Validity of Fitbit activity monitoring for adults with progressive muscle diseases

Sarah F. Roberts-Lewis, Claire M. White, Mark Ashworth and Michael R. Rose

School of Population Health and Environmental Sciences, King’s College London, London, UK; Neurology Department, King’s College Hospital, London, UK

ABSTRACT

Purpose: Measuring physical activity informs activity recommendations in clinical practice and provides outcomes in clinical trials that are meaningful to patients. Activity assessment in muscle disease is challenging and there is insufficient evidence to support any single activity measure; however, multi-modal activity measurement might have potential.

Materials and methods: This two-part study included 20 and 95 adults with progressive muscle diseases with mobility ranging from independent to assisted, including wheelchair users. Their activity was measured using a multi-sensor Fitbit activity monitor, for which criterion validity and acceptability were tested in study 1 and validity, reliability, and responsiveness were tested in the longitudinal, home-based study 2.

Results: Study 1: Fitbit was acceptable and had strong criterion validity (rho/kappa ≥0.90), although up to 15% measurement error. Study 2: Fitbit had satisfactory concurrent and construct validity, reliability, and responsiveness. However, Fitbit active minutes registered 75 min more activity per week than gold standard moderate and vigorous physical activity (MVPA) time.

Conclusions: Fitbit had satisfactory measurement properties for monitoring physical activity in adults with progressive muscle diseases. However, Fitbit should not be considered an exact step counter, heart rate monitor or calorimeter and Fitbit active minutes are not synonymous with MVPA time.

Introduction

Measuring physical activity is important for adults with progressive muscle diseases to inform activity recommendations in clinical practice and for outcomes in clinical trials that are meaningful to patients [1,2]. In muscle disease, functional heterogeneity and abnormalities in gait, heart rate and metabolism make activity assessment particularly challenging.

Currently, there is insufficient evidence to support the use of any single physical activity measure for adults with progressive muscle diseases [1,3]. To compensate for limitations in physical activity measurement approaches, combined approaches using multiple activity metrics have been proposed [4,5]. Several multi-sensor activity monitors, with smart phone connectivity, are commercially available and the technology is regularly evolving. Some smart activity monitors, including Fitbits, have demonstrated satisfactory measurement properties in healthy populations [6]. However, it is not known whether Fitbit multisensor activity monitoring is suitable for adults with progressive muscle diseases.

The aim of this two-part study was to evaluate a smart, multi-sensor Fitbit activity monitor as a measure of physical activity in adults with progressive muscle diseases. The objective of the first study was to test the criterion validity and acceptability of Fitbit. The objective of the second study was to test the validity, reliability, and responsiveness of Fitbit compared to a research-grade activity monitor.

Methods

Study design and setting

Ethical approval for the first prospective, observational study was granted in 2017 by King’s College London Research Ethics Committee (LRS-16/17-4226). From October 2017 to April 2018, participants were invited for activity monitoring, during lying, sitting, walking, and cycling tasks with a physiotherapist at King’s College London, Guy’s Hospital Campus (n = 16). For those unable
to travel to London, activity monitoring was carried out at a gym close to their address (n = 4).

Ethical approval for the second prospective, home-based observational study was granted in 2019 (LRS-18/19-10909). Free-living activity monitoring was carried out remotely; from March 2019 until August 2020 (see Figure 1). Longitudinal Fitbit, questionnaire, and accelerometry data were collected at baseline in 2019 and followed up six to nine months later in 2020. Little muscle disease-related functional deterioration was anticipated in under a year [7,8]; thus, any changes measured at follow up were expected to reflect activity behaviour change predominantly.

Participants

In study 1, participants were a purposive sample including 10 people who mobilised with assistance of sticks, crutches, wheeled walkers, and wheelchairs, and 10 who mobilised independently. They were recruited via Muscular Dystrophy UK (MDUK) advertisement (www.musculardystrophyuk.org/progress-in-research/research-projects), newsletter, social media sharing, and word of mouth. Volunteers contacted the King’s College London (KCL) researcher (SRL), who remotely provided study information and followed up with eligibility screening by telephone. Informed consent for study participation, and for contacting their doctor, was obtained by return of signed consent form by post or email. Eligibility was confirmed by physiotherapist (SRL) when sufficient supporting information was provided by the potential participant. In the case of any discrepancy, eligibility was resolved by a neurologist (MR) and/or the potential participant’s GP/consultant. Finally, participants were screened for contraindications to exercise testing by medical clearance from their usual doctor.

In study 2, participants were a convenience sample who opted-in to wear a Fitbit in addition to longitudinal activity monitoring by questionnaire and accelerometry. Participants in the main activity monitoring study [9] were recruited via national muscle disease registries (https://newcastle-muscle.org/clinical) and MDUK. They underwent the same eligibility screening and consent procedures described above, excluding clearance for exercise testing. Screening order was determined chronologically, based on the date of initial contact only, to ensure no selection bias based on demographic information.

In both studies, participants were included if they were UK resident adults with a confirmed diagnosis of inclusion body myositis, myotonic dystrophy, or muscular dystrophy (including facioscapulohumeral dystrophy, limb girdle muscular dystrophy, dysferlinopathy, dystrophinopathy (including manifesting female carriers), or specific congenital myopathies lasting into adulthood). Participants were excluded if they were not UK residents (because KCL research participant insurance only covers UK residents), cognitively impaired, unable to wear an activity monitor, aged
<18 years, did not have a confirmed diagnosis or had muscle weakness resulting from other nervous system dysfunction.

**Study 1**

Participants completed four supervised 10-min tasks (sitting, lying, walking/mobilising, and a cycle ergometer submaximal exercise tolerance test [10]). They wore a Fitbit Charge 2 (Fitbit Inc., San Francisco, CA, discontinued). Fitbit is a wrist-worn, tri-axial accelerometer with continuous optical heart rate monitoring. It syncs with the Fitbit smart phone app. It yields physical activity metrics including steps, calories (converted to intensity in metabolic equivalent minutes (METs) by multiplying Fitbit calories by body weight and dividing by time), active minutes (in bouts of ≥10 min), sleep, and hourly activity (frequency measured by the number of hours between 09:00 and 17:00 daily when ≥250 steps or significant arm movements were registered). Fitbase (an independent research data platform (Small Steps Labs, LLC, San Diego, CA) was used to access minute-by-minute Fitbit data.

For criterion validity, the gold standard comparators each minute during each task were direct observation of physical activity (including independent-observer, video-verified step counting, and vicarious rating of perceived exertion (RPE) [11]) and miniature electrocardiogram heart rate measurement by Polar H10 (Polar Electro, Espoo, Finland). Testing was repeated a week later. Participants wore their Fitbit during the week between testing sessions and gave feedback on their experiences. Feedback was analysed qualitatively using a framework analysis. Satisfactory criterion validity, for both independent and assisted mobility groups, and no excessive or unresolvable negative feedback in terms of acceptability were prerequisites for study 2 to proceed.

**Study 2**

At baseline, participants completed demographic and questionnaire data including anthropometrics, employment, disability (Health Assessment Questionnaire (HAQ) [12]) and quality of life (Individualised Neurological Quality of Life (INQoL) [13]). Self-reported physical activity time and estimated metabolic expenditure was measured using a modified version of the International Physical Activity Questionnaire (IPAQ) [14]. Modifications included equivalent wheelchair activities in the vigorous, moderate, and walking activity categories, and “inactive” replacing the word “sedentary” in the final question.

The comparator measure for Fitbit validity and responsiveness was a research-grade tri-axial accelerometer, GENEActiv (ActivInsights Ltd, Kibolton, UK). The GENEActiv has been validated for adults with progressive muscle diseases [15]. It was posted to participants to wear continuously on their non-dominant wrists for a week, removing it only for washing. The accelerometer sampled continuously at a frequency of 10 Hz. Data were processed automatically using the GGRi package in R (version 3.6.0) (R Foundation for Statistical Computing, Vienna, Austria) [16]. Acceleration data were processed in 1-min epochs in milli-gravitational units (milli-g) with gravitational correction [17]. Overall activity intensity was expressed in mean accelerations (milli-g) per minute each day. Activity frequency was the percentage per day of hourly, non-consecutive movements ≥80 milli-g/min for ≥5 min each hour between 09:00 and 17:00 daily. Intensity cut-points were light ≥30 milli-g/min, moderate ≥100 milli-g/min and vigorous ≥400 milli-g/min [18]. These yielded time in minutes of sleep, inactivity, light activity, and bouts of ≥10 min of moderate and vigorous physical activity (MVPA). Objective physical activity was measured using GENEActiv at baseline and nine months later at follow up.

After baseline, participants were posted a Fitbit Inspire HR (Fitbit Inc., San Francisco, CA; like the Fitbit described above, but lighter weight, waterproof, with contact charger and no altimeter). After receiving their Fitbit, participants set it up (with support as required) and allowed the researcher remote access to their activity data. Participants were asked to wear their Fitbit continuously on their non-dominant wrists, removing it only for washing.

At time point 1 (two months after baseline), participants had worn their Fitbit for a week. Their activity data were collected remotely, and they completed the IPAQ and HAQ for the corresponding week. At time point 2, the following week, Fitbit and questionnaire data collection were repeated. For reliability testing, stability between weeks was established by lack of change in IPAQ and HAQ scores.

At follow up (nine months after baseline), participants’ Fitbit data were collected for a final week. During the follow up week, participants simultaneously wore a GENEActiv and a Fitbit, and all baseline questionnaires were repeated. A single researcher (SRL) was responsible for data collection and analysis. To minimise risk of data entry errors, Fitbit, GENEActiv, and electronic questionnaire data were collected into automatically populated spreadsheets using R (R Foundation for Statistical Computing, Vienna, Austria).

**Statistical analysis**

Statistical analyses were carried out using standardised scripts in R (version 3.6.0) (R Foundation for Statistical Computing, Vienna, Austria). Non-parametric analyses were used if data were not normally distributed. Significance level was set at α = 0.05. Satisfactory thresholds for validity correlation coefficients were based on recommendations for measure evaluation studies [19] (see Table 1).

**Study 1**

A sample size of 20 was calculated to yield 95% power for a predicted criterion correlation coefficient of 0.70 [20] (see Table 1 for a summary of criterion validity and measurement error testing). Categorical agreement of activity level (inactive, light, moderate, and vigorous) between Fitbit and direct observation was tested using Cohen's kappa (see Table 1). Categorical cut-points applied for Fitbit activity level in METs and directly observed RPE each minute were: inactive (<1.5 METs/RPE), light (<3.0 METs/RPE), moderate (<6.0 METs/RPE), and vigorous exceeding these. A kappa of ≥0.75 was considered satisfactory [21]. Absolute measurement error was calculated using the difference between Fitbit and gold standard comparator; this was divided by the gold standard measurement to calculate percentage error. Minutes with missing data were excluded from analyses. Minute-by-minute analyses were carried out for the whole sample, and by mobility subgroup (independent and assisted) and task subgroup (inactive lying and sitting and active walking/mobilising and cycling).

**Study 2**

The desired sample size was 100, as recommended for measure evaluation studies [19]. A retrospective power calculation was performed. Validity, measurement error, reliability, and responsiveness testing are summarised in Table 1. For discriminative validity, wheelchair users were expected to be significantly less active. Subgroup analyses were carried out by disability (according to HAQ reported ambulant and non-ambulant wheelchair user status). Missing accelerometer data of ≤10 min were included in daily means; days with <23 h monitored were excluded from
analyses. Questionnaires were scored using available items; questionnaires with >10% of items missing were excluded from analyses. Participants lost to follow up were excluded from validity and responsiveness analyses.

**Results**

**Study 1**

Fitbit data (1.9% missing) from all 20 participants were included. Fitbit had strong criterion validity compared to gold standard comparators for minute-by-minute steps, heart rate and categorised METs based activity intensity level (rho = 0.97, 0.96 and kappa = 0.90, respectively). Mean systematic, absolute and percentage measurement errors for step count were ±0.8, 2.2, and 9%, respectively. For heart rate bpm, ±3.2, ±3.7, and ±4%, respectively.

Subgroup analysis revealed consistently acceptable criterion validity between mobility and active/inactive task subgroups. Measurement error for heart rate was also consistent between independent and assisted mobility groups. However, Fitbit step error was lower during independent mobility than for those who mobilised with assistance (5% versus 15%). Step error was also lower during inactive lying and sitting compared to active walking and cycling (0% versus 10%). Fitbit heart rate error was lower during lying than the other tasks (2% versus 4–5%) (see Table 2).

Acceptability of the Fitbit was supported by largely positive feedback from participants (see Figure 2). Concerns raised, such as technophobia and dexterity challenges were easily resolved by additional technical support and new equipment for straps and charging. Two minor adverse events were reported; minor skin irritation and bruising from a tight strap.

**Study 2**

Of 511 participants who responded, 147 were chronologically screened for eligibility until the study reached its target of 110 participants. At baseline, 103 participants completed data collection and of these, 95 opted-in to the Fitbit part of the study. At time points 1 and 2, 94 participants completed Fitbit data collection and, of these, 90 completed data collection at follow-up (6.9 months later (range 6.8–7.5)) (see Figure 1). For n = 90 validity correlations, a retrospective power calculation yielded 96% power. Table 3 summarises participants’ demographic information. All participants included had at least five days of useable activity monitoring data.

| Table 1. Measurement properties summary of analyses. |
|-----------------------------------------------|
| **Property**                                      | **Fitbit metric** | **Comparator measure** | **Analysis** | **Threshold** |
| Criterion validity (n = 20)                       | Steps (per minute) | Direct observation (counted steps per minute) | Spearman’s Rho | ≥0.70 |
|                                                   | Heart rate (bpm)  | Polar heart rate (bpm) | Spearman’s Rho | ≥0.70 |
|                                                   | Intensity (METs intensity level) | Direct observation (vicarious RPE intensity level each minute) | Cohen’s kappa (1-way random model, absolute agreement type with linear weighting) | ≥0.75 |
| Measurement error (study 1) (n = 20)              | Steps (per minute) | Direct observation (counted steps per minute) | Bland–Altman, systematic, and absolute error | ≤5% |
|                                                   | Heart rate (bpm)  | Polar heart rate (bpm) | As above | ≤5% |
| Concurrent validity (n = 90)                      | Frequency (%) hourly activity at follow up | GENEActiv frequency of hourly movement (% per day) | Spearman’s Rho | ≥0.70 |
|                                                   | Intensity (steps per day) at follow up | GENEActiv Intensity (daily mean accelerations per minute per day) at follow up | Spearman’s Rho | ≥0.70 |
|                                                   | Intensity (METs per week) at follow up | GENEActiv Intensity (weekly mean accelerations per minute per week) at follow up | Spearman’s Rho | ≥0.70 |
|                                                   | Time (active minutes per week) at follow up | GENEActiv MVPA time (minutes per week) at follow up | Spearman’s Rho | ≥0.70 |
|                                                   | Sleep (hours per night) at follow up | GENEActiv sleep time (hours per night) at follow up | Pearson’s r | ≥0.70 |
| Convergent validity (n = 94)                      | Frequency, steps, METs, active minutes at time point 1 and 2 | Disability (HAQ), self-reported physical activity (modified IPAQ) quality of life (INQoL) | Spearman’s Rho | ≥0.40 |
| Divergent validity (n = 94)                       | Frequency, steps, METs, active minutes at time point 1 and 2 | Height, handness, gender, and age (collected at baseline, also INQoL above) | Spearman’s Rho | <0.20 |
| Discriminative validity                           | Ambulant participants Fitbit metrics at time point 1 and 2 (n = 56) | Wheelchair users Fitbit metrics (frequency, intensity (steps and METs), time (active minutes)) (n = 38) | Independent t-test or Mann–Whitney’s U | p ≤ 0.001 |
| Measurement error (study 2)                       | Time (active minutes) at follow up (n = 90) | GENEActiv MVPA time (minutes per week) at follow up | Bland–Altman, systematic and absolute error | N/A |
| Test–retest reliability (n = 94)                  | Fitbit metrics (frequency, intensity, and time) at time point 1 | Fitbit metrics (frequency, intensity (steps and METs), time (active minutes and sleep)) at time point 2 | Intra-class correlation coefficient (2-way mixed effects consistency model) | ≥0.75 |
|                                                   | HAQ and IPAQ at time point 1 | Change from baseline to follow up in GENEActiv frequency of hourly movement (% per day), intensity (mean accelerations per minute per day), MVPA time (minutes per week) | Mann–Whitney’s U | p > 0.50 |
| Responsiveness (n = 90)                           | Change in frequency, intensity (steps, METs), time (active minutes) from timepoint 1 and 2 to follow up | Change from baseline to follow up in GENEActiv frequency of hourly movement (% per day), intensity (mean accelerations per minute per day), MVPA time (minutes per week) | Area under curve (AUC) of receiver operating characteristic (ROC) curve | ≥0.70 |

bpm: beats per minute; METs: metabolic equivalents; RPE: rating or perceived exertion; %: percentage/percent; HAQ: Health Assessment Questionnaire; INQoL: Individualised Neurological Quality of Life; IPAQ: International Physical Activity Questionnaire; MVPA: moderate and vigorous physical activity.
Fitbit had generally satisfactory concurrent validity. There was a strong correlation between Fitbit and GENEActiv frequency of hourly movements per day (rho = 0.77). Similarly, there were strong correlations between GENEActiv intensity (daily mean hourly movements per day) and Fitbit intensity (steps and METs) (rho = 0.74 to 0.89). Discriminative validity was demonstrated by significant differences in Fitbit activity metrics between ambulant participants and wheelchair users (see Figure 3). The limits of agreement were moderate; however, there was proportional bias (R² = 0.30, linear regression coefficient = 0.55, p < 0.000) indicating significantly more measurement error in Fitbit active minutes at greater MVPA time. There was a 72-min systematic error of Fitbit active minutes in excess of GENEActiv MVPA minutes. The absolute measurement error was 75 min per week.

Test-retest reliability was satisfactory between time point 1 and 2 for Fitbit frequency, intensity (steps and METs) and time (active minutes and sleep) (ICC = 0.94, 0.95 and 0.89, 0.78 and 0.77). Stability between testing weeks was confirmed by no significant change in self-reported physical activity (IPAQ) or disability (HAQ). Satisfactory responsiveness to change was demonstrated for Fitbit frequency, intensity (steps and METs), and time (active minutes) (AUC = 0.85, 0.85, 0.89, and 0.78, respectively) (see Figure 4).

Subgroup analysis of Fitbit steps revealed stronger validity for ambulant participants versus wheelchair users (concurrent validity with GENEActiv rho = 0.76 versus 0.53, convergent validity with disability rho = 0.56 versus −0.48, quality of life rho = −0.32 versus −0.24 and IPAQ score rho = 0.41 versus 0.29). Whereas validity was similar between disability sub-groups for Fitbit frequency, METs, and time. Reliability and responsiveness were similar for all Fitbit metrics between disability sub-groups. Although, there was a trend for slightly stronger measurement properties in the larger, n = 56 ambulant sub-group (see Table 4).

**Discussion**

Fitbit had strong criterion validity and was an acceptable activity monitor. It had broadly satisfactory validity, reliability, and
responsiveness for the assessment of physical activity in adults with progressive muscle diseases. This is in concurrence with other studies reporting satisfactory measurement properties in healthy adults and people with neuromuscular diseases and altered mobility [6,22,23]. However, we found Fitbit metrics, including steps, heart rate, and active minutes have measurement error which must be considered when interpreting data.

Fitbit steps had the strongest validity in our study. Other studies have also reported satisfactory validity for Fitbit steps [6,22,23]. Although, most of these studies had <50 participants and some used waist or ankle-worn devices. Conversely, several larger studies questioned the accuracy of Fitbit steps [24–27]. All these studies tested older Fitbit models and the evolving nature of activity monitoring technology has been noted [22]. However, our findings concur regarding considerable Fitbit steps measurement error. We also noted greater measurement error during active walking and cycling tasks compared to inactive lying and sitting tasks. Furthermore, we found a general trend for slightly inferior measurement properties for the non-ambulant and assisted-mobility subgroups. Variation in gait patterns, mobility aids, and device movement artefacts may impact Fitbit accuracy [6,28]. Fitbit steps are reportedly less accurate, especially when wrist-worn [24], for impaired mobility [25,26] and in short walking tests compared to longer, free-living assessments [27]. Free-living activity monitoring typically lasts >24 h and is, therefore, inclusive of more inactive than active time. Thus, it is likely that Fitbit performs better as a free-living activity monitor, over days or weeks including both active and inactive time, than as a precise step counter for short active tasks, especially for those with mobility impairments.

Fitbit heart rate criterion validity and measurement error were satisfactory. Although the systematic underestimation of 3.7 bpm and improved accuracy in lying mean, the Fitbit should not be used for absolute cardiac monitoring. Activity data derived from heart rate monitoring alone might also be imprecise secondary to reduced heart rate variability in some adults with muscle diseases [29,30]. However, Fitbit algorithms combine heart rate and accelerometer data to estimate physical activity which may temper the effect of individual metric measurement errors.

Fitbit active minutes had unsatisfactory concurrent validity with GENEActiv (rho = 0.68). Fitbit active minutes are derived from combined accelerometry and heart rate data. Therefore, they would not be expected to correlate completely with GENEActiv (solely accelerometer derived) MVPA minutes. In contrast to our findings, a study in healthy adults suggested Fitbit activity time might have stronger validity than steps over a week [27]. Most other studies, in concurrence with our findings, reported good to strong correlations supporting validity of Fitbit active minutes but highlighted a tendency for overestimation [6]. In our study, despite a trend for greater error at higher MVPA times, Fitbit active minutes absolute measurement error of 75 min was nearly equal to the systematic overestimation of 72 min, indicating very little random error. Based on this finding, we can conclude that Fitbit active minutes include light, moderate and vigorous activity and are not synonymous with MVPA time. This caveat must be considered when interpreting Fitbit active minutes data.

Fitbit frequency of hourly movement had satisfactory validity, reliability, and responsiveness in our study. This is a novel finding, as no other studies have evaluated this Fitbit metric, to the authors’ knowledge. Discriminative validity testing revealed Fitbit frequency was potentially more sensitive than GENEActiv to wheelchair users’ hourly movements (6.3% versus 0.0%). Thus, frequency of hourly movement might be a useful activity metric in clinical practice and trials for both able-bodied and disabled adults with neuromuscular disease. Activity frequency also links to activity recommendations such as the World Health Organisation recommendations, which state “every move counts” [31].

Fitbit METs (converted from Fitbit calories) had satisfactory validity, reliability, and responsiveness. These measurement properties were similar across mobility, task, and disability sub-groups.
Table 3. Demographics and baseline information (N = 94).

| Characteristic                  | All                  | Ambulant             | Wheelchair users     |
|--------------------------------|----------------------|----------------------|----------------------|
| Sample size                    | 94                   | 56                   | 38                   |
| Diagnoses                      | 31 LGMD, 24 FSHD, 12 IBM, 10 MD, 10 DM, 7 CM | 16 LGMD, 17 FSHD, 7 IBM, 5 MD, 9 DM, 2 CM | 15 LGMD, 7 FSHD, 5 IBM, 5 MD, 1 DM, 5 CM |
| Gender                         | 46 male, 48 female   | 25 male, 31 female   | 21 male, 17 female   |
| Employment status              | 39 FT, 15 PT, 40 NW  | 29 FT, 10 PT, 17 NW  | 10 FT, 5 PT, 23 NW   |
| Handedness                     | 84 right, 10 left    | 50 right, 6 left     | 34 right, 4 left     |
| Age (years, mean (SD))         | 46.7 (15.1)          | 44.7 (15.1)          | 49.7 (14.8)          |
| Body mass index (mean (SD))    | 25.7 (6.0)           | 25.3 (5.6)           | 26.4 (6.5)           |
| HAQ Disability Index (median (IQR)) | 1.75 (0.91–2.38)   | **1.19 (0.63–1.64)** | **2.50 (2.03–2.75)** |
| Quality of life INQoL (median (IQR)) | 45.3 (34.9–50.5)   | **40.5 (29.0–47.8)** | **48.9 (43.8–53.7)** |
| Self-reported activity IPAQ (median (IQR)) | 2162 (1086–4163) | **3051 (1988–5685)** | **1042 (519–1930)** |
| IPAQ score (estimated weekly metabolic expenditure MET mins) | 113.3 (20.0–233.5) | **149.6 (72.5–327.5)** | **36.7 (0.0–103.1)** |
| GENEActiv accelerometer (mean (SD)) | 27.0 (9.3)       | **32.5 (7.4)**       | **19.7 (6.0)**       |
| Sleep hours per night          | 7.20 (1.78)          | **6.87 (1.62)**      | **7.69 (1.90)**      |
| GENEActiv accelerometer (median (IQR)) | 19.4 (3.7–40.7) | **33.3 (20.4–52.3)** | **0.0 (0.0–11.1)**  |
| Frequency of hourly movement per day (%) | 0.7 (0.0–50.8) | **18.9 (0.0–109.6)** | **0.0 (0.0–0.0)**   |
| Time: moderate and vigorous active minutes per week | **Fitbit (mean (SD))** | 7.18 (1.22) | **7.00 (0.73)** | **7.65 (1.69)** |
| Time: moderate and vigorous active minutes per week | **Fitbit (median (IQR))** | 4608 (1662–7124) | **6614 (5470–8645)** | **1395 (472–2602)** |
| Frequency of hourly movement per day (%) | 42.9 (8.5–62.7) | **58.7 (46.2–70.8)** | **6.3 (0.2–19.2)**  |
| Time: Fitbit active minutes per week | 650.0 (113.3–213.5) | **147.0 (67.6–286.6)** | **14.8 (0.0–36.5)** |
| Time: Fitbit estimated weekly metabolic expenditure (MET mins) | 405 (248–611) | **539.6 (405.0–673.9)** | **240 (124–328)** |

LGMD: limb girdle muscular dystrophy (including 2i,a,b, I, miyoshi myopathy, 1a); FSHD: facioscapulohumeral muscular dystrophy; IBM: inclusion body myositis (including hereditary, GNE myopathy, and sporadic); MD: muscular dystrophy (Becker/Duchenne including manifesting female carriers); DM: myotonic dystrophy (including types 1 and 2); CM: congenital myopathies (including centronuclear myopathy, Bethlem, Emery Dreifuss, and Ullrich muscular dystrophies); FT: full time; PT: part time; NW: not working; SD: standard deviation; IQR: interquartile range; INQoL: Individualised Neurological Quality of Life percentage score (higher values indicative of worse quality of life); HAQ: Health Assessment Questionnaire; IPAQ: International Physical Activity Questionnaire (modified to include activity examples for ambulant and non-ambulant respondents, addition of walking “or equivalent light activities”, sedentary changed to “inactive”); MET: metabolic equivalents.

*p < 0.001, *p < 0.05 significant difference between ambulant and wheelchair users according to Chi square for frequency data, independent t-test and Mann–Whitney’s U for non-parametric data.

Figure 3. Bland–Altman’s plot comparing Fitbit active minutes with GENEActiv moderate and vigorous physical activity (MVPA) minutes.

suggesting that combined multisensor inputs might have tempered individual measurement discrepancies. Fitbit METs had strong agreement with observed physical activity level (kappa = 0.90); however, our study did not test the measurement error of Fitbit metabolic expenditure estimations against an energy criterion. Studies, of healthy adults, have reported a tendency for Fitbit metabolic expenditure overestimation during activity, underestimation at rest, and underestimation over free-living days/weeks [6]. Thus, Fitbit should not be considered an accurate calorimeter. Rather, Fitbit is a smart, multisensor activity monitor capable of relative energy expenditure estimation indicative of physical activity level. In muscle disease, any energy expenditure estimations should be treated with caution, as they are potentially inaccurate because of altered muscle metabolism in some muscle diseases [32].

Fitbit was generally acceptable to study participants. Commercial, wrist-worn multisensor devices, like Fitbits, provide user-friendly activity metrics and appear to be convenient, valid, reliable, and responsive. The Fitbit models used in this study were approximately 50% lower in cost than GENEActiv monitors. However, Fitbit accuracy was questionable, and their activity metric algorithms and thresholds are not publicly available. Whereas research-grade devices are transparent for accessing and interpreting raw data and their accuracy is established. Some research-grade devices, like GENEActiv, use known light, moderate, and vigorous activity thresholds to determine active time. Also, the activity measurement epoch is adjustable (e.g., instead of 10-min bouts, 1-min bouts could be used to align better with current WHO activity recommendations [31]). Other research grade monitors have been refined for accuracy in specific metrics, such as heart rate (Actiheart, Camntech) or steps (StepWatch, Cymatech). Research-grade activity monitors might, therefore, be considered...
Table 4. Measurement properties of Fitbit (home-based activity monitoring results from study 2).

| Property | Fitbit metric | Statistic | 95% CI      | p Value |
|----------|---------------|-----------|-------------|---------|
| Concurrent validity (n = 90, mean data per participant per week) | Frequency (% hourly activity) | Rho = 0.77** | 0.64–0.87 | <0.001 |
| | Intensity (steps per day) | Rho = 0.83** | 0.75–0.88 | <0.001 |
| | Intensity (METs per week) | Rho = 0.72** | 0.65–0.80 | <0.001 |
| | Time (Active minutes per week) | Rho = 0.68* | 0.57–0.79 | <0.001 |
| | Sleep (hours per night) | Rho = 0.55* | 0.31–0.74 | <0.001 |
| Concurrent validity disability sub-analyses (ambulant n = 56, non-ambulant n = 38, mean data per participant per week) | Frequency (ambulant subgroup) | Rho = 0.46* | 0.19–0.68 | <0.001 |
| | Steps (ambulant subgroup) | Rho = 0.76** | 0.60–0.87 | <0.001 |
| | Steps (non-ambulant subgroup) | Rho = 0.53* | 0.20–0.76 | 0.009 |
| | METs (ambulant subgroup) | Rho = 0.63* | 0.47–0.74 | <0.001 |
| | METs (non-ambulant) | Rho = 0.62* | 0.36–0.77 | 0.012 |
| | Time (ambulant subgroup) | Rho = 0.51* | 0.29–0.69 | <0.001 |
| | Time (non-ambulant subgroup) | Rho = 0.47* | 0.19–0.73 | 0.041 |
| Convergent validity with HAQ, INQoL, IPAQ (n = 94, mean data per participant per week) | Frequency (% activity) | Rho = 0.69*** | 0.50 to 0.83 | <0.001 |
| HAQ | Rho = 0.46** | 0.26 to 0.62 | <0.001 |
| INQoL | Rho = 0.43** | 0.10–0.71 | <0.001 |
| IPAQ | Rho = 0.47** | 0.30 to 0.60 | <0.001 |
| Intensity (steps per day) | Rho = 0.33*** | 0.12–0.69 | <0.001 |
| HAQ | Rho = 0.58** | 0.32 to 0.71 | <0.001 |
| INQoL | Rho = 0.40** | 0.18 to 0.62 | <0.001 |
| IPAQ | Rho = 0.56*** | 0.25–0.80 | <0.001 |
| Time (active mins/week) | Rho = 0.67*** | 0.51 to 0.70 | <0.001 |
| HAQ | Rho = 0.49*** | 0.27 to 0.65 | <0.001 |
| INQoL | Rho = 0.46** | 0.21–0.68 | <0.001 |
| IPAQ | Rho = 0.55** | 0.37 to 0.68 | <0.001 |
| Frequency (ambulant group) | HAQ | Rho = 0.35* | 0.09 to 0.57 | 0.013 |
| INQoL | Rho = 0.16 | −0.13 to 0.42 | 0.272 |
| IPAQ | Rho = 0.51** | −0.21 to −0.76 | 0.001 |
| Steps (ambulant subgroup) | HAQ | Rho = 0.34* | −0.07 to 0.57 | 0.035 |
| INQoL | Rho = 0.02 | −0.29 to 0.37 | 0.901 |
| IPAQ | Rho = 0.56** | −0.30 to −0.77 | <0.001 |
| Steps (non-ambulant group) | HAQ | Rho = 0.32** | −0.09 to 0.57 | 0.026 |
| INQoL | Rho = 0.41** | 0.12–0.64 | 0.026 |
| IPAQ | Rho = 0.56** | −0.23 to −0.67 | 0.001 |
| METs (ambulant subgroup) | HAQ | Rho = 0.18 | −0.12 to 0.47 | 0.219 |
| INQoL | Rho = 0.05 | −0.23 to 0.30 | 0.726 |
| IPAQ | Rho = 0.13 | −0.19 to 0.43 | 0.375 |
| METs (non-ambulant) | HAQ | Rho = 0.01 | −0.30 to 0.28 | 0.935 |
| INQoL | Rho = 0.14 | −0.14 to 0.41 | 0.395 |
| IPAQ | Rho = 0.11 | −0.20 to 0.39 | 0.494 |
| Time (ambulant subgroup) | HAQ | Rho = 0.58** | −0.36 to −0.75 | <0.001 |
| INQoL | Rho = 0.39* | −0.07 to 0.64 | 0.005 |

(continued)
“specialists”, whereas commercial, multisensor activity monitors, like Fitbit, might be considered “generalists” for broader activity estimation. Both commercial and research-grade activity monitoring technology is regularly evolving; future models have the potential to address shortcomings of their predecessors.

The key strengths of this study included thorough exploration of multiple measurement properties, inclusion of wheelchair users (previously underrepresented in neuromuscular disease studies of physical activity [33]) and study design inclusive of people with a broad range of function, activity levels, ages, occupations, and locations across the UK. Study 2 also included a larger sample size than previous objective measurement studies in this population [1]. However, despite inclusion of participants from a range of backgrounds, generalisability in terms of co-morbidities, ethnicity, socioeconomic status, and education level is unknown because these data were not formally collected. Other limitations include recruitment of volunteers introducing potential selection bias, and a single researcher being responsible for data collection and analysis. However, risk of selection bias was minimised by the chronological eligibility screening and risk of information/researcher bias was minimised by independent verification of collected data and using automatic data processing and analyses. The study could have been improved by comparing Fitbit with an energy expenditure criterion. A healthy comparison group would have helped further clarify Fitbit measurement properties between populations and a larger sample size of adults with progressive muscle diseases would have allowed subgroup analyses by diagnosis and increased generalisability.

The smart, multisensor Fitbit activity monitor had satisfactory validity, reliability, responsiveness, and acceptability for the assessment of physical activity in adults with progressive muscle diseases. Smart, multisensor monitoring using Fitbit is suitable for the assessment of free-living physical activity intensity, frequency, and time over days or weeks. It might be generalisable for use with able-bodied and disabled adults with a range of comparable conditions. However, Fitbit should not be considered an exact step counter, heart rate monitor, or calorimeter and Fitbit active minutes are not synonymous with MVPA time.

Acknowledgements
We would like to thank The John Walton Muscular Dystrophy Research Centre at Newcastle University and the Muscular Dystrophy UK charity for their help with advertising the study.

Authors contributions
Dr Mark Ashworth, Dr Claire White, and Dr Michael Rose: PhD supervision, monitoring progress at inception, implementation,
analysis, and writing up. Sarah Roberts-Lewis: guarantor with responsibility for design, implementation, honesty, accuracy, and writing up.

**Disclosure statement**
The authors report no conflicts of interest.

**Funding**
This work is supported by Medical Research Council (MRC) International Centre for Genomic Medicine in Neuromuscular Disease PhD studentship at King’s College London (MR/K50130X/1).

**ORCID**
Sarah F. Roberts-Lewis [http://orcid.org/0000-0001-5579-2701](http://orcid.org/0000-0001-5579-2701)

**References**

[1] Roberts-Lewis SF, Rose MR, White CM, et al. Measurement of physical activity in adults with muscular dystrophy: a systematic review. J Phys Activ Res. 2019;4(1):1–25.

[2] Mercuri E, Mayhew A, Muntoni F, et al. Towards harmonisation of outcome measures for DMD and SMA within TREAT-NMD; report of three expert workshops: TREAT-NMD/ENMC workshop on outcome measures, 12th–13th May 2007, Naarden, The Netherlands; TREAT-NMD workshop on outcome measures in experimental trials for DMD, 30th June–1st July 2007, Naarden, The Netherlands; conjoint Institute of Myology TREAT-NMD meeting on physical activity monitoring in neuromuscular disorders, 11th July 2007, Paris, France. Neuromuscul Disord. 2008;18(11):894–903.

[3] Jimenez-Moreno AC, Newman J, Charman SJ, et al. Measuring habitual physical activity in neuromuscular disorders: a systematic review. J Neuromuscul Dis. 2017;4(1):25–52.

[4] Prahm KP, Witting N, Vissing J. Decreased variability of the 6-minute walk test by heart rate correction in patients with neuromuscular disease. PLoS One. 2014;9(12):e114273.

[5] Nightingale TE, Rouse PC, Thompson D, et al. Measurement of physical activity and energy expenditure in wheelchair users: methods, considerations and future directions. Sports Med Open. 2017;3(1):10.

[6] Feehan LM, Geldman J, Sayre EC, et al. Accuracy of Fitbit devices: systematic review and narrative syntheses of quantitative data. JMIR Mhealth Uhealth. 2018;6(8):e10527.

[7] Andersen G, Dahlqvist JR, Vissing CR, et al. MRI as outcome measure in facioscapulohumeral muscular dystrophy: 1-year follow-up of 45 patients. J Neurol. 2017;264(3):438–447.

[8] Leung DG, Bocchieri AE, Ahalwatt S, et al. Longitudinal functional and imaging outcome measures in FKRP limb-girdle muscular dystrophy. BMC Neurol. 2020;20(1):196.

[9] Roberts-Lewis SF, Ashworth M, White CM, et al. COVID-19 lockdown impact on the physical activity of adults with progressive muscle diseases. BMJ Neurol Open. 2021;3(1):e000140.

[10] Gappmaier E. The submaximal clinical exercise tolerance test (SXTT) to establish safe exercise prescription parameters for patients with chronic disease and disability. Cardiopulm Phys Ther J. 2012;23(2):19–29.

[11] Borg G. Psychophysical scaling with applications in physical work and the perception of exertion. Scand J Work Environ Health. 1990;16(Suppl.1):55–58.

[12] Bruce B, Fries JF. The Stanford Health Assessment Questionnaire: a review of its history, issues, progress, and documentation. J Rheumatol. 2003;30(1):167–178.

[13] Vincent KA, Carr AJ, Walburn J, et al. Construction and validation of a quality of life questionnaire for neuromuscular disease (INQoL). Neurology. 2007;68(13):1051–1057.

[14] Craig CL, Marshall AL, Sjostrom M, et al. International Physical Activity Questionnaire: 12-country reliability and validity. Med Sci Sports Exerc. 2003;35(8):1381–1395.

[15] Jimenez-Moreno AC, Charman SJ, Nikolenko N, et al. Analyzing walking speeds with ankle and wrist worn accelerometers in a cohort with myotonic dystrophy. Disabil Rehabil. 2018;10:1–7.

[16] van Hees VT, Sabia S, Anderson KN, et al. A novel, open access method to assess sleep duration using a wrist-worn accelerometer. PLoS One. 2015;10(11):e0142533.

[17] van Hees VT, Gorzelniak L, Leon ECD, et al. Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. PLoS One. 2013;8(4):e01691.

[18] van Hees VT, Fang Z, Langford J, et al. Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents. J Appl Physiol (1985). 2014;117(7):738–744.

[19] Mokkink LB, Prinsen CAC, Patrick DL, et al. COSMIN methodology for systematic reviews of patient-reported outcome measures (PROMs): user manual. In: Mokkink LB, editor. COSMIN methodoloy for systematic reviews of patient-reported outcome measures (PROMs): user manual. Version 1.0 ed. Amsterdam: Amsterdam Public Health Research Institute; 2018.

[20] Mondal H, Mondal S. Sample size calculation to data analysis of a correlation study in Microsoft Excel: a hands on guide with example. Int J Clin Exp Physiol. 2016;3(4):180–189.

[21] Sim J, Wright CC. The kappa statistic in reliability studies: use, interpretation, and sample size requirements. Phys Ther. 2005;85(3):257–268.

[22] Evinson KR, Goto MM, Furberg RD. Systematic review of the validity and reliability of consumer- wearable activity trackers. Int J Behav Nutr Phys Act. 2015;12:159.

[23] Tedesco S, Sica M, Ancillao A, et al. Validity evaluation of the Fitbit Charge2 and the garmin vivosmart HR+ in free-living environments in an older adult cohort. JMIR Mhealth Uhealth. 2019;7(6):e13084.

[24] Chu AH, Ng SH, Paknezhad M, et al. Comparison of wrist-worn Fitbit flex and waist-worn ActiGraph for measuring steps in free-living adults. PLoS One. 2017;12(2):e0172535.

[25] Treacy D, Hassett L, Schurr K, et al. Validity of consumer- wellbeing devices: systematic review and narrative syntheses of quantitative data. JMIR Mhealth Uhealth. 2018;6(8):e10527.

[26] Floegel TA, Florez-Pregonero A, Hekler EB, et al. Validation of consumer-based hip and wrist activity monitors in older adults with varied ambulatory abilities. J Gerontol A Biol Sci Med Sci. 2017;72(2):229–236.
[27] Brewer W, Swanson BT, Ortiz A. Validity of Fitbit’s active minutes as compared with a research-grade accelerometer and self-reported measures. BMJ Open Sport Exerc Med. 2017;3(1):e000254.

[28] O’Connell S, O’Laighin G, Quinlan LR. When a step is not a step! Specificity analysis of five physical activity monitors. PLoS One. 2017;12(1):e0169616.

[29] Yotsukura M, Fujii K, Katayama A, et al. Nine-year follow-up study of heart rate variability in patients with Duchenne-type progressive muscular dystrophy. Am Heart J. 1998;136(2):289–296.

[30] Marca GD, Frusciante R, Scatena M, et al. Heart rate variability in facioscapulohumeral muscular dystrophy. Funct Neurol. 2010;25(4):211–216.

[31] WHO. Guidelines on physical activity and sedentary behaviour. Geneva: World Health Organization; 2020.

[32] Cruz Guzman Odel R, Chavez Garcia AL, Rodriguez-Cruz M. Muscular dystrophies at different ages: metabolic and endocrine alterations. Int J Endocrinol. 2012;2012:485376.

[33] Jones K, Hawke F, Newman J, et al. Interventions for promoting physical activity in people with neuromuscular disease. Cochrane Database Syst Rev. 2021;5:CD013544.