BMJ Open

How is the usability of commercial activity monitors perceived by older adults and by researchers? A cross-sectional evaluation of community-living individuals

Lena M Hofbauer , Francisca S Rodriguez

To cite: Hofbauer LM, Rodriguez FS. How is the usability of commercial activity monitors perceived by older adults and by researchers? A cross-sectional evaluation of community-living individuals. *BMJ Open* 2022;**12**:e063135. doi:10.1136/bmjopen-2022-063135

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2022-063135).

Received 22 March 2022
Accepted 13 October 2022

Research Group Psychosocial Epidemiology & Public Health, German Center for Neurodegenerative Diseases Site Rostock/Greifswald, Greifswald, Germany

Correspondence to
Lena M Hofbauer;
lena.hofbauer@dzne.de

ABSTRACT

Objectives Using commercial activity monitors may advance research with older adults. However, usability for the older population is not sufficiently established. This study aims at evaluating the usability of three wrist-worn monitors for older adults. In addition, we report on usability (including data management) for research.

Design Data were collected cross-sectionally. Between-person of three activity monitor type (Apple Watch 3, Fitbit Charge 4, Polar A370) were made.

Setting The activity monitors were worn in normal daily life in an urban community in Germany. The period of wear was 2 weeks.

Participants Using convenience sampling, we recruited N=27 healthy older adults (≥60 years old) who were not already habitual users of activity monitors.

Outcomes To evaluate usability from the participant perspective, we used the System Usability Scale (SUS) as well as a study-specific qualitative checklist. Assessment further comprised age, highest academic degree, computer proficiency and affinity for technology interaction. Usability from the researchers' perspective was assessed using quantitative data management markers and a study-specific qualitative check-list.

Results There was no significant difference between monitors in the SUS. Female gender was associated with higher SUS usability ratings. Qualitative participant-usability reports revealed distinctive shortcomings, for example, in terms of battery life and display readability. Usability for researchers came with problems in data management, such as completeness of the data download.

Conclusion The usability of the monitors compared in this work differed qualitatively. Yet, the overall usability ratings by participants were comparable. Conversely, from the researchers' perspective, there were crucial differences in data management and usability that should be considered when making monitor choices for future studies.

STRENGTHS AND LIMITATIONS OF THIS STUDY

⇒ This study reports on a direct comparison of usability of three popular models of activity monitors for older non-habitual users.
⇒ Usability data reported is of high ecologic validity, that is, based on use in daily life.
⇒ Results also give valuable insights into usability from a researcher perspective, which will allow for evidence-based decision making in future research designs.
⇒ Generalisability of results is hampered by the specific use case under investigation (ie, 2 weeks of monitor wear without ongoing mobile device synchronisation).

BACKGROUND

Physical activity is linked to better physical health as well as improved cognition and functional capacity in older adults.[1] Thus, researchers have an interest in accurately measuring physical activity in this population. In the past, researchers have relied on self-reports. Due to social desirability bias and inaccurate recall, especially among people with cognitive impairments, these have limited reliability.[2] Commercially available activity monitors have opened up new possibilities. Commonly wrist-worn, these monitors allow for continued and unobtrusive wear and real-time recording of activity in daily life.

Commercial activity monitors record a range of activity indicators. The most commonly used metric is step counts.[3] Reviews find an acceptable-to-high accuracy of step estimates across a variety of activity monitors,[3 4] also for older adults.[5] Not only are steps the least error-prone activity metric,[4] they allow clear-cut recommendations for physical activity.

To additionally determine the intensity of activity, heart rate (HR) measurements can be used. HR measurements are traditionally performed with an ECG. However, wearable ECGs are not suitable for everyday use.[6] Commercial activity monitors may provide researchers with a good alternative. The loss in reliability compared with traditional ECG is minimal and can be neglected as long as heart disease is not the focus of the investigation.[4 7]

While the accuracy of activity monitors has been validated, their usability for older adults has not been sufficiently established. One study with habitual users of activity monitors found no age effect.[8] Yet, habitual user reports may not be representative. Two recent systematic reviews included a total of 20 studies with older adults, including with non-habitual users. The results indicate that activity monitor use is feasible for older adults.[5 9] Indeed, it was reported that the majority of older adults described the use of fitness monitors as 'easy' and wore them reliably for several days.[5] Only two studies directly assessed usability in normal daily life for older adults. Results were mixed. While a commercial monitor was preferred over a research-grade device in a study lasting 1 day,[10] in the second study, five out of nine users reported not wanting to continue using the activity monitors after 2 weeks.[11]

Further, the usability from the researcher perspective must be considered. From the research perspective, data management is central to usability. Given that the software of commercial activity monitors varies between brands and models and is updated with little transparency,[12 13] it is unclear whether research requirements are met.

Aim

Thus, we aimed at evaluating the usability of three commercially available activity monitors in older adults. More specifically, we studied two aspects, namely (1) usability from the participants' perspective and (2) usability from the research perspective (including data management).

METHODS

Participants

The study took place in an urban community in Germany. We recruited community-dwelling older adults via convenience sampling. Inclusion criteria were (1) ≥60 years old, (2) sufficient vision, hearing and language comprehension, and (3) not being a habitual user of an activity monitor. Reasons for exclusion were (1) insufficient capacity to participate in the questionnaires and interview due to dementia, mental illness and/or severe physical limitation, (2) having a motor disorder that would interfere with the use of a monitor or (3) having a cardiovascular disease that would seriously interfere with the monitors' optical pulse measure. From those who were interested, no one had to be excluded from participation.

In total, n=27 participants were included in the study. The average age was 72.96 (SD: 6.32). For more details see table 1 and figure 1. One participant discontinued wearing the monitor after 4 days due to skin irritation. As this participant completed all study material, his/her data were included in the analysis. One participant did not provide data on the Affinity for Technology Interaction (ATI) score. Excluding the data of this participant made no difference to results, thus, we decided to include his/her data in the analysis.

Table 1 Sample demographics

	Descriptive statistics
Gender n (%)	
Male	9 (33.33)
Female	18 (66.66)
Age	
Mean (SD)	72.96 (6.32)
Range	60–87
Level of education n (%)	
Primary school (I)	1 (3.7)
Secondary school (Realschule, III)	6 (22.22)
Secondary school (Gymnasium, IV)	1 (3.7)
Vocational school (V)	1 (3.7)
University of applied science (VI)	13 (48.15)
University (VII)	5 (18.52)
CPQ Score	
Mean (SD)	20.30 (7.23)
Range	5.78–30
ATI Score*	
Mean (SD)	3.59 (SD: 1.09)
Range	1.66–6
Days between monitor drop-off and pick-up	
Mean (SD)	15.33 (2.70)
Range	4–20
Activity monitor wear days (recording-based)	
Mean (SD)	12.05 (4.29)
Range	1–16

Descriptive statistics of demographic information, activity monitor wear information and psychometric Questionnaire Scores.
*Data available for n=26.
ATI, Affinity for Technology Interaction Scale; CPQ, Computer Proficiency Questionnaire.

Researchers

The authors of this study provided the researcher perspective on the qualitative measures after reaching consensus in discussion.

Procedure

Each participant attended two sessions in his/her home, a public place, or at the research institute. In the first session, participants completed questionnaires and received their assigned activity monitor. Participants received only minimal instructions on monitor use. They were instructed how to put on and take off the monitor, how to activate the display, how to recognise that the monitor needed to be charged, and how to charge the monitor. The researcher provided them with contact details so that they could reach out if they needed assistance during the wear-period (none did). Participants were asked to wear the monitor as often as possible unless

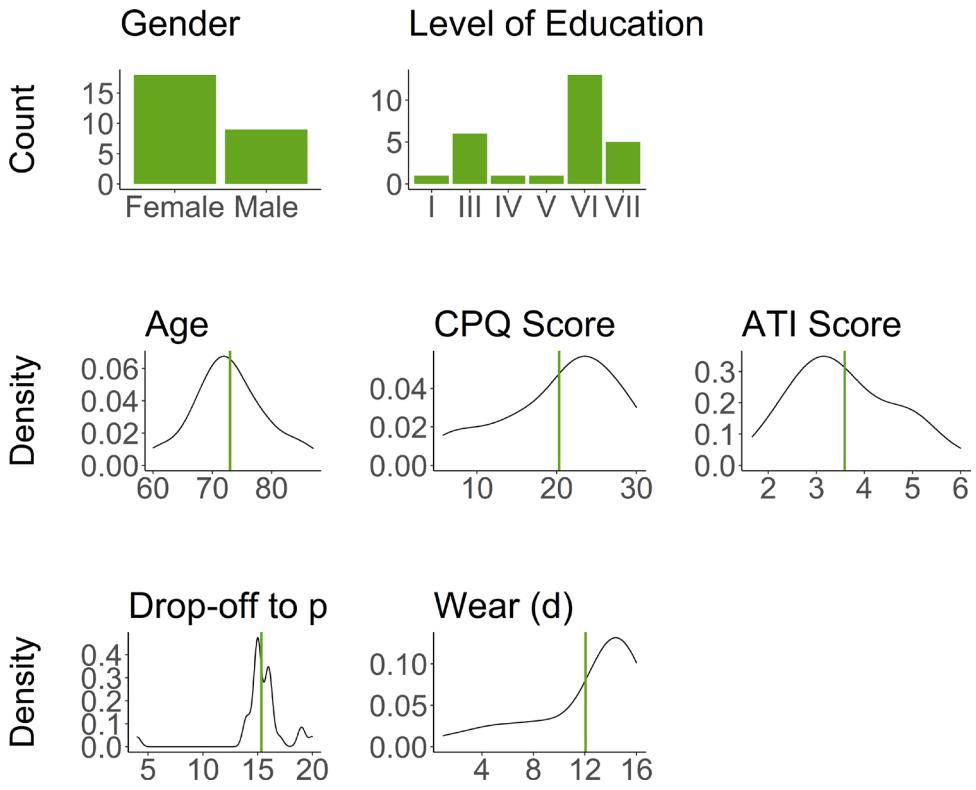

Figure 1 Distribution/counts of demographic variables in the sample. Distribution plots provided for continuous variables, counts for categorical variables. The mean value is marked with a vertical line in distribution plots. I, primary school; III, secondary school (Realschule); IV, secondary school (gymnasium); V, vocational school; VI, university of applied science; VII, university; (d), days.

showering/bathing. However, they were also advised to remove the monitor in case of discomfort. For comparability with the previous study by Fausset *et al*,[11] the target wear-period was 2 weeks. In the second session, the monitors were collected and participants reported on usability.

Activity monitors

We used three popular activity monitors: (1) Apple Watch 3, (2) Fitbit Charge 4 and (3) Polar A370 (referred to as Apple/Fitbit/Polar in the following). Each participant received one activity monitor. Monitor distribution had a pseudorandom order. That is, random order was deviated from whenever participants were living together and would have otherwise received the same monitor. This was done once to ensure that the participant's feedback would not be influenced by the experience of the other. To protect participants, data of a standard avatar (male, 65 years, 176 cm tall, weight 86.5 kg, right-handed) were entered when setting up the monitors. The mobile phone that was connected with the monitor remained with the researchers, that is, monitors were out of reach of the Bluetooth connection during wear.

Activity monitor data

When activity monitors were returned by the participants, they were synced with the respective app on the mobile phone. While Apple and Fitbit provided access to comprehensive data downloads, Polar did not. For this reason, we were unable to review Polar data beyond documenting the number of days for which data were recorded.

For Apple and Fitbit data, in cases where there were more than one HR measurement per minute, we calculated average beats per minute (bpm).[14 15] Step data were summed up for per minute values. To avoid bias due to the putting on of the activity monitor in the mornings, we removed the first five measurements of each day.

In accordance with previous sensor research,[16] we took a multidimensional approach to data quality. Our quality markers on the participant level were the following:

Date volume

We identified the number of days recorded and calculated the average number of measurements per day for each participant.

Correctness

Instead of using a reference measurement, we approximated data correctness by identifying implausible values. That is, we flagged HR measurements which differed more than 10 bpm from the rolling average over the last three measurements ('HR jumps') and calculated the percentage of total measurements flagged. For step data, we flagged any measure of more than 200 steps per minute as implausible ('high speed'); this definition has been used previously.[17]

	Apple	Fitbit	Polar
Participant Perspective			
Wristband On & Off	+	+	-
Comfortable Wear	+/-	+/-	+/-
Unobtrusive Wear	+	-	+
Battery Life	--	+	+
Charging Cable & Connector	+	+	--
Display Readability	+	--	-
Signal quality	+	-	+
Researcher Perspective			
Data management	-	+	--
Monitor & Account Set-Up	-	+	+
Synchonisation	-	+	+
Erasing Backups	++	-	+
Reset & Reuse	-	+	+

Figure 2 Visual synthesis of participant and researcher reports. Synthesis takes into account quantitative and qualitative data. +represents a positive evaluation, –a negative evaluation, ++ and --represent particularly positive or negative evaluations. ± represents a mixed evaluation.

Completeness

We flagged instances in which the time lag between a measurement and the previous measurement was 6–9 min long ('drop instances'). The upper limit was chosen because it is unlikely that the activity monitor was taken off and put down for less than 10 min. The lower limit was chosen because Apple HR measurements were recorded every 5 min. We then calculated which percentage of total observations were drop instances. Further, we identified the number of days for which there was step data but no HR data and vice versa ('missing days') as this would suggest a measurement, synchronisation or export error rather than non-wear.

Questionnaires

Demographic characteristics

Participants reported their gender (male, female, non-binary), their level of education (elementary school, secondary school (German Hauptschule), secondary school (German Realschule), secondary school (German Gymnasium), vocational school, University of Applied Sciences, university or none) and their age.

Affinity for Technology Interaction Scale

The ATI[18] surveys the tendency to engage in intensive technology interaction. For nine items, participants rate their agreement on a Likert scale of 1–6. Three items are inversely scored. The average of the item scores is calculated.

Computer Proficiency Questionnaire

The Computer Proficiency Questionnaire (CPQ)[19] contains 33 questions grouped in 6 domains: (1) computer basics, (2) printing, (3) communication, (4) internet, (5) scheduling software and (6) multimedia use. Respondents rate how easily they could complete a given task on a 5-point scale. The total score is calculated by first averaging scores within a subscale and then taking the average of subscale averages.

System Usability Scale

The System Usability Scale (SUS)[20] contains 10 questions (Likert scale of 1–5) about the usability of a system. Item scores are added and multiplied by 2.5. The score indicates the user-friendliness in values up to 100.

Usability Checklist for Participants (P-checklist)

A checklist was designed for this study (see online supplemental table S1) that allows for a detailed report of usability from the participants' perspective. The checklist consists of a 13 items in yes/no format (eg, 'Were there any situations in which wearing the [activity monitor] was uncomfortable or hindered you in your activity?'), eight text-based items for participants to expand on their answers, and one open-ended item to allow for further comments.

Usability Checklist for Researchers (R-checklist)

The specifically designed R-checklist (see online supplemental table S2) comprised 13 items. On three of these, researchers gave a qualitative evaluation of the quality of the data in terms of data volume, completeness and correctness. On eight items, they rated the ease of use from one to ten (eg, 'How effortful was it to set up a user account for [the activity monitor]?'). On the 12th item, researchers reported how long the manually deleted participant data remained with the provider. These questions are used to guide a text-based evaluation of researcher usability on the final 13th item.

Analysis

Data analysis was conducted using R (V.4.0.2) in RStudio (V.1.2.5042).[21] We used a significance level of $p<0.05$.

To test for group differences at baseline, we used one-way analysis of variances (ANOVAs) (Age, CPQ Score, ATI Score) and Fisher's exact tests (FETs) (Gender, Level of Education).

Correlation analyses were performed to determine whether any baseline demographics or psychometric test scores were associated with participant-reported usability (SUS). Pearson's correlations were used for continuous variables, Spearman's correlations for categorical variables (gender, level of education).

To determine whether there was a statistically significant difference between activity monitors in system usability according to the participant report on the SUS, we used a between-groups one-way ANOVA. Due to the presence of a potential outlier in the Polar group, as determined by residual inspection, non-parametric comparison by

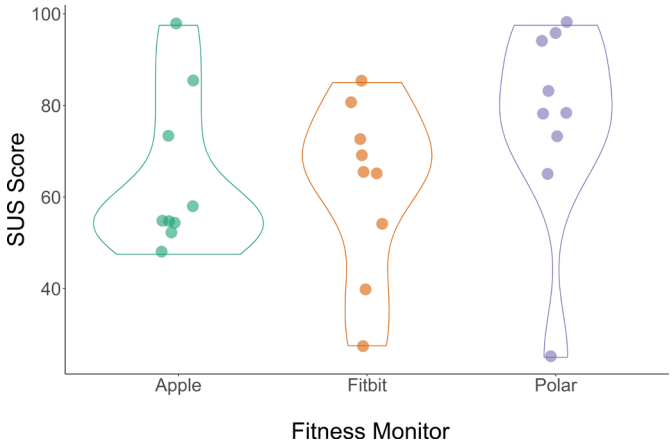

Figure 3 System Usability Scale (SUS) scores by activity monitor.

Kruskal-Wallis testing was also completed. The results did not differ, hence we present the parametric test results.

In order to contextualise SUS scores, we further calculated one sample t-tests and compared the group values to two available norms: (1) the median score (68) observed across usability research[22] and (2) the grand average SUS score of 64.3 from a range of previous investigations of habitual monitor users of varying ages.[8] To assess whether there were statistically significant differences between activity monitors for yes/no items on the P-checklist, we completed FETs.

Text-based reports on the P-checklist and the R-checklist were evaluated qualitatively.

Patient and public involvement

Patients and the public were not involved in the design, reporting or dissemination plans of this research. Patients and community service programmes aided recruitment by disseminating study information.

RESULTS

In this between-person comparison, there were no significant differences in any baseline characteristics between the groups (gender: FET p=0.268; age: F (2, 24) = 1.39, p=0.268, η^2=0.10; education: FET p=0.824; CPQ score: F (2,24) = 0.83, p=0.447, η^2=0.06; ATI score: F (2,23) = 1.27, p=*0.300*, η^2=0.10). Differences in perceived usability of the three activity monitors are reported below. For a visual synthesis, see figure 2.

Table 2 Results of correlation analyses between individual differences and SUS Scores

Variable	Correlation with SUS	95% CI	P value
Age	0.12	−0.27 to 0.48	0.535
Gender	**−0.46**	−0.71 to 0.01	**0.017**
Education level	−0.02	−0.40 to 0.35	0.906
CPQ Score	−0.15	−0.50 to 0.24	0.450
ATI Score	−0.12	−0.38 to 0.39	0.546

Significant result in bold.
.ATI, Affinity for Technology Interaction Scale; CPQ, Computer Proficiency Questionnaire; SUS, System Usability Scale.

Usability from the participants' perspective

The only significant association was between SUS score and gender (Spearman's r=−0.46, p=0.017; table 2), indicating that SUS scores reported by men were lower than those reported by women. There was no significant difference in SUS scores between activity monitors (F (2, 24) = 1.41, p=0.265, η^2=0.10; figure 3) and no differences between activity monitors on any of the P-checklist items (table 3). SUS score for the different monitors did not differ significantly from either the norm values for usability in general (Apple: t (8) = −0.02, p=0.982, d=0.01; Fitbit: t (8) = −0.93, p=0.379, d=0.31; Polar: t(8) = 1.13, p=0.291, d=0.38) or usability of activity monitors specifically (Apple: t (8) = −0.67, p=0.519, d=0.22; Fitbit: t (8) = −0.34, p=0.746, d=0.11; Polar: t(8) = 1.63, p=0.142, d=0.54).

Qualitative analysis revealed some common themes across activity monitors. Nine participants reported removing the monitor at night. Other activities for which the monitors were reportedly removed were gardening, swimming, paintwork and housework involving water. Five participants used the monitors to check their HR data, three reported tracking steps and exercise. Three believed that the activity tracking was not of interest to them or only of use to very physically active people. Monitor-specific results were the following:

Apple

Participants reported charging the activity monitor daily or every 2 days. One participant complained that the '(b)attery runs out far too quickly'. One participant had to terminate the trial early after developing a rash from the wristband. One participant reported that the watch synced with their personal phone, which should not have been possible.

Fitbit

Participants reported charging the monitor between one and three times during the 2 weeks, two times being the most commonly (5/9 participants) reported frequency. Three participants reported that the wristband irritated their skin. Three reported issues with the display: Two reported the display being too dark to read in daylight, one of them stating that the 'display (was) almost unreadable in daylight'. Two would have preferred bigger font size. One further participant appeared to struggle with display; (s)he had trouble accessing the battery status information and could no longer use the display at all after charging.

Polar

The frequency of charging the wrist band varied from two times during the 2-week period to every day. The most commonly reported charging frequency was two times.

Table 3 Usability from the participants' perspective

SUS	Activity monitor	Mean (SD)	
	Apple	62.22 (18.60)	
	Fitbit	64.17 (17.05)	
	Polar	76.39 (22.26)	
P-Checklist		**Y/N (n)**	**P value***
Able to put on/take off (the activity monitor) without any problems	Apple	8/1	0.751
	Fitbit	9/0	
	Polar	7/2	
Easily remembered to put (the activity monitor) back on after taking it off	Apple	9/0	0.999
	Fitbit	9/0	
	Polar	8/1	
Wearing (the activity monitor) was generally comfortable	Apple	7/2	0.842
	Fitbit	8/1	
	Polar	6/3	
Clear what signals meant	Apple	6/3	0.425
	Fitbit	3/6	
	Polar	6/3	
If necessary, able to turn off signals	Apple	2/2/5†	0.220
	Fitbit	1/1/7†	
	Polar	0/0/9†	
Used any of (the activity monitor's) functions besides wearing it	Apple	3/6	0.999
	Fitbit	2/7	
	Polar	2/7	
Had to charge (the activity monitor)	Apple	8/1	0.999
	Fitbit	9/0	
	Polar	9/0	
Any difficulty charging (the activity monitor)	Apple	1/8	0.751
	Fitbit	0/9	
	Polar	2/7	
Had to take off (the activity monitor) for comfort	Apple	1/8	0.281
	Fitbit	4/5	
	Polar	4/5	
Any situations in which wearing (the activity monitor) was uncomfortable or hindered an activity	Apple	3/6	0.121
	Fitbit	7/2	
	Polar	3/6	
Any instances in which using (the activity monitor) was frustrating	Apple	3/6	0.999
	Fitbit	3/6	
	Polar	2/7	
Bothered by any signals (whether sound or vibration)	Apple	0/9	NA
	Fitbit	0/9	
	Polar	0/9	
Got help from anyone aside from researchers	Apple	0/9	0.999
	Fitbit	0/9	
	Polar	1/8	

Descriptive statistics of SUS Scores and P-Checklist Items.
*P values according to Fisher's exact tests
†For this item there was a third answer option 'not necessary'.
.NA, not available; P-Checklist, Participant Checklist; SUS, System Usability Scale.

Two participants had trouble putting on and/or taking off the wristband. Two further participants complained about the fastening mechanism, saying that the eyelets were 'too fiddly' and that wrist straps were 'somewhat annoying'. Two complained about skin irritation. One participant reported having to get assistance for charging the monitor. Two participants expressed dissatisfaction with the small connector for the charging cable, one of which stated: 'The connector of the charging cable is very small and probably difficult to handle for elderly people with handicaps'. One participant pointed out the overall poor make of the wristband, which was 'detaching from the watch-part'. Two participants reported having issues with the display, either it being incidentally unresponsive or it changing appearance irreversibly. Two participants reported that the LED-lights caused nuisance by not switching off even after the monitor had been put down.

Usability from the Researchers' Perspective

Monitor and account setup was unproblematic for all monitors. Instances of temporary synchronisation errors occurred with Fitbit and Polar monitors. These could be resolved. Handling the Apple monitor was most time consuming. Most time took the set up and deleting of data. For Polar and Fitbit, data were deleted by closing the account. This required email confirmation. Resetting the monitors to allow for reuse was time-consuming for Apple, error-prone in Polar, and went most smoothly with Fitbit. Apple did not store any server backups. Fitbit stored backups for 30–90 days, Polar for 14 days. Fitbit and Polar required new email accounts for every participant. Apple users were connected to the same mobile phone account, which resulted in major synchronisation problems when multiple monitors were worn in overlapping time periods. In these cases, attempting to synchronise one monitor after the other failed and led to massive data loss for all monitors involved.

Data management

Step and HR data were collected for all Fitbit and Polar users. In addition, HR data were available for 2/9 and step data for 4/9 Apple users. For Polar users, there were zero missing days. For Fitbit and Apple users, some days were missing, both in terms of steps and in terms of HR. Compared with Apple monitors, Fitbit monitors recorded more data, there were fewer drop instances, fewer instances of improbably high step speed, and more HR jumps. For more details, see table 4.

The Fitbit was able to store around 2 weeks of HR and step data internally. Further, HR data also contained values indicating confidence in the estimates. However, in some instances, either HR or step days were lost inexplicably. Apple monitors recorded a maximum of 9 days of step and HR data simultaneously). However, when the data were read out after 31 May 2021, no HR data were contained in the export. Based on the visual inspection, there appeared to be only few gaps in the Polar data. Beyond this, we are unable to comment on the quality

Table 4 Descriptive statistics for data management markers

	Apple	Fitbit	Polar
Step data			
Participants w/ any data	n=4	n=9	n=9
Days recorded			
Mean (SD)	2.78 (3.56)	13.78 (2.54)	12.44 (4.39)
Range	0–9	9–16	1–15
Obs. minutes per day			
Mean (SD)	90.18 (60.16)	599.35 (159.76)	NA
Range	36.50–160.43	397.87–873.55	NA
% drop instances			
Mean (SD)	7.49 (2.67)	1.44 (1.02)	NA
Range	5.56–11.42	0.26–2.82	NA
% high speed			
Mean (SD)	15.96 (9.67)	0.17 (0.19)	NA
Range	4.63–27.4	0–0.51	NA
Missing days			
Mean (SD)	0.00 (0.00)	0.44 (0.88)	0
Range	0	0–2	0
HR data			
Participants w/any data	n=2	n=9	n=9
Days recorded			
Mean (SD)	1.44 (3.13)	13.56 (2.24)	12.44 (4.39)
Range	0–9	8–15	1–15
Obs. minutes per day			
Mean (SD)	97.71 (62.17)	1015.19 (250.33)	NA
Range	53.75–141.67	713.20 (1363.23)	NA
% drop instances			
Mean (SD)	30.74 (5.32)	0.01 (0.01)	NA
Range	26.98–34.51	0–0.03	NA
% HR Jump			
Mean (SD)	0.47 (0.00)	1.05 (1.21)	NA
Range	0.47–0.47	0.22–4.48	NA
Missing days			
Mean (SD)	1.33 (2.69)	0.67 (1.00)	0
Range	0.00–7.00	0.00–3.00	0

.HR, heart rate; NA, not available; Obs., observed; w/, with.

of data because Polar did not allow for a comprehensive download of data.

DISCUSSION

We aimed at comparing commercial, wrist-worn activity monitors in terms of usability for older adults. We explored usability from the participants' perspective as well as from the researchers' perspective, the latter including data management. We found modest participant usability across brands and, from the researchers' perspective, benefits of using the Fitbit monitor compared with the Polar or the Apple monitor.

Usability from the Participants' Perspective

In quantitative reports, we found that usability as measured by participants' reports on the SUS, did not differ significantly between monitors. This is in line with a previous study with habitual-users, which did not find significant differences between various monitors.[8] Indeed, SUS scores in this study did not differ from those found in this previous study. Our SUS scores were also not different from the established 'average' usability, which represents a C grade.[22] Taken together, these findings suggest that the brand of monitor chosen may not make any notable difference to usability and that usability could be improved across brands.

Indications on aspects that need further development can be found in the qualitative usability data. The major shortcoming of the Apple monitor was the limited battery life, while the major criticism of the Fitbit monitor was the poor readability of the display due to limited brightness and small font size. Addressing this combination of complaints may constitute a difficult task, as maximising display brightness and size while maintaining a good battery life is a well-recognised manufacturing challenge.[23] Criticism of the Polar monitor regarding poorly made straps and charging interface may be easier to address. Indeed, all manufacturers may want to improve the monitors' straps as skin irritation occurred across monitors. Concerns regarding battery life and comfort of wear appear to generalise across research uses of commercial monitors,[13] while the issues described with the Polar charger and the Fitbit readability may be specific to the older population in this study and/or specific to the monitors.

Usability concerns described in the qualitative data are also important for researchers to consider. If prolonged wear is required or skin sensitivities are known, it may be advisable to provide participants with replacement wrist bands made from materials that are less irritating to the skin than the standard straps. This can also avoid aesthetical concerns, as has been previously reported.[24] Whether to prioritise display readability or battery life may depend on the research design: if participants are required to keep track of their activity statistics or if they only use monitors for a short period of time, display readability could be maximised by using the Apple monitor. If wear is prolonged and tracking activity is not a concern, the Fitbit monitor may be preferred, especially considering its data management advantages.

Based on the present results, we cannot make any recommendations as to which individual factors to take into account when planning activity monitor research. While we did find an association between gender and usability, we have limited confidence in this finding as only one third of our sample (9/27) was male. Neither level of education, nor age, nor computer experience, nor ATI predicted usability ratings on the SUS, in spite of diversity within the sample on at least the latter three aspects. Previously observed age effects in mobile technology adoption[25] may therefore not stem from low

perceived usability, a general lack of experience with, or aversion to technology. Rather, age effects might be based on low personal interest in activity monitoring. Qualitative reports from a previous investigation[11] and this study support this idea. In our study, multiple participants indicated that activity monitoring has little value to them personally and only about one-third (8/27) reported that they had monitored either their steps or HR. Only 6/27 (22%) rated the SUS item 'I think that I would like to use this [activity monitor] frequently' with a score indicating clear agreement. This indicates even lower potential for adoption than previously reported (56% stated that they wanted to continue use).[11]

Going forward, researchers may consider providing participants with motivating factors, such as extensive device training, personal activity goal setting and peer support.[26] Gamification, which has shown promising benefits when incorporated into health technology for older adults,[27] may be another promising avenue.

Usability from the researchers' perspective and data management

There were substantial differences between activity monitors in terms of data management and researcher usability. Polar did not allow a comprehensive data download. This precludes the Polar activity monitors from future use in research studies that require access to multiple days of real-time activity data. Data download from Apple was characterised by many missing days, reflecting a very limited internal storage capacity of the monitor. Further, HR data were excluded if the data were downloaded after a specific date. This poses major problems for research studies. Moreover, syncing multiple Apple monitors with one mobile phone caused data loss, meaning that multiple phones would have to be used, which is not financially feasible in many research settings. Therefore, from the researcher perspective, using the Fitbit seems most convenient. An advantage of Fitbit data was the report of confidence levels for HR data, indicating whether the HR estimate may have been influenced by movement or an impeded optical signal.[28] For a researcher, these confidence levels are important quality indicators.

In general, as others have remarked previously,[12 13] researchers' experience would be greatly improved by more comprehensive data access and greater transparency with regards to the algorithms used to estimate activity markers. Currently, lacking access to raw data and poor documentation of algorithms—including updates to algorithms—force researchers to work with data of unknown quality. Confidence indicators may be an acceptable compromise that, on one hand, protects proprietary algorithms while given researchers some indication as to which data are useable for analysis and, on the other hand, could easily be implemented by manufacturers. In designs such as the present one, the option to extend internal storage of monitors would also help to prevent data loss in prolonged periods without syncing.

LIMITATIONS

While our study has the advantage of being set in normal daily life, which increases ecological validity, there still are limits to the generalisability of the results. Findings are based on a use scenario of 2 weeks without continuous synchronisation with a mobile device. Other use cases might yield diverging findings. For instance, synchronisation issues experienced when syncing multiple Apple monitors with one mobile phone would almost certainly be avoided if each device was synched with a separate phone. Moreover, issues with monitor readability may have been avoided if participants had ongoing access to the synced mobile device, which showed them an overview of the data. Indeed, a recent review of qualitative studies investigating wearable technology use found that limited feedback or visualisation of device measurements can negatively affect attitudes towards use.[26] Usability in this study may thus be biased towards lower ratings. However, as ratings did not differ from those in previous research with habitual monitor users, we believe that such a bias, if present, was minimal.

Conclusion

Usability was comparable across monitors and in line with usability reports from previous studies. Modest usability ratings and qualitative reports suggest that monitors still require further development for the older population. Brand-specific complaints, for instance, regarding display readability (Fitbit), battery life (Apple) and ease of charging the battery (Polar) should be considered when planning monitor use in research. We observed critical differences in data management, an important aspect for researchers that must also be taken into account when planning new studies. Further, participants may require external motivation for prolonged use of monitors, for example, through gamified feedback and peer support, another aspect that researchers should keep in mind.

Twitter Lena M Hofbauer @LenaHofbauer

Acknowledgements We want to thank our colleague, Sabrina D. Ross, for aiding data collection. We also thank participants for their time.

Contributors LMH and FSR conceived and designed this study. LMH acquired data and carried out the statistical analysis, which was checked by FSR. Both authors interpreted the data. LMH produced the initial draft of the manuscript which was further revised by both LMH and FSR. Both authors approved of final submission. The corresponding author attested all authors meet authorship criteria and that no others meeting the criteria have been omitted. LMH is responsible for the overall content as guarantor.

Funding The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

Competing interests None declared.

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

Patient consent for publication Not applicable.

Ethics approval This research was ethically approved by University Medicine Greifswald's ethics committee (BB 231/20). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available on reasonable request. Depersonalised data that underlie the results reported in this article are available on request to the corresponding author.

ORCID iDs
Lena M Hofbauer http://orcid.org/0000-0003-1789-711X
Francisca S Rodriguez http://orcid.org/0000-0003-2919-5510

REFERENCES

1 Pasanen T, Tolvanen S, Heinonen A, *et al*. Exercise therapy for functional capacity in chronic diseases: an overview of meta-analyses of randomised controlled trials. *Br J Sports Med* 2017;51:1459–65.
2 Sattler MC, Jaunig J, Tösch C, *et al*. Current evidence of measurement properties of physical activity questionnaires for older adults: an updated systematic review. *Sports Med* 2020;50:1271–315.
3 Henriksen A, Johansson J, Hartvigsen G, *et al*. Measuring physical activity using Triaxial wrist worn polar activity Trackers: a systematic review. *Int J Exerc Sci* 2020;13:438.
4 Germini F, Noronha N, Borg Debono V, *et al*. *Accuracy and acceptability of Wrist-Wearable Activity-Tracking devices: systematic review of the* literature. *J Med Internet Res* 2022;24:e30791.
5 Straiton N, Alharbi M, Bauman A, *et al*. The validity and reliability of consumer-grade activity trackers in older, community-dwelling adults: a systematic review. *Maturitas* 2018;112:85–93.
6 Serhani MA, T El Kassabi H, Ismail H. Ecg monitoring systems: review, architecture, processes, and key challenges. *Sensors* 20201796;20. doi:10.3390/s20061796
7 Dobbs WC, Fedewa MV, MacDonald HV, *et al*. The accuracy of acquiring heart rate variability from portable devices: a systematic review and meta-analysis. *Sports Med* 2019;49:417–35.
8 Liang J, Xian D, Liu X, *et al*. Usability study of mainstream wearable fitness devices: feature analysis and system usability scale evaluation. *JMIR Mhealth Uhealth* 2018;6:e11066.
9 Alharbi M, Straiton N, Smith S, *et al*. Data management and wearables in older adults: a systematic review. *Maturitas* 2019;124:100–10.
10 Boeselt T, Spielmanns M, Nell C, *et al*. Validity and usability of physical activity monitoring in patients with chronic obstructive pulmonary disease (COPD). *PLoS One* 2016;11:e0157229.
11 Fausset CB, Mitzner TL, Price CE. *Older Adults' Use of and Attitudes toward Activity Monitoring Technologies. in Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. Los Angeles, CA: SAGE Publications Sage CA, 2013.
12 Woolley SI, Collins T, Mitchell J, *et al*. Investigation of wearable health tracker version updates. *BMJ Health Care Inform* 2019;26:e100083.
13 Wright SP, Hall Brown TS, Collier SR, *et al*. How consumer physical activity monitors could transform human physiology research. *Am J Physiol Regul Integr Comp Physiol* 2017;312:R358–67.
14 Nelson BW, Allen NB. Accuracy of consumer wearable heart rate measurement during an ecologically valid 24-hour period: Intraindividual validation study. *JMIR Mhealth Uhealth* 2019;7:e10828.
15 Gorny AW, Liew SJ, Tan CS, *et al*. *Fitbit charge HR wireless heart rate monitor: validation study conducted under free-living* conditions. *JMIR Mhealth Uhealth* 2017;5:e8233:e157.
16 Cheng H, Feng D, Shi X. Data quality analysis and cleaning strategy for wireless sensor networks EURASIP. *Journal on Wireless Communications and Networking* 2018;2018:1–11.
17 Tudor-Locke C, Johnson WD, Katzmarzyk PT. Accelerometer-determined steps per day in US adults. *Med Sci Sports Exerc* 2009;41:1384–91.
18 Franke T, Attig C, Wessel D. A personal resource for technology interaction: development and validation of the affinity for technology interaction (Ati) scale. *Int J Hum Comput Interact* 2019;35:456–67.
19 Boot WR, Charness N, Czaja SJ, *et al*. Computer proficiency questionnaire: assessing low and high computer proficient seniors. *Gerontologist* 2015;55:404–11.
20 Brooke J. *Sus: A "Quick and Dirty'usability*. Usability evaluation in industry 1996;189.
21 RStudio Team. *Integrated development for R*. Boston, MA: PBC, 2020.
22 Sauro J, Lewis JR. *Quantifying the User Experience: Practical Statistics for User Research*. Morgan Kaufmann, 2016.
23 Schuchhardt M, Jha S, Ayoub R. Association for Computing Machinery. In: *Caped: Context-Aware Personalized Display Brightness for Mobile Devices, in Proceedings of the 2014 International Conference on Compilers, Architecture and Synthesis for Embedded Systems*. New Delhi, India, 2014: 19.
24 Balbim GM, Marques IG, Marquez DX, *et al*. Using Fitbit as an mHealth intervention tool to promote physical activity: potential challenges and solutions. *JMIR Mhealth Uhealth* 2021;9:e25289.
25 Fox G, Connolly R. Mobile health technology adoption across generations: narrowing the digital divide. *Information Systems Journal* 2018;28:995–1019.
26 Moore K, O'Shea E, Kenny L, *et al*. *Older adults' experiences with using wearable devices: qualitative systematic review and Meta-*synthesis. *JMIR Mhealth Uhealth* 2021;9:e23832.
27 Koivisto J, Malik A. Gamification for older adults: a systematic literature review. *Gerontologist* 2021;61:e360–72.
28 Russell A, Heneghan C, Venkatraman S. Investigation of an estimate of daily resting heart rate using a consumer wearable device. *medRxiv* 2019:19008771.