and refined in new studies with more patients.

Average acceleration represents the movement of the device wherever it is placed. Different devices subjected to the same movement should provide similar acceleration data. However, it cannot be assumed that every device measures acceleration with equal accuracy.[7] The acceleration accuracy and interdevice agreement of Fitbit, Apple Watch, Actigraph, and Garmin devices have been tested by using a shaker table at frequencies that are consistent with human walking and running.[60] All devices exhibited errors that were small but statistically significant, except for the Garmin device, which had a moderate error. Furthermore, the intraclass correlation coefficient (ICC) of Apple and Actigraph were ICCs = 0.99 and 0.97, respectively, whereas the ICC of Fitbit and Garmin was 0.88. In a similar experiment,[61] the DynaPort MiniMod (McRoberts; an older device by the same manufacturer as the MoveMonitor) exhibited excellent accuracy and an ICC of 0.99. Regarding body placement, a wrist-worn device would provide very different values from a device simultaneously worn on the hip.[62] The movement of the hip better reflects walking, which is of key importance because the ability to walk reasonable distances at pace is required for social or work-related activities outside of the home. Otherwise, the movement of the wrist better reflects the household work[62] required by most but is also dependent on age, social norms and sick roles adopted by persons with HF and their care partners. Whereas average acceleration recorded by a device worn on the wrist or hip can provide relevant information, walking may be characterized most adequately by step count; in this regard, as previously discussed, some in-built algorithms achieve high performance despite using wrist acceleration as input, possibly because they were trained with large amounts of data. Thus, wrist average acceleration may be better suited as a complementary outcome variable.

Wearing Time

In trials, limiting device data to that recorded during daytime hours (eg, 8 AM−10 PM) is a possibility. However, whereas most people are most active during the day, a significant percentage of patients with HF are active very early in the morning or during the night.[44] A potential solution to cover all participants would be to have the device worn for 24 hours and to consider all data. However, some patients with HF have reduced activity when awake and increased movement when resting, due to poor sleep quality.[63] Therefore, treatment-related increases in activity during active periods may be partially offset by simultaneous reductions in movement when resting (ie, due to treatment-related improved sleep quality), confounding the assessment of benefit. For this reason, the best option might be to instruct patients to wear the device during their awake times. This would result in greater between-individual variability in wear time; nevertheless, this can be addressed in the statistical analysis. There is no gold standard, but in most published clinical trials of HF,[9–13] patients were instructed to wear the device 24 hours a day through the entire duration of the trial. The exceptions were the METEORIC-HF (Multicenter Exercise Tolerance Evaluation of Omecamtiv Mecarbil Related to Increased Contractility in Heart Failure) trial,[15] which used 10 hours during awake time for 2 weeks, and the EMPIRE-HF (Empagliflozin in heart failure patients with reduced ejection fraction)[12] and DETERMINE (Dapagliflozin Effect on Exercise Capacity Using a 6-Minute Walk Test in Patients With Heart Failure) trials,[16] which required only 1 week of measurements at baseline and at the last visit.

Beyond within-day recordings, small but statistically significant differences between weekdays and Sundays were found by Schwendinger and colleagues.[44] This may vary with different cultures or age groups, which have different resting days. Furthermore, healthy older adults reported being less active in winter compared to summer.[64] In patients with HF, Klompstra and colleagues[65] reported similar data obtained by using a questionnaire. In the same way, a modest decrease in activity was observed during winter in a cohort of healthy participants and those with HF in a Switzerland-based study,[44] which, however,

did not have enough power to show statistical significance. This seasonal difference may depend on geography and seasonal weather patterns.

When deciding on wear time, it is important to take reactivity into account. Reactivity, or the Hawthorne effect, occurs when individuals modify their behavior because they are aware of being monitored. This effect can be magnified if participants are able to view activity metrics in their devices.[7] The OUTSTEP-HF (randOmized stUdy Using accelerometry to Compare Sacubitril/valsarTan and Enalapril in Patients With Heart Failure) study[11,66] found substantial differences between the beginning and end of the trial (which could indicate reactivity), whereas other HF trials did not exhibit similar findings. Interestingly, a study involving patients with HF wearing wrist-worn devices for 2 weeks showed no evidence of reactivity.[44] Nevertheless, to minimize this effect, it is advisable to establish baseline values after a certain wear-time, such as 2 weeks; this assumes reactivity is reduced as the patient gets used to being monitored.

Limitations

Algorithms are likely to underperform when estimating proposed measures in situations that result in reduced wrist movement, such as when patients are using walking aids,[67] carrying objects or pushing a pram.[68] Additionally, the limited number of HF trials that have employed actigraphy to date leaves many questions unanswered.

Statistical Considerations

Protocol deviations, such as a device not being worn according to instructions, and technical malfunctions can result in missing data. Requirements in which days are included in the calculation of, for example, mean steps per day or mean daily nonsedentary time, can help to mitigate the impact of these issues. Additionally, data-driven missing-data handling rules, such as imputation of specific time windows (eg, days, hours or even seconds) or applying different daily weights (eg, based on wear-time or weekdays vs weekends) need to be decided. A proper missing-data strategy can minimize the impact of within-individual variability. Missing data can vary substantially among studies, and there is little literature to help design a proper strategy. Catellier and colleagues[69] approached this topic and recommended imputation for PA measured by accelerometers.

Once missing data are handled, actigraphy measurements can represent building blocks for endpoints. What is different from other COA-based endpoints is that actigraphy is data-dense, and data aggregation has more possibilities than, for example, PROs. In order to assess *meaningful change*, the aggregate (summary) statistic at the 2 time points (eg, mean or maximum or percent of days with a value $>$ X, etc.) need to have a well-defined period over which to aggregate. This choice should be based on qualitative data, a period that matters to most patients.

Construct Validity for Actigraphy Measures as Endpoints

For COAs, in addition to defining the endpoint, the measurement properties of the underlying parameter need to be studied, documented and understood. This includes demonstrating that the parameter (eg, daily number of steps or daily nonsedentary time), is reliable (ie, stable across timepoints between which no change is expected) and responsive (ie, sensitive to changes that are known to have occurred). Thresholds for clinically meaningful differences in the COA are also necessary to aid interpretation and contextualization of results. This can be estimated by anchor-based analysis, anchoring to well-understood outcomes; for example, if it can be demonstrated that a certain change in mean daily number of steps over a month postpones time-to-hospitalization by a year. Alternatively, it can be anchored to a specifically designed patient global impression of change if it targets the same underlying concept. For example, a 4-week assessment of mean steps per day may be anchored by a patient global impression of change that asks, "How has the amount you have been limited in your walking ability changed over the past 4 weeks?"

Relation to known hard outcomes also supports construct validity. In this regard, there is a clear relationship between the endpoints proposed in this review and events of cardiovascular disease,[23] all-cause mortality,[24,70–73] cardiovascular death,[74,75] and hospitalization.[76] In 2 comprehensive meta-analyses, Paluch and colleagues found a strong association of age with the connection between steps and all-cause mortality[24] and cardiovascular disease events.[23] In both cases, there was a lower risk of adverse outcomes with increased daily steps; there was a plateauing in this relationship, with risk at approximately 6000−8000 steps for those aged $\geq$ 60 years and 8000−10000 steps for those aged $<$ 60 years.

Unfortunately, comorbidities and other extracardiac factors may limit the potential for improvement in physical activity, despite an improvement in HF status.[77,78] Adding an exercise program or support program to drug treatment is a potential strategy to mitigate motivational barriers. In this case, wearable devices showing activity metrics can be helpful.[7]

Additional Considerations for Integrating Actigraphy in Trials

Minimizing the burden actigraphy represents for the various stakeholders is crucial for the successful implementation of actigraphy in a trial. Critical aspects include, but are not limited to, how user-friendly the devices are for

trial participants (wearability and user interface) and site staff (set-up interface and trouble-shooting), data transfer, patient privacy, cost of device and supporting systems, and technical support. Key practical factors for patients, sites and sponsors are summarized in Fig. 2.

Burden for Patients and Study Sites

It is important that participants be able to use the device easily during the required data-collection period. This includes the physical design and shape of the wearable (ie, the form factor) so that it minimizes inconvenience for the patient and ensures that the device is not removed during parts of the day. Thus, the design of the device should be attractive to wear, and issues, such as gender differences in the acceptability of a device, its position and appearance and its wear period, should be considered. Usability factors, such as requirement for charging during a collection period, possibility to shower or bathe and easy or automated data transfer or storage, are also important considerations.[7]

The wearing of an activity monitor (unless watch-based for someone used to wearing a watch) is an additional activity for individuals and is likely to inconvenience them, add to their daily routines and mean that they must remember to do specific tasks, so the setup, use and training need to be as easy and seamless as possible. This would also minimize site burden and avoid extended or additional site visits. If successfully implemented, actigraphy can reduce the number of required site visits. The need for technical support should also be minimized, but its availability should be maximized to avoid study sites' becoming service desks. Furthermore, data transfer, storage and management should be transparent to patients and sites, while potentially allowing for monitoring of missing data.

In most trials, the devices are given only to individuals taking part in the trial during the trial period—partly because, during this period, device-specific support is also provided. The devices are usually destroyed after the study has been completed. It is interesting to consider whether patients could be allowed to keep the devices and whether that might increase patients' engagement and motivation.

Trial Sponsors

As a sponsor, it is important to ensure that selected vendors can comply with data-integrity requirements and clarity of data ownership and that they provide reassurance regarding the privacy of patients' data. Software integration to facilitate the transfer, visualization and analysis of data is another important factor to consider. Additional aspects are global device availability, availability of and access to technical support, initial procurement costs, and maintenance costs. There also needs to be consideration of software longevity. Actigraphy device software undergoes regular patching and updating, and there has to be a guarantee that data are comparable, regardless of when it was collected during the potentially long duration of the clinical trial.

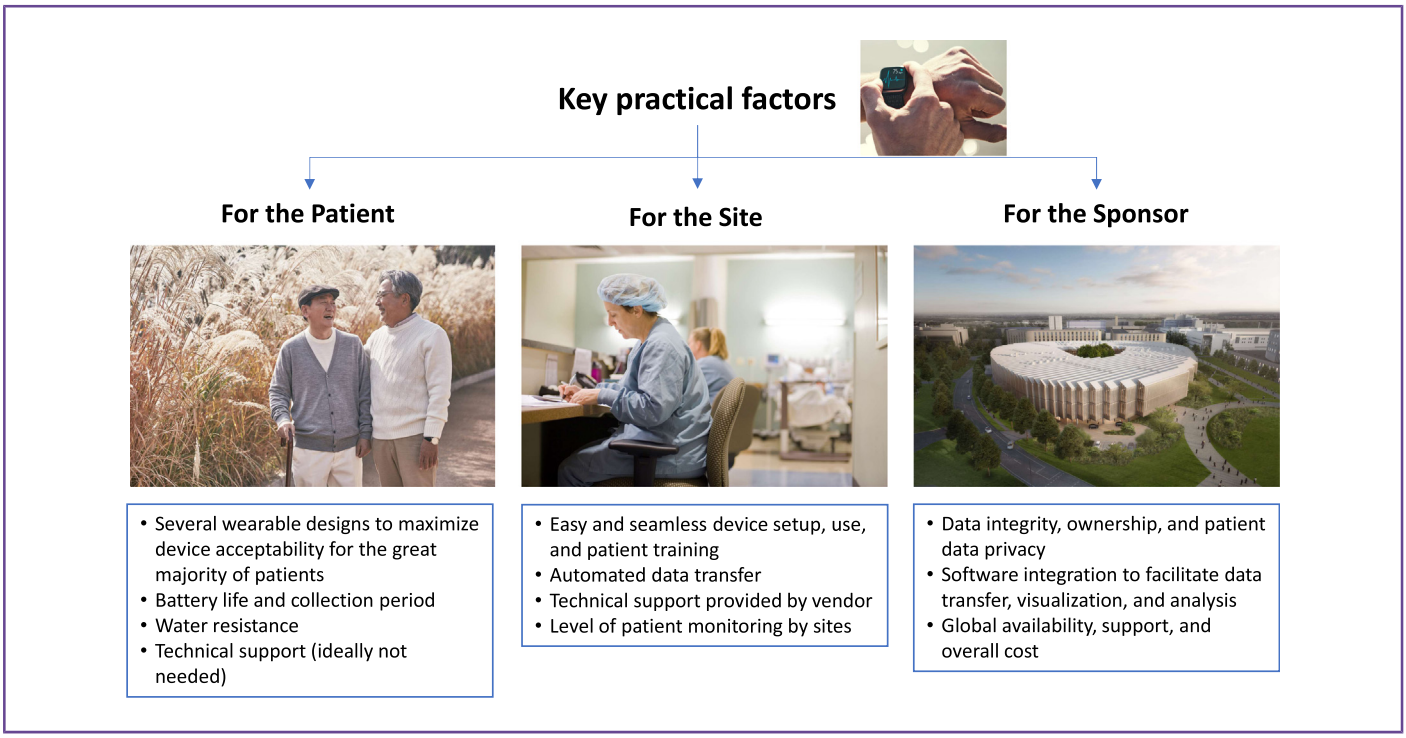

Fig. 2. Key practical factors to consider for integrating actigraphy in trials.

Pragmatic Trials

Pragmatic trials would present additional sources of heterogeneity, such as different devices, different patterns of wear-time and nonrandom missing data. Given that many devices and algorithms are not suitable, and that wearing conditions are critical, actigraphy might not be well-suited for pragmatic trials in HF.

Regulatory Agencies' Position on Actigraphy Endpoints

Regulatory agencies play a critical role in guiding the development of novel COAs and endpoints. Both the Food and Drug Administration (FDA) and the European Medicines Agency (EMA) have encouraged sponsors to explore the development of functions and symptoms assessment.[79] As an example, the EMA recently qualified Stride velocity 95th centile as a primary endpoint (ie, an alternative to the 6MWT) in studies of ambulatory Duchenne muscular dystrophy. Although that disease is very different from HF, the qualification opinion[80] expressed a favorable position by the EMA concerning actigraphy.

In HF, regulatory agency recommendations underscore the possibility of incorporating actigraphy-based endpoints in clinical trials so as to demonstrate improvements in (or prevention or deterioration of) exercise and/or symptoms,[2] rather than having to pursue exclusively long-term cardiovascular outcomes trials that are intended to demonstrate long-term survival-benefit indications. The FDA states that improvements in walking, exercising or performing other activities of daily living can be the basis of showing the effectiveness of an HF drug. Also, measures of activity and activities of daily living, including accelerometry data, can be acceptable endpoints.[2] It should be noted that both the FDA and the EMA employ and apply qualification procedures when assessing the clinical meaningfulness of novel devices and endpoints, and these can take considerable time to validate, because the relevant evidence has to be generated.

The FDA patient-focused drug-development guidance series provides comprehensive guidance about COA development,[6,79] extending to parameters derived from actigraphy. The incorporation of actigraphy-based clinical trial endpoints necessitates adherence to established standards while navigating additional complexities. The regulatory landscape and considerations for actigraphy-based endpoints must appreciate the significance of aligning outcomes with studys' objectives.

Conclusions and Recommendations

Actigraphy has the potential to provide relevant information about the functional status of patients with HF and should, therefore, be more frequently implemented in HF trials. One of the main advantages over other measurements of physical function is its continuous nature and, as such, long measurement periods would realize the full potential of actigraphy to provide more comprehensive and important information. We suggest the optimal clinical trial setup may be that in which patients wear a wristband during awake time for as long as the trial runs, provided such measurements do not impose an inappropriate burden on the people in the study.

We propose that step count is the best candidate to be the outcome variable that constitutes the main building block of the primary endpoint. Nonsedentary time and average acceleration would support secondary or exploratory endpoints. Data from the weeks 1−2 (or a run-in period) can be collected and excluded from the main analysis to mitigate reactivity; weeks 3−6 may serve as baseline, and the last 4 weeks can be used to determine change from baseline.

Mean differences, or median differences relative to baseline, have the advantage of being straightforward measurements of average treatment effect that are easy to communicate. Depending on the duration of the trial, seasonal variations should be considered in the analysis. Finally, all data should be included in supplemental analyses and made available to the research community.

Disclosures

RB, MK, FF, RH, HS and LL are employees of AstraZeneca and may own stock or stock options. KD, the University of Glasgow, has been remunerated by AstraZeneca for work relating to clinical trials; he has received speaker's honoraria from AstraZeneca, Pharmacosmos and Radcliffe Cardiology, served on an advisory board for Us2.ai and Bayer AG, served on a clinical endpoint committee for Bayer AG, and received research grant support from Boehringer Ingelheim, Roche Diagnostics and AstraZeneca (paid to his institution). MRC is a Senior Vice President of AstraZeneca and may own stock or stock options.

Lay Summary

People living with heart failure experience high burdens of symptoms and functional limitations, but ways to measure and report them are currently limited. Actigraphy, using wearable devices, provides an opportunity to measure objectively continuous physical activity and changes over time, including those in clinical trials of drug therapy. Despite the potential advantages, the use of actigraphy to evaluate treatment benefits remains in its infancy. This state-of-the-art study provides comprehensive recommendations about how to implement actigraphy in HF clinical trials successfully, including the types of devices, the measurements to be taken, how to analyze them, and how best to describe the changes seen.

Ruben Buendia

CRediT authorship contribution statement

RUBEN BUENDIA: Conceptualization, Investigation, Methodology, Visualization, Writing — original draft, Writing — review & editing, Project administration. **MARTIN KARPEFORS:** Project administration, Supervision, Visualization, Writing — original draft, Writing — review & editing. **FOLKE FOLKVALJON:** Writing — original draft, Writing — review & editing. **ROBERT HUNTER:** Writing — original draft. **HENRIK SILLEN:** Investigation, Writing — review & editing. **LONG LUU:** Investigation, Writing — review & editing. **KIERAN DOCHERTY:** Supervision, Validation, Writing — review & editing. **MARTIN R COWIE:** Conceptualization, Project administration, Supervision, Validation, Writing — review & editing.

Supplementary materials

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.cardfail.2024.01.016.

References

1. Barker J, Smith Byrne K, Doherty A, Foster C, Rahimi K, Ramakrishnan R, et al. Physical activity of UK adults with chronic disease: cross-sectional analysis of accelerometer measured physical activity in 96 706 UK Biobank participants. Int J Epidemiol 2019;48:1386. https://doi.org/10.1093/ije/dyz148.
2. Food and Drug Administration. Treatment for Heart Failure: Endpoints for Drug Development: Guidance for Industry. 2019.
3. Acebo C, LeBourgeois MK. Actigraphy. Respir Care Clin N Am 2006;12:23–30. https://doi.org/10.1016/j.rcc.2005.11.010.
4. Green CP, Porter CB, Bresnahan DR, Spertus JA. Development and evaluation of the Kansas City Cardiomyopathy Questionnaire: a new health status measure for heart failure. J Am Coll Cardiol 2000;35:1245–55. https://doi.org/10.1016/s0735-1097(00)00531-3.
5. Psotka MA, Abraham WT, Fiuzat M, Filippatos G, Lindenfeld J, Ahmad T, et al. Functional and Symptomatic Clinical Trial Endpoints: the HFC-ARC Scientific Expert Panel. JACC Heart Fail 2022;10:889–901. https://doi.org/10.1016/j.jchf.2022.09.012.
6. Food and Drug Administration. Patient-focused drug development: selecting, developing, or modifying fit-for-purpose clinical outcome assessments. Guidance for Industry. 2022.
7. Food and Drug Administration. Digital Health Technologies for Remote Data Acquisition in Clinical Investigations Guidance for Industry, Investigators, and Other Stakeholders. 2023.
8. Niklasson A, Maher J, Patil R, Sillen H, Chen J, Gwaltney C, et al. Living with heart failure: patient experiences and implications for physical activity and daily living. ESC Heart Fail 2022;9:1206–15. https://doi.org/10.1002/ehf2.13795.
9. Redfield MM, Anstrom KJ, Levine JA, Koepp GA, Borlaug BA, Chen HH, et al. Isosorbide mononitrate in heart failure with preserved ejection fraction. N Engl J Med 2015;373:2314–24. https://doi.org/10.1056/NEJMoa1510774.
10. Borlaug BA, Anstrom KJ, Lewis GD, Shah SJ, Levine JA, Koepp GA, et al. Effect of inorganic nitrite vs placebo on exercise capacity among patients with heart failure with preserved ejection fraction: the INDIE-HFpEF randomized clinical trial. JAMA 2018;320:1764–73. https://doi.org/10.1001/jama.2018.14852.
11. Piepoli MF, Hussain RI, Comin-Colet J, Dosantos R, Ferber P, Jaarsma T, et al. OUTSTEP-HF: randomised controlled trial comparing short-term effects of sacubitril/valsartan versus enalapril on daily physical activity in patients with chronic heart failure with reduced ejection fraction. Eur J Heart Fail 2021;23:127–35. https://doi.org/10.1002/ejhf.2076.
12. Jensen J, Omar M, Kistorp C, Poulsen MK, Tuxen C, Gustafsson I, et al. Twelve weeks of treatment with empagliflozin in patients with heart failure and reduced ejection fraction: a double-blinded, randomized, and placebo-controlled trial. Am Heart J 2020;228:47–56. https://doi.org/10.1016/j.ahj.2020.07.011.
13. Khandwalla RM, Grant D, Birkeland K, Heywood JT, Fombu E, Owens RL, et al. The AWAKE-HF study: sacubitril/valsartan impact on daily physical activity and sleep in heart failure. Am J Cardiovasc Drugs 2021;21:241–54. https://doi.org/10.1007/s40256-020-00440-y.
14. Spertus JA, Birmingham MC, Nassif M, Damaraju CV, Abbate A, Butler J, et al. The SGLT2 inhibitor canagliflozin in heart failure: the CHIEF-HF remote, patient-centered randomized trial. Nat Med 2022;28:809–13. https://doi.org/10.1038/s41591-022-01703-8.
15. Lewis GD, Voors AA, Cohen-Solal A, Metra M, Whellan DJ, Ezekowitz JA, et al. Effect of omecamtiv mecarbil on exercise capacity in chronic heart failure with reduced ejection fraction: the METEORIC-HF randomized clinical trial. JAMA 2022;328:259–69. https://doi.org/10.1001/jama.2022.11016.
16. McMurray JJV, Docherty KF, de Boer RA, Hammarstedt A, Kitzman DW, Kosiborod MN, et al. Effect of dapagliflozin versus placebo on symptoms and 6-minute walk distance in patients with heart failure: the DETERMINE randomized clinical trials. Circulation 2024;149:825–38. https://doi.org/10.1161/CIRCULATIONAHA.123.065061.
17. Vetrovsky T, Siranec M, Frybova T, Gant I, Svobodova I, Linhart A, et al. Lifestyle walking intervention in patients with heart failure with reduced ejection fraction: the WATCHFUL trial. Circulation 2024;149:177–88. https://doi.org/10.1161/CIRCULATIONAHA.123.067395.
18. Mak J, Rens N, Savage D, Nielsen-Bowles H, Triggs D, Talgo J, et al. Reliability and repeatability of a smartphone-based 6-min walk test as a patient-centred outcome measure. Eur Heart J Digit Health 2021;2:77–87. https://doi.org/10.1093/ehjdh/ztab018.
19. Kirk C, Küderle A, Micó-Amigo ME, et al. Mobilise-D insights to estimate real-world walking speed in multiple conditions with a wearable device. Sci Rep 2024;14:1754. https://doi.org/10.1093/ehjdh/ztab018.
20. Mico-Amigo ME, Bonci T, Paraschiv-Ionescu A, Ullrich M, Kirk C, Soltani A, et al. Assessing real-world gait with digital technology? Validation, insights and recommendations from the Mobilise-D consortium. J Neuroeng Rehabil 2023;20:78. https://doi.org/10.1186/s12984-023-01198-5.
21. Taoum A, Chaudru S, DEM PY, Congnard F, Emily M, Noury-Desvaux B, et al. Comparison of activity monitors accuracy in assessing intermittent outdoor walking. Med Sci Sports Exerc 2021;53:1303–14. https://doi.org/10.1249/MSS.0000000000002587.
22. Bassett Jr DR, Ainsworth BE, Leggett SR, Mathien CA, Main JA, Hunter DC, et al. Accuracy of five electronic pedometers for measuring distance walked. Med Sci Sports Exerc 1996;28:1071–7. https://doi.org/10.1097/00005768-199608000-00019.
23. Paluch AE, Bajpai S, Ballin M, Bassett DR, Buford TW, Carnethon MR, et al. Prospective association of daily steps with cardiovascular disease: a harmonized meta-analysis. Circulation 2023;147:122–31. https://doi.org/10.1161/CIRCULATIONAHA.122.061288.
24. Paluch AE, Bajpai S, Bassett DR, Carnethon MR, Ekelund U, Evenson KR, et al. Daily steps and all-cause mortality: a meta-analysis of 15 international cohorts. Lancet Pub Health 2022;7:e219–28. https://doi.org/10.1016/S2468-2667(21)00302-9.
25. Vidan MT, Blaya-Novakova V, Sanchez E, Ortiz J, Serra-Rexach JA, Bueno H. Prevalence and prognostic impact of frailty and its components in non-dependent elderly patients with heart failure. Eur J Heart Fail 2016;18:869–75. https://doi.org/10.1002/ejhf.518.
26. Femiano R, Werner C, Wilhelm M, Eser P. Validation of open-source step-counting algorithms for wrist-worn tri-axial accelerometers in cardiovascular patients. Gait Posture 2022;92:206–11. https://doi.org/10.1016/j.gaitpost.2021.11.035.
27. Migueles JH, Rowlands AV, Huber F, Sabia S, Hees VTv. GGIR: a research community-driven Open Source R package for generating physical activity and sleep outcomes from multi-day raw accelerometer data. J Measure Phys Behav 2019;2:188–96.
28. Patterson MR. Development of an algorithm to count steps from 24hr wrist accelerometry data 2020. Unpublished research.
29. Rowlands AV, Maylor B, Dawkins NP, Dempsey PC, Edwardson CL, Soczawa-Stronczyk AA, et al. Stepping up with GGIR: validity of step cadence derived from wrist-worn research-grade accelerometers using the verisense step count algorithm. J Sports Sci 2022;40:2182–90. https://doi.org/10.1080/02640414.2022.2147134.
30. Genovese V, Mannini A, Sabatini M. A smartwatch step counter for slow and intermittent ambulation. IEEE Access 2017;5:13028–37.
31. Storm FA, Heller BW, Mazza C. Step detection and activity recognition accuracy of seven physical activity monitors. PLoS One 2015;10:e0118723. https://doi.org/10.1371/journal.pone.0118723.
32. Fokkema T, Kooiman TJ, Krijnen WP, CP VDS, M DEG. Reliability and validity of ten consumer activity trackers depend on walking speed. Med Sci Sports Exerc 2017;49:793–800. https://doi.org/10.1249/MSS.0000000000001146.
33. Chen CK, Cheng LY, Hsu SW, Liao MT, Ku PW, Liu YB. Comparative analysis of physical activity detected via an external accelerometer and cardiac implantable electronic devices. Front Cardiovasc Med 2022;9:898086. https://doi.org/10.3389/fcvm.2022.898086.

34. Feito Y, Bassett DR, Thompson DL. Evaluation of activity monitors in controlled and free-living environments. Med Sci Sports Exerc 2012;44:733–41. https://doi.org/10.1249/MSS.0b013e3182351913.
35. Vetrovsky T, Siranec M, Marencakova J, Tufano JJ, Capek V, Bunc V, et al. Validity of six consumer-level activity monitors for measuring steps in patients with chronic heart failure. PLoS One 2019;14:e0222569. https://doi.org/10.1371/journal.pone.0222569.
36. Mora-Gonzalez J, Gould ZR, Moore CC, Aguiar EJ, Ducharme SW, Schuna Jr JM, et al. A catalog of validity indices for step counting wearable technologies during treadmill walking: the CADENCE-adults study. Int J Behav Nutr Phys Act 2022;19:117. https://doi.org/10.1186/s12966-022-01350-9.
37. Ngueleu AM, Barthod C, Best KL, Routhier F, Otis M, Batcho CS. Criterion validity of ActiGraph monitoring devices for step counting and distance measurement in adults and older adults: a systematic review. J Neuroeng Rehabil 2022;19:112. https://doi.org/10.1186/s12984-022-01085-5.
38. Nuss KJ, Hulett NA, Erickson A, Burton E, Carr K, Mooney L, et al. Comparison of energy expenditure and step count measured by ActiGraph accelerometers among dominant and nondominant wrist and hip sites. J Measure Phys Behav 2020;3:315–22.
39. Gall N, Sun R, Smuck M. A Comparison of wrist- versus hip-worn actigraph sensors for assessing physical activity in adults: a systematic review. J Measure Phys Behav 2022;5::252–62.
40. Veerabhadrappa P, Moran MD, Renninger MD, Rhudy MB, Dreisbach SB, Gift KM. Tracking steps on Apple Watch at different walking speeds. J Gen Intern Med 2018;33:795–6. https://doi.org/10.1007/s11606-018-4332-y.
41. Wallen MP, Gomersall SR, Keating SE, Wisloff U, Coombes JS. Accuracy of heart rate watches: implications for weight management. PLoS One 2016;11:e0154420. https://doi.org/10.1371/journal.pone.0154420.
42. Esliger DW, Rowlands AV, Hurst TL, Catt M, Murray P, Eston RG. Validation of the GENEA accelerometer. Med Sci Sports Exerc 2011;43:1085–93. https://doi.org/10.1249/MSS.0b013e31820513be.
43. Dibben GO, Gandhi MM, Taylor RS, Dalal HM, Metcalf B, Doherty P, et al. Physical activity assessment by accelerometry in people with heart failure. BMC Sports Sci Med Rehabil 2020;12:47. https://doi.org/10.1186/s13102-020-00196-7.
44. Schwendinger F, Wagner J, Infanger D, Schmidt-Trucksass A, Knaier R. Methodological aspects for accelerometer-based assessment of physical activity in heart failure and health. BMC Med Res Methodol 2021;21:251. https://doi.org/10.1186/s12874-021-01350-6.
45. Hildebrand M, VANH VT, Hansen BH, Ekelund U. Age group comparability of raw accelerometer output from wrist- and hip-worn monitors. Med Sci Sports Exerc 2014;46:1816–24. https://doi.org/10.1249/MSS.0000000000000289.
46. Rowlands AV, Mirkes EM, Yates T, Clemes S, Davies M, Khunti K, et al. Accelerometer-assessed physical activity in epidemiology: are monitors equivalent? Med Sci Sports Exerc 2018;50:257–65. https://doi.org/10.1249/MSS.0000000000001435.
47. Hildebrand M, Hansen BH, van Hees VT, Ekelund U. Evaluation of raw acceleration sedentary thresholds in children and adults. Scand J Med Sci Sports 2017;27:1814–23. https://doi.org/10.1111/sms.12795.
48. Neishabouri A, Nguyen J, Samuelsson J, Guthrie T, Biggs M, Wyatt J, et al. Quantification of acceleration as activity counts in ActiGraph wearable. Sci Rep 2022;12:11958. https://doi.org/10.1038/s41598-022-16003-x.
49. Brage S, Westgate K, Franks PW, Stegle O, Wright A, Ekelund U, et al. Estimation of free-living energy expenditure by heart rate and movement sensing: a doubly-labelled water study. PLoS One 2015;10:e0137206. https://doi.org/10.1371/journal.pone.0137206.
50. Rowlands AV, Edwardson CL, Davies MJ, Khunti K, Harrington DM, Yates T. Beyond cut points: accelerometer metrics that capture the physical activity profile. Med Sci Sports Exerc 2018;50:1323–32. https://doi.org/10.1249/MSS.0000000000001561.
51. Vetrovsky T, Clark CCT, Bisi MC, Siranec M, Linhart A, Tufano JJ, et al. Advances in accelerometry for cardiovascular patients: a systematic review with practical recommendations. ESC Heart Fail 2020;7:2021–31. https://doi.org/10.1002/ehf2.12781.
52. Rowlands AV, Fairclough SJ, Yates T, Edwardson CL, Davies M, Munir F, et al. Activity intensity, volume, and norms: utility and interpretation of accelerometer metrics. Med Sci Sports Exerc 2019;51:2410–22. https://doi.org/10.1249/MSS.0000000000002047.
53. Nuss KJ, Thomson EA, Courtney JB, Comstock A, Reinwald S, Blake S, et al. Assessment of accuracy of overall energy expenditure measurements for the Fitbit Charge HR 2 and Apple Watch. Am J Health Behav 2019;43:498–505. https://doi.org/10.5993/AJHB.43.3.5.
54. Chevance G, Golaszewski NM, Tipton E, Hekler EB, Buman M, Welk GJ, et al. Accuracy and precision of energy expenditure, heart rate, and steps measured by combined-sensing Fitbits against reference measures: systematic review and meta-analysis. JMIR Mhealth Uhealth 2022;10:e35626. https://doi.org/10.2196/35626.
55. Troiano RP, McClain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity research. Br J Sports Med 2014;48:1019–23. https://doi.org/10.1136/bjsports-2014-093546.
56. Doherty A, Jackson D, Hammerla N, Plotz T, Olivier P, Granat MH, et al. Large-scale population assessment of physical activity using wrist worn accelerometers: the UK Biobank Study. PLoS One 2017;12:e0169649. https://doi.org/10.1371/journal.pone.0169649.
57. Freedson PS, John D. Comment on "Estimating activity and sedentary behavior from an accelerometer on the hip and wrist.". Med Sci Sports Exerc 2013;45:962–3. https://doi.org/10.1249/MSS.0b013e31827f024d.
58. Toftager M, Kristensen PL, Oliver M, Duncan S, Christiansen LB, Boyle E, et al. Accelerometer data reduction in adolescents: effects on sample retention and bias. Int J Behav Nutr Phys Act 2013;10:140. https://doi.org/10.1186/1479-5868-10-140.
59. Keogh A, Alcock L, Brown P, Buckley E, Brozgol M, Gazit E, et al. Acceptability of wearable devices for measuring mobility remotely: observations from the Mobilise-D technical validation study. Digit Health 2023;9:20552076221150745. https://doi.org/10.1177/20552076221150745.
60. James III WW, Olivia F, Nick T, Srihari N, David III EB, Russ P, et al. Comparison of raw accelerometry data from ActiGraph, Apple Watch, Garmin, and Fitbit using a mechanical shaker table. medRxiv 2023:2023.05.25.23290556. https://doi.org/10.1101/2023.05.25.23290556.
61. Van Hees VT, Slootmaker SM, De Groot G, Van Mechelen W, Van Lummel RC. Reproducibility of a triaxial seismic accelerometer (DynaPort). Med Sci Sports Exerc 2009;41:810–7. https://doi.org/10.1249/MSS.0b013e31818ff636.
62. Karas M, Bai J, Straczkiewicz M, Harezlak J, Glynn NW, Harris T, et al. Accelerometry data in health research: challenges and opportunities. Stat Biosci 2019;11:210–37. https://doi.org/10.1007/s12561-018-9227-2.
63. Hastings PC, Vazir A, O'Driscoll DM, Morrell MJ, Simonds AK. Symptom burden of sleep-disordered breathing in mild-to-moderate congestive heart failure patients. Eur Respir J 2006;27:748–55. https://doi.org/10.1183/09031936.06.00063005.
64. Aspvik NP, Viken H, Ingebrigtsen JE, Zisko N, Mehus I, Wisloff U, et al. Do weather changes influence physical activity level among older adults? The Generation 100 study. PLoS One 2018;13:e0199463. https://doi.org/10.1371/journal.pone.0199463.
65. Klompstra L, Jaarsma T, Stromberg A, van der Wal MHL. Seasonal variation in physical activity in patients with heart failure. Heart Lung 2019;48:381–5. https://doi.org/10.1016/j.hrtlng.2019.04.003.
66. Docherty KF, Vaduganathan M. OUTSTEP-HF: re-evaluating the role of physical activity measures in drug and device development in heart failure. Eur J Heart Fail 2021;23:136–9. https://doi.org/10.1002/ejhf.2106.
67. Kooner P, Schubert T, Howard JL, Lanting BA, Teeter MG, Vasarhelyi EM. Evaluation of the effect of gait aids, such as canes, crutches, and walkers, on the accuracy of step counters in healthy individuals. Orthop Res Rev 2021;13:1–8. https://doi.org/10.2147/ORR.S292255.
68. Grym K, Niela-Vilen H, Ekholm E, Hamari L, Azimi I, Rahmani A, et al. Feasibility of smart wristbands for continuous monitoring during pregnancy and one month after birth. BMC Preg Childbir 2019;19:34. https://doi.org/10.1186/s12884-019-2187-9.
69. Catellier DJ, Hannan PJ, Murray DM, Addy CL, Conway TL, Yang S, et al. Imputation of missing data when measuring physical activity by accelerometry. Med Sci Sports Exerc 2005;37:S555–62. https://doi.org/10.1249/01.mss.0000185651.59486.4e.
70. Hansen BH, Dalene KE, Ekelund U, Wang Fagerland M, Kolle E, Steene-Johannessen J, et al. Step by step: association of device-measured daily steps with all-cause mortality-a prospective cohort study. Scand J Med Sci Sports 2020;30:1705–11. https://doi.org/10.1111/sms.13726.
71. Zaghi A, Holm H, Korduner J, Dieden A, Molvin J, Bachus E, et al. Physical inactivity is associated with post-discharge mortality and re-hospitalization risk among Swedishheart failure patients:the HARVEST-Malmo Study. Front Cardiovasc Med 2022;9:843029. https://doi.org/10.3389/fcvm.2022.843029.
72. Lear SA, Hu W, Rangarajan S, Gasevic D, Leong D, Iqbal R, et al. The effect of physical activity on mortality and cardiovascular disease in 130,000 people from 17 high-income, middle-income, and low-income countries: the PURE study. Lancet 2017;390:2643-5-4. https://doi.org/10.1016/S0140-6736(17)31634-3.
73. Jeong SW, Kim SH, Kang SH, Kim HJ, Yoon CH, Youn TJ, et al. Mortality reduction with physical activity in patients with and without cardiovascular disease. Eur Heart J 2019;40:3547–55. https://doi.org/10.1093/eurheartj/ehz564.
74. Hall KS, Hyde ET, Bassett DR, Carlson SA, Carnethon MR, Ekelund U, et al. Systematic review of the prospective association of daily step counts with risk of mortality, cardiovascular disease, and dysglycemia. Int J Behav Nutr Phys Act 2020;17:78. https://doi.org/10.1186/s12966-020-00978-9.
75. Kraus WE, Powell KE, Haskell WL, Janz KF, Campbell WW, Jakicic JM, et al. Physical activity, all-cause and cardiovascular mortality, and cardiovascular disease. Med Sci Sports Exerc 2019;51:1270–81. https://doi.org/10.1249/MSS.0000000000001939.
76. Tan MKH, Wong JKL, Bakrania K, Abdullahi Y, Harling L, Casula R, et al. Can activity monitors predict outcomes in patients with heart failure? A systematic review. Eur Heart J Qual Care Clin Outcomes 2019;5:11–21. https://doi.org/10.1093/ehjqcco/qcy038.
77. Chudasama YV, Khunti KK, Zaccardi F, Rowlands AV, Yates T, Gillies CL, et al. Physical activity, multimorbidity, and life expectancy: a UK Biobank longitudinal study. BMC Med 2019;17:108. https://doi.org/10.1186/s12916-019-1339-0.
78. Yang M, Kondo T, Adamson C, Butt JH, Abraham WT, Desai AS, et al. Impact of comorbidities on health status measured using the Kansas City Cardiomyopathy Questionnaire in patients with HFrEF and HFpEF. Eur J Heart Fail 2023. https://doi.org/10.1002/ejhf.2962.
79. Food and Drug Administration. Patient-focused drug development: incorporating clinical outcome assessments into endpoints for regulatory decision-making guidance for industry, Food and Drug Administration staff, and other stakeholders. 2023.
80. European Medicines Agency. Qualification opinion for stride velocity 95th centile as primary endpoint in studies in ambulatory Duchenne muscular dystrophy studies. Committee for Medicinal Products for Human Use (CHMP) 2023.
81. Snipelisky D, Kelly J, Levine JA, Koepp GA, Anstrom KJ, McNulty SE, et al. Accelerometer-measured daily activity in heart failure with preserved ejection fraction: clinical correlates and association with standard heart failure severity indices. Circ Heart Fail 2017;10:e003878. https://doi.org/10.1161/CIRCHEARTFAILURE.117.003878.
82. German CA, Brubaker PH, Nelson MB, Fanning J, Ye F, Kitzman DW. Relationships between objectively measured physical activity, exercise capacity, and quality of life in older patients with obese heart failure and preserved ejection fraction. J Card Fail 2021;27:635–41. https://doi.org/10.1016/j.cardfail.2020.12.025.