Habitual physical activity levels of adults with heart failure: systematic review and meta-analysis

Cara Jordan 1, Sarah J Charman 1,2,3, Alan Mark Batterham,1 Darren Flynn,4 David Houghton,2,5 Linda Errington,6 Guy MacGowan,3,6 Leah Avery1

ABSTRACT
Objective To conduct a systematic review and meta-analysis to quantify habitual physical activity (PA) levels of patients with heart failure (HF) and assess the quality of reporting of device-assessed PA.
Methods Eight electronic databases were searched up to 17 November 2021. Data on the study and population characteristics, method of PA measurement and PA metrics were extracted. A random-effects meta-analysis (restricted maximum likelihood with Knapp-Hartung SE adjustment) was conducted.
Results Seventy-five studies were included in the review (n=7775 patients with HF). Meta-analysis was restricted to mean steps per day, encompassing 27 studies (n=1720 patients with HF). Pooled mean steps per day were 5040 (95% CI: 4272 to 5807). The 95% prediction interval for mean steps per day in a future study was 1262 to 8817. Meta-regression at the study level revealed that a 10-year increment in the mean age of patients was associated with 1121 fewer steps per day (95% CI: 258 to 1984).
Conclusions Patients with HF are a low-active population. These findings have implications for the way in which PA is targeted in patients with HF, and interventions should focus on addressing the age-related decline observed as well as increasing PA to improve HF symptoms and quality of life.
PROSPERO registration number CRD42020167786.

INTRODUCTION
Heart failure (HF) is a complex clinical syndrome, caused by structural cardiac abnormalities resulting in impaired cardiac function.1 Common symptoms include breathlessness and fatigue. It is estimated that 64 million people are living with HF worldwide.2 Within the UK, the prevalence of HF is 654 521.3 The prognosis of HF is poor with survival rates of 73.9%, 45.5% and 24.5% at 1, 5 and 10 years, respectively.4 Patients experience repeated hospitalisations, particularly during 30 days from discharge.5 To reduce healthcare costs and improve patient outcomes, tailored treatment approaches are required.6
Cardiac rehabilitation (CR) elicits improvements in physical activity (PA) levels.7 Prior to the COVID-19 pandemic, the number of patients offered CR in the UK was suboptimal (<15%) despite its demonstrated efficacy for reducing mortality and hospitalisations8 and of those referred <10% attended.9 Increasing PA in combination with standard pharmacological treatment has been shown to elicit beneficial outcomes in patients with HF, reducing all-cause and cardiovascular-related mortality.10 The American College of Cardiology, the American Heart Association1 and secondary prevention guidelines support the usage of PA as part of CR for patients with HF, with guidelines similar to those for the general population11 and those with chronic health conditions.12 Critically, an understanding of the habitual PA levels of patients with HF will help to inform the development of interventions to promote PA, enabling clinicians to engage patients in self-management of HF using PA.13 Our primary aim was to conduct a systematic review and meta-analysis to quantify habitual PA levels of patients with HF. A secondary aim was to assess the quality of reporting of PA measurement in device-based studies examining PA levels of patients with HF.
METHODS
This systematic review was conducted in accordance with a published protocol\textsuperscript{13} and adhered to the Preferred Reporting Items for Systematic Reviews and Meta-Analysis (PRISMA) guidelines.\textsuperscript{14} The review team included male and female members, including three professors, one cardiologist consultant, a PhD student (lead author), a medical librarian and two mid-career researchers.

Search strategy
Eight bibliographic databases (CDSR, CINAHL, CENTRAL, Embase, MEDLINE, PsycINFO, Scopus and Web of Science) were searched (up to 17 November 2021) by a Medical Research Librarian (LE) using a structured search strategy including a combination of Medical Subject Headings and keywords (online supplemental file 1). Search results were restricted to English language and human studies. Hand searching of end reference lists and citation searches of included studies (using Endnote X9) were also conducted to identify relevant studies that were potentially missed by the database search.

Eligibility criteria
Inclusion criteria
Included articles were any primary quantitative research studies that reported data on PA levels of adults (≥18 years old, community dwelling or residents in supported care environments (eg, retirement complex or nursing care homes that provide assisted living)) with a confirmed diagnosis of HF defined by the New York Heart Association (NYHA) classification and/or physician diagnosed via left ventricular ejection fraction (LVEF) through echocardiogram imaging. Studies using self-based methods of PA assessment were eligible for inclusion. The inclusion criteria for device-based studies were a minimum wear time of ≥10 hours per day\textsuperscript{15} for ≥4 consecutive days.\textsuperscript{16,17} Criteria for questionnaire-based studies included those that focused solely on PA amount (eg, minutes) and intensity (eg, moderate-vigorous PA (MVPA)). Multiarm trials that involved adults with other health conditions (eg, chronic obstructive pulmonary disease), left ventricular assist devices and coronary artery bypass graft were included providing one arm consisted of patients with HF with baseline PA data as defined above.

Study selection
Two reviewers (CJ and SJC) independently screened 20% of titles and abstracts of articles retrieved by the search (stage one). Intercoder agreement between reviewers was assessed using Kappa (\(\kappa\)), whereby the threshold of ≥0.80 was set.\textsuperscript{18} If the agreement threshold was met, the first author (CJ) continued to screen the remaining 80% of articles retrieved. If ≥0.80 agreement was not achieved, the same two reviewers independently screened a further 20% of titles and abstracts and repeated this process until a level of ≥0.80 agreement was achieved. Full-text articles of studies retained at stage one were obtained for further scrutiny (stage two). The same two reviewers independently assessed 20% of full-text articles using a study selection form (online supplemental file 2).

Data extraction
The following data were extracted from included studies using a standardised data extraction form (online supplemental file 3): study characteristics, population characteristics, device wear time and calibration for data processing, and PA outcomes. Where studies referred to supplementary data files or previously published papers (ie, study protocols), they were retrieved and relevant data were extracted. The data extraction form was piloted with two studies. One reviewer (CJ) extracted data from 100% of full-text studies, two reviewers (SJC and DH) each extracted data from 20% of full-text studies and data extraction was checked and discussed to ensure accuracy. In situations where studies reported SD for steps per day, those were converted to mean with an SE using the formulae recommended by the Cochrane Collaboration Handbook.\textsuperscript{19} Similarly, where studies had multiple HF groups, the groups were combined (ie, group one combined with group two) using the formulae recommended by Cochrane Collaboration Handbook (Section 7.7.3.8) to create a single baseline PA outcome or single study-level characteristic (eg, age).

Quality of reporting
The habitual level of PA at baseline was the primary outcome in this review, and this was not influenced by the study design. A checklist was developed to assess the quality of device-based reporting of PA (online supplemental file 4). An expert in PA measurement (AMB) and two health psychologists (LA and DF) with expertise in developing and evaluating PA interventions (and informed by the work of Montoye and colleagues\textsuperscript{20}) developed a 13-item checklist to establish the quality of reporting across a range of devices for device-based measurement of PA: device details, wear time, device calibration and outcomes (online supplemental file 4). For each item, a yes=1 point/no=0 point was applied, with a total score ranging from 0 to 13.

Synthesis of results
Descriptive details of study characteristics including year of publication, study design, PA measurement type and metrics used are presented in a narrative format. Descriptive statistics for specific population characteristics are presented as mean and SD.

Meta-analysis
Extracted summary data (mean daily steps and SE) were meta-analysed using Stata software.\textsuperscript{21} For studies that reported median and IQR only, we assumed that the authors reported these measures of central tendency and variability due to their PA data being substantially right-skewed. Therefore, we converted the median and IQR to mean and SD assuming a log-normal distribution.\textsuperscript{22} In one instance in Guder et al.,\textsuperscript{23} this process resulted in an estimated mean smaller than the stated median. Therefore, for this study, we assumed a normal distribution, with the mean equal to the median and the SD approximated using the formula IQR/1.35 (Cochrane Collaboration Handbook Section 6.5.2.5).\textsuperscript{24}

We conducted a random-effects meta-analysis using restricted maximum likelihood and Knapp-Hartung-Sidik-Jonkman adjustment of SE to calculate the weighted pooled mean effect with a corresponding 95% CI. This pooled mean is an estimate of the centre of the distribution of true effects, assuming heterogeneity of effect size. Heterogeneity was quantified using the tau statistic—an SD reflecting the typical variability in effect size (mean daily steps) between studies. Using the between-study variance and the SE for the mean effect, we derived a 95% prediction interval.\textsuperscript{25} This interval provides a plausible range of effect sizes, compatible with the data and model, for a future study conducted in similar (exchangeable) settings. Using the SE of the prediction interval, we derived the probability\textsuperscript{23} that the mean steps per day in a new study
would exceed the threshold of 7000 associated with meeting the minimum recommended amount of MVPA for healthy older adults.26 An approximate CI for this probability was estimated by first calculating the 95% CI for tau-squared using the Q-profile method27 and using these values to derive the SE of the prediction interval. No other meta-analyses were performed due to the inconsistencies in the data reported (ie, other metrics were unsuitable for meta-analysis).

Meta-regression was conducted to explore heterogeneity using three study-level covariates: mean age, proportion of men and whether steps were measured using a pedometer or other device (binary variable). A simple linear model was used. There were insufficient data to properly explore non-linearity. Small study effects and outlying studies were investigated by plotting the residuals from the meta-analysis against the SE for each study, with bias manifesting as non-uniformly distributed residuals for studies with large SEs. In a sensitivity analysis, any such studies were removed, and the meta-analysis was repeated.

RESULTS

The database search generated 15 606 records. Title and abstract and full-text screening led to 100% agreement following discussions. Seventy-five studies were included in the review (see figure 1 for PRISMA diagram). For a full reference list, see online supplemental file 5.

Study characteristics

Study characteristics for the 75 included studies are reported in online supplemental file 6. Of the 75 included studies, the year of publication ranged from 1991 to 2021, with the majority (n=64, 85%) published between 2010 and 2021. Nineteen studies were randomised controlled trials, twenty-two were cross-sectional studies, twenty-two were observational studies, six were case–control studies, three were pilot/feasibility studies and three were quasi-experimental studies.

Nine studies used self-report (questionnaire-based) methods of PA measurement. Four studies used self-report and device-based measures of PA, and three studies used the doubly labelled water technique. The remaining 59 studies used various devices to assess PA (online supplemental file 7).

Population characteristics

A total of 7775 patients with HF were included across 75 studies. The mean age across 74 studies was 65.6±8.4 years; one study did not report age. Sex was reported in 72 studies; the recruitment rate of men was higher than women in 64 studies. Across 72 studies the numbers of men and women were 5531 (71%) and 2139 (28%), respectively. Time since diagnosis was reported in 11 studies, with a mean time of 59.4±38.6 months (4.9±3.2 years) and a range of 3 to 120 months.

LVEF was reported in 51 studies, with a mean HF with reduced ejection fraction (HFrEF), HF with preserved ejection fraction...
Heart failure and cardiomyopathies

<table>
<thead>
<tr>
<th>Study</th>
<th>Steps/day with 95% CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alosco et al. (2015)</td>
<td>4349 [3769, 4928]</td>
</tr>
<tr>
<td>Andrade et al. (2021)</td>
<td>6450 [5621, 7279]</td>
</tr>
<tr>
<td>Andreass et al. (2019)</td>
<td>4623 [4137, 5109]</td>
</tr>
<tr>
<td>Borland et al. (2014)</td>
<td>3871 [2872, 4770]</td>
</tr>
<tr>
<td>Cowie et al. (2010)</td>
<td>4666 [4006, 5326]</td>
</tr>
<tr>
<td>Da Silva et al. (2013)</td>
<td>9229 [7751, 10707]</td>
</tr>
<tr>
<td>Fioelset et al. (2017)</td>
<td>4890 [4028, 5752]</td>
</tr>
<tr>
<td>German et al. (2014)</td>
<td>3785 [3415, 4155]</td>
</tr>
<tr>
<td>Gue et al. (2021)</td>
<td>10566 [7279, 13853]</td>
</tr>
<tr>
<td>Hoodless et al. (1983)</td>
<td>4973 [3818, 6028]</td>
</tr>
<tr>
<td>Houghton et al. (2000)</td>
<td>2137 [1441, 2833]</td>
</tr>
<tr>
<td>Houghton et al. (2002)</td>
<td>2171 [1482, 2890]</td>
</tr>
<tr>
<td>Izewa et al. (2011)</td>
<td>6154 [5715, 6593]</td>
</tr>
<tr>
<td>Izwet al. (2012)</td>
<td>6388 [5907, 6869]</td>
</tr>
<tr>
<td>Izwet al. (2013a)</td>
<td>6332 [5952, 6712]</td>
</tr>
<tr>
<td>Izwet al. (2013b)</td>
<td>5987 [5575, 6399]</td>
</tr>
<tr>
<td>Jaiclovic et al. (2014)</td>
<td>5490 [3417, 7563]</td>
</tr>
<tr>
<td>Jehn et al. (2009)</td>
<td>7257 [6393, 8151]</td>
</tr>
<tr>
<td>McCarthy et al. (2015)</td>
<td>6590 [5457, 7723]</td>
</tr>
<tr>
<td>Miyahara et al. (2018)</td>
<td>1854 [1298, 2410]</td>
</tr>
<tr>
<td>Nguyen et al. (2012)</td>
<td>7464 [6522, 8408]</td>
</tr>
<tr>
<td>Okwose et al. (2017)</td>
<td>5108 [3765, 6451]</td>
</tr>
<tr>
<td>Shiraiishi et al. (2021)</td>
<td>3751 [1631, 5871]</td>
</tr>
<tr>
<td>Stocker et al. (2019)</td>
<td>3892 [3271, 4513]</td>
</tr>
<tr>
<td>Werhahn et al. (2019)</td>
<td>3612 [1560, 5664]</td>
</tr>
<tr>
<td>Yano et al. (2020)</td>
<td>4005 [3343, 4667]</td>
</tr>
<tr>
<td>Yvarri et al. (2017)</td>
<td>3099 [2394, 3804]</td>
</tr>
<tr>
<td>Overall</td>
<td>5040 [4272, 5807]</td>
</tr>
</tbody>
</table>

Figure 2  Forest plot illustrating mean steps per day in 27 studies. The midpoint of the diamond represents the overall pooled weighted mean steps per day (5040) with the end points of the diamond depicting the 95% prediction interval. REML, restricted maximum likelihood.

(HFpEF) and HF with midrange ejection fraction (HFmrEF) of 30.2±5.0%; mean, 58.2±5.5% and 45.6±3.1%, respectively. Of the 51 studies that reported LVEF, HFpEF was diagnosed across 41 studies, with the remaining 10 studies reporting HFpEF (n=5) or HFmrEF (n=5). The NYHA classification was reported in 59 studies. The proportion of patients in NYHA class II (n=3410, 49%) was higher than the other NYHA functional classes: NYHA class I (n=1144, 17%), class III (n=2133, 31%) and class IV (n=220, 3%). Data for individual population characteristics and study characteristics are presented in online supplemental file 6.

Quality of reporting

Of the 75 included studies, 63 reported PA using device-based measures and 3 of those studies used more than one device. Therefore, 66 devices were assessed using the quality of reporting checklist (online supplemental file 8).

‘Domain 1: Device details’—all studies reported details of all devices used (n=66). Authors provided sufficient information to indicate that researchers had received training for PA measurement (n=51), and device placement was reported by 53 studies. The most common placement was the hip/waist.

‘Domain 2: Device wear time’—the wear time (hours) commonly used was continuously/24 hours (n=26) or waking hours (n=14), and the wear time (days) commonly used was 7 days (n=32). Despite all included studies requesting a minimum/target wear time, only 15 studies reported on valid hours and 16 studies reported on valid days (n=14, 21% and n=16, 24%, respectively) that would constitute an evidence-based valid wear time for obtaining sufficient PA data representative of a week. The most common valid days and hours were ≥4 days and ≥10 hours. Fewer than half of the studies reported actual device wear time (n=26). This ranged from 5 days to 14 days and 12.7 to 16.2 hours per day. Implantable cardiac devices (ICDs) measure activity continuously and have 100% wear time. Eight studies used ICDs, and the range in duration of collection time was 7 days to 1 year.

‘Domain 3: Device calibration’—threshold cut-points to detect intensities of PA were reported in 20 studies (30%). Epoch length to detect accelerometer signalling was reported in 26 studies and ranged from 1 s to 15-min epochs; 60 s were used most frequently (n=10, 15%).

‘Domain 4: Outcomes’—all 63 studies reported on PA metrics and PA outcomes for each device.

PA metrics and outcomes

All 75 studies reported PA outcomes and metrics and the number of metrics used by self-report and/or device-based studies (see online supplemental file 9). Online supplemental file 10 provides details of PA metrics for individual studies.

Meta-analysis

Two studies were excluded from the step data meta-analysis (see online supplemental file 11). One study reported the median daily steps with no measure of variability, and the other study reported the median and the range, with no SD or IQR. Therefore, 27 studies were included in the meta-analysis (n=1720 patients with HF). The mean age across these 27 studies was 64.2±8.3 years, and the majority of participants were men (n=1274, 74%). For individual population and study characteristics of the 27 meta-analysed studies, refer to online supplemental file 11. The analysis revealed a pooled mean of 5040 steps per day (95% CI: 4272 to 5807, p<0.0001). Between-study variability (tau) was ±1796 steps per day (figure 2).

Prediction interval

The 95% prediction interval for the mean steps per day in a future study was 7 days to 1 year.

Exploration of small study effects and outlying studies revealed one moderate outlier. Repeating the meta-analysis with this study removed made no material difference to the findings, with a pooled mean effect of 4916 (95% CI: 4201 to 5630) steps per day and a between-study variability (tau) of 1796 steps per day.

Meta-regression

Meta-regression by study-level covariates of age (range 46–79 years), sex (proportion male) and device type (pedometer vs other) revealed that only age accounted for a substantial proportion (29%) of the between-study variability in mean steps per...
day. A 10-year increment in the mean age of patients was associated with 1121 fewer steps per day (95% CI: 258 to 1984; p=0.013).

**DISCUSSION**

This is the first systematic review and meta-analysis to quantify habitual PA of patients with HF, indicating that they are representative of a low-active population. Quantifying habitual PA helps to inform PA recommendations for this specific clinical population.

Furthermore, the prediction interval for a new study conducted in similar contexts reveals that it is unlikely that the average steps per day would meet the minimum recommended level for older adults. A previous review of steps per day in populations living with chronic disease and disability reported high variability ranging from 1214 to 8008 steps per day. This simple range is consistent with the prediction interval reported in our analysis.

Steps per day is a useful indicator of daily PA levels, as well as a feasible metric to assess the prognosis of patients with HF, with ≤4889 steps/day indicative of poor prognosis. With reference to the graduated step index, more than half (56%) of studies included reported mean steps per day that would be reflective of a ‘sedentary’ population (<5000 steps per day). This cut-point denoting a very low active population closely approximates the pooled mean steps per day in our meta-analysis (the midpoint of the prediction interval).

PA measured by steps per day has shown to decrease in those living with chronic conditions due to their debilitating effects. In addition, the degenerative ageing process comes with functional limitations and lower levels of PA. This observation is consistent with our findings from the meta-regression of >1100 fewer steps per day with every 10-year increment in age, though caution is warranted in drawing conclusions from meta-regression using study-level covariates. A cohort study exploring PA trajectories on cardiovascular disease risk in adults aged ≥65 years reported lower HF incidence rates in active men and women compared with low-active or inactive counterparts. Such findings have implications for clinicians to be trained in the benefits of PA and to be provided with evidence-informed tools to promote PA behavioural change, to delay the onset of age-related decline and to provide patients with HF with physiological and psychological benefits. However, we acknowledge that, depending on disease severity, increasing PA may not be an option for some patients with HF.

The quality of methods reporting in studies was identified as poor in this review, with the exception of ‘domain 1—device details’. With reference to our quality of reporting checklist, the under-reported domains included device wear time and device calibration. Selection of evidence-based target and actual device wear time are important variables that should be reported, as this reflects an individual’s or a group’s habitual PA over a specified period in time. Previously, researchers have developed recommendations for best practices in using accelerometry. It is recommended that authors report on the brand and location of accelerometer placement, sampling periods, methods for determining wear time, criteria for a valid day and quality control checks. Despite frameworks having been developed, there is no requirement for researchers to use them when reporting studies.

**Limitations**

No previous study has aimed to precisely quantify the habitual PA levels of the HF patient population. Nevertheless, this study has limitations. First, many of the included studies failed to report key methodological items for PA measurement in device-based studies, creating difficulties with between-study comparisons and decreasing the reproducibility of findings. As we have demonstrated, under-reporting on key items, or simply not calibrating devices to the relevant population makes it more difficult to interpret findings and may produce considerable variability across PA outcomes. It should also be acknowledged that the quality assessment of PA only concerns studies selected for the primary objective of this study.

For a PA intervention to be effective, the assessment of PA is required to enable clinicians to make appropriate recommendations for patients, specifically those identified as low active. Due to the methodological heterogeneity across studies with a wide variety of PA outcomes, our meta-analysis was restricted to steps/day. A detailed analysis of the other PA outcomes (minutes, metabolic equivalents (METs), MET-minute, active calories, MVPA, activity counts, vector magnitude unit (VMU), milligravity and miles) would have enabled a more comprehensive assessment of true baseline levels reflecting the multidimensionality of PA. Although the available data suggest that this population is low active according to other PA outcomes, we cannot definitively conclude this. Ideally, multiple PA metrics should be reported to generate a PA profile across physiologically important dimensions that reflect health-related outcomes. It has been suggested that reliance on one PA descriptor could lead to an inaccurate representation of true PA levels. This issue should be addressed in future research on patients with HF.

**CONCLUSIONS**

The main finding from this review is that patients with HF are typically low active, and the probability of this population meeting the daily recommended level of moderate PA of 7000 steps/day is low (unlikely). It is, therefore, imperative to develop interventions that address the age-related and condition-related decline in PA of this population, and work to increase PA with the overall aim of achieving levels that are beneficial for physical and mental health and well-being. The findings of this systematic review have implications for intervention design and evaluation, including clinician training on PA behavioural change to improve HF symptoms, reduce hospital admissions and enhance the quality of life.

**Twitter Cara Jordan @carajordan95**

**Acknowledgements** We would like to express our gratitude to the authors of included studies who responded positively to requests for additional information and data.

**Contributors** LA, AMB and DF designed the review protocol. LE conducted the search strategy with input from CJ, LA and AMB. CJ conducted the study selection and data extraction with assistance from SJC and DH. AMB, LA, DF and CJ developed the methodological quality assessment form. CJ completed the methodological quality assessment with assistance from SJC. CJ drafted the initial manuscript. All authors contributed to the initial and final drafts of the manuscript. LA, AMB and DF provided ongoing supervision to CJ. LA is the guarantor of this work and accepts full responsibility for the work and controlled the decision to publish.

**Funding** Funding for this research was provided by Teesside University via a Doctoral Studentship.

**Competing interests** None declared.

**Patient and public involvement** Patients and/or the public were not involved in the design, conduct, reporting or dissemination plans of this research.

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

**Data availability statement** All data relevant to the study are included in the article or uploaded as supplementary information.
Heart failure and cardiomyopathies

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

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

ORCID iDs
Cara Jordan http://orcid.org/0000-0003-4853-3511
Sarah J Charman http://orcid.org/0000-0002-3029-7672

REFERENCES
21 StataCorp L. Stata statistical software: release 15 college station. TX StataCorp LP, 2013.
22 Anderson Cancer Center De8, Parameter solver v.3.0. University of Texas MD.