Digital Biomarkers

Research Report

Digit Biomark 2025;9:40–51
DOI: 10.1159/000543898

Received: February 1, 2024
Accepted: January 26, 2025
Published online: February 3, 2025

Analysis Method of Real-World Digital Biomarkers for Clinical Impact in Cancer Patients

Ingrid Oakley-Girvan[a, b] Yaya Zhai[c] Reem Yunis[a] Raymond Liu[d, e] Sharon W. Davis[a] Ai Kubo[d] Sara Aghaee[d] Jennifer M. Blankenship[c] Kate Lyden[c] Elad Neeman[f]

[a]Medable Inc., Strategy and Science Departments, Palo Alto, CA, USA; [b]Public Health Institute, The Data and Technology Proving Ground, Oakland, CA, USA; [c]Vivosense, Inc., Newport Coast, CA, USA; [d]Kaiser Permanente Northern California, Division of Research, Oakland, CA, USA; [e]Kaiser Permanente Northern California, Department of Hematology Oncology, San Francisco, CA, USA; [f]Kaiser Permanente Northern California, San Rafael Medical Center, San Rafael, CA, USA

Keywords
Bring your own device · Cancer · Wearables · Physical activity · Adverse clinical events · Digital health · Daily step count · Sufficient step count

Abstract
Introduction: Wearable technologies can enhance measurements completed from home by participants in decentralized clinical trials. These measurements have shown promise in monitoring patient wellness outside the clinical setting. However, there are challenges in handling data and its interpretation when using consumer wearables, requiring input from statisticians and data scientists. This article describes three methods to estimate daily steps to address gaps in data from the Apple Watch in cancer patients and uses one of these methods in an analysis of the association between daily step count estimates and clinical events for these patients. ***Methods:*** A cohort of 50 cancer patients used the DigiBioMarC app integrated with an Apple Watch for 28 days. We identified different gap types in watch data based on their length and context to estimate daily steps. Cox proportional hazards regression models were used to determine the association between step count and time to death or time to first clinical event. Decision tree modeling and participant clustering were also employed to identify digital biomarkers of physical activity that were predictive of clinical event occurrence and hazard ratio to clinical events, respectively. ***Results:*** Among the three methods explored to address missing steps, the method that identified different step data gap types according to their duration and context yielded the most reasonable estimate of daily steps. Ten hours of waking time was used to differentiate between sufficient and insufficient measurement days. Daily step count on sufficient days was the most promising predictor of time to first clinical event ($p = 0.068$). This finding was consistent with participant clustering and decision tree analyses,

Ingrid Oakley-Girvan and Yaya Zhai are co-first authors and contributed equally to this study.

karger@karger.com
www.karger.com/dib

Correspondence to:
Ingrid Oakley-Girvan, oakley@stanford.edu

where the participant clusters emerged naturally based on different levels of daily steps, and the group with the highest steps on sufficient days had the lowest hazard probability of mortality and clinical events. Additionally, daily steps on sufficient days can also be used as a predictor of whether a participant will have clinical events with an accuracy of 83.3%. ***Conclusion:*** We have developed an effective way to estimate daily steps of consumer wearable data containing unknown data gaps. Daily step counts on days with sufficient sampling are a strong predictor of the timing and occurrence of clinical events, with individuals exhibiting higher daily step counts having reduced hazard of death or clinical events.

Introduction

Clinical trials that allow patients to complete at least some of the required measurements from home have become more common [1, 2]. Decentralized clinical trials may reduce patient and sponsor burden by reducing the number of required visits to the clinical site [3, 4] and are financially viable [5]. Decentralized clinical trials that leverage wearable and connected technologies support the FDA's Patient-Focused Drug Development guidance [6, 7] by enhancing patient engagement and enabling the assessment of real-world measures that matter to patients and reflect their life experience [8, 9]. Enabling patients to use their own mobile device in trials ("Bring Your Own Device" [BYOD]) has demonstrated equivalence in measurements obtained compared with a provisioned device, along with high patient acceptance [10–13].

Patient-reported outcomes collected remotely can predict adverse events, including hospitalization, and addressing them may extend survival [14, 15]. Wearable sensors also have the potential to add health information with clinical implications to home-based monitoring [16]. In oncology patients, step counts measured at home were associated with chemotherapy symptoms [17, 18], hospitalization, and treatment delays during radiation therapy [19, 20], health-related quality-of-life indicators [21], postoperative physical recovery [22], and survivorship [23]. Passively measured steps from consumer wearables have been established as predictors of clinical outcomes as well as quality of life in cancer patients [24].

Utilizing consumer wearable devices for continuous sensor data collection offers a distinct advantage over episodic assessments in clinical settings as it provides comprehensive and real-time insight into patients' health behaviors and physiological responses in a low-burden and sensitive way [25, 26]. It is feasible to collect data from wrist-worn smartwatches inside the home environment without clinical oversight [27], even among senior citizens [28].

However, there is a lack of consensus on how to process, analyze, and interpret data from consumer wearable sensors for clinical purposes [24, 29]. Analyzing and interpreting sensor data requires input from statisticians and data scientists to ensure clinical validity [30, 31]. This article aimed to (1) describe our method to estimate daily steps from the Apple Watch used at home by patients undergoing treatment for cancer and (2) evaluate the impact of our method by analyzing the associations between estimates of daily step counts and clinical events for these individuals.

Methods

Study Design

This single-arm, longitudinal, mixed-methods, prospective usability study enrolled a cohort of cancer patients receiving treatment at Kaiser Permanente Northern California (KPNC), an integrated healthcare delivery system. The study population and recruitment process have been detailed elsewhere [32]. In brief, participants were eligible for participation if they were ≥18 years of age, actively receiving infusion therapy, had an informal caregiver (≥18 years), had an iPhone 6 and above, and had sufficient cognitive capacity to provide informed consent. Participants used an iPhone app known as Digital BioMarkers for Clinical Impact or DigiBioMarC, which was integrated with the Apple Watch.

The study participants (demographic data available in the online supplementary material; for all online suppl. material, see https://doi.org/10.1159/000543898) were instructed to utilize the mobile app and watch for at least 28 days, during which the mobile app delivered surveys and activity requests, and the Apple Watch captured specific digital activity data concurrently. Among the 50 participants, 12 already owned their own Apple Watch (version 3 or above), while the remaining participants were provisioned with a loaner Apple Watch (version 3 or 4) that was mailed to them at no cost. Participants who completed the study received a USD 100 Amazon gift certificate for their time. The study was reviewed and approved by the KPNC Institutional Review Board. All data collected using the app was stored in a secure HIPAA-compliant cloud that was managed by KPNC project staff.

Estimation of Daily Steps

Participants' step counts were passively collected during the study on the Apple Watch using the Apple HealthKit software native to the device. In an at-home study with a consumer device such as the Apple Watch, gaps in the step counts can arise due to participants forgetting to wear or charge their watch, or extended periods of inactivity. Because of this, we employed three distinct methods to estimate the number of daily steps. These methods made different assumptions about the gaps in sampling. Online supplementary Figure S1A illustrates an example of gaps in sampling, with time stamps modified to ensure privacy. A summary of each method is provided below.

Method 1: Assumed all gaps represented inactivity and calculated the sum of all recorded steps to estimate the daily step count. This method is also considered the method without preprocessing.

Method 2: Assumed all gaps were because the participant was not wearing the watch (non-wear) and, thus, employed a normalization approach by averaging the daily steps to impute the number of steps during gaps with the assumption that participants were awake for 16 h during each 24-h period.

Method 3: Identified different gap types according to their length and context. Utilized gap type-based normalization of steps by extracting the measurement period of all the channels from Apple HealthKit except for Basal Energy Burned (BEB) and Apple Exercise Time (AET). Gaps were identified when no channels registered measurement. The longest gap for each noon-to-noon period was recognized as sleep (SLEEP), allowing for short measurements of 2 min for any movement or awakening during sleep. All gaps with lengths shorter than 3.5 h were considered short gaps with inactivity (SED for short of sedentary). Any gaps longer than 3.5 h were categorized as non-wear (NONWEAR). We chose the threshold of 3.5 h because this duration is well above the typical maximum sedentary duration for healthy populations, which range from 100 to 120 min [33, 34]. Daily waking wear time (DWWT) was defined as the sum of time with measurement (RECORDED) and short gaps of inactivity (SED). Days with at least 10 h of waking wear time were considered days with sufficient measurement [35]. Daily waking time (DWT) was defined as the sum of daily waking wear time (DWWT) and long non-wear gaps (NONWEAR).

The recorded daily steps were collected over daily waking wear time (RECORDED + SED); therefore, the estimation of daily steps for Method 3 was derived through the following calculation:

$$\frac{Recorded\ daily\ steps}{Daily\ waking\ wear\ time\ (min)} \times Daily\ waking\ time\ (min)$$

where,

$$Daily\ waking\ wear\ time\ (DWWT) = RECORDED + SED$$

$$Daily\ waking\ time\ (DWT) = DWWT + NONWEAR$$

Digital Biomarkers of Physical Activity

We determined eight digital biomarkers of physical activity, considering both daily steps and the availability of activity data itself:

(1) Number of days with sufficient measurement: this refers to the count of days with waking wear time equal to or exceeding 10 h, hereafter referred to as "sufficient days."
(2) Number of days with insufficient measurement: this indicates the count of days with waking wear time less than 10 h, hereafter referred to as "insufficient days."
(3) Median daily steps on sufficient measurement days: for days with sufficient measurement, the median daily steps were adjusted for estimated non-wear time using Method 3.
(4) Median daily steps on insufficient measurement days: similar to the previous digital biomarker (#3), the median daily steps were adjusted for estimated non-wear time (Method 3) for days with insufficient measurement.
(5) Daily unmeasured time on sufficient days: this represents the unmeasured time on days with sufficient measurement, calculated as the difference between 24 h and the daily waking wear time, estimated as the sum of time with measurement and short gaps.
(6) Daily unmeasured time on insufficient days: similar to digital biomarker 5, this denotes the unmeasured time on days with insufficient measurement.
(7) Count of days with no measurement: this indicates the number of days with no recorded measurements.
(8) Insufficient day to sufficient day ratio (ISR): this is computed as the ratio of insufficient days to sufficient days.

Definition of Clinical Events

The timing of death and other adverse clinical events was collected from the participants' electronic health records (EHRs) for a period of up to 6 months following the end of the study period. The clinical events of interest included palliative care referrals, hospice referrals, and death within 6 months and emergency department visits and hospital visits if within 3 months.

Analysis of Clinical Events and Digital Biomarkers of Physical Activity

To determine whether digital biomarkers of physical activity predicted time to death and time to first clinical event, Cox proportional hazards regression models were implemented using the coxph and Surv function from the R package survival [36, 37]. Kaplan-Meier survival curves were generated using the ggsurvplot function in the R package survminer [36–38]. A power analysis was performed on the Cox proportional hazards regression model with nonbinary covariates.

To assess the utility of digital biomarkers of physical activity and identify the digital biomarker that was most predictive of the occurrence of clinical events, a decision tree model was used. Specifically, DecisionTreeClassifier in the module sklearn.tree of the scikit-learn library [39] was deployed to select the digital biomarker that distinguished the participants with at least one clinical event (Class = 1) from those having no reported clinical events (Class = 0). Gini impurity [40], the probability of incorrect classification when all labels are randomly chosen, was automatically generated with this same method as a part of the output.

We also examined whether participant groups naturally emerging from patterns in digital biomarkers of physical activity were associated with the time to death or time to clinical events. Participants were grouped into clusters using the hclust function in the R base [41] based on their median daily step count on sufficient and insufficient days. Clusters were then analyzed for their hazard ratio to death or clinical events using Surv and Survfit functions from the R package survival [36], and Kaplan-Meier survival curves were generated using the ggsurvplot function in the R package survminer [38].

Step Estimation Method Comparisons: Impact of Methods on Clinical Analyses

The impact of the data preprocessing steps deployed using Method 3 was assessed by conducting parallel analyses with step data derived directly from Apple Watch (i.e., Method 1). p values were compared to evaluate whether the data preprocessing procedures contributed additional insights into clinical event prediction compared to the use of raw recorded steps.

Compliance of previously owned versus provisioned Apple Watches. To assess whether compliance varied by whether participants already owned an Apple Watch (n = 12) or were provisioned an Apple Watch (n = 37), a t-test was used to compare the daily unmeasured time between these two participant groups.

Results

Estimation of Daily Steps: Comparison of Proposed Methods

We used three methods to treat gaps in sampling when estimating daily steps using consumer-grade wearable device data. Online supplementary Figure S1B illustrates how the three methods handle gaps, using a representative example. Simply summing up daily recorded steps (Method 1) resulted in an average of 3,981 ± 11,041 steps per day, while a simple normalization of recorded stepping time to a predefined 16 h of waking time (Method 2) estimated an average of 30,487 ± 18,125 daily steps. On average, ± SD, 5,011 ± 3,354 daily steps were estimated using Method 3. The three methods to estimate the daily step count of an example participant are shown in online supplementary Figure S2.

The distinction between Method 1 and Method 3 lies in their treatment of prolonged gaps in collected data during the daytime. Method 1 assumes all gaps imply no steps, whereas Method 3 reintroduces steps for long gaps during waking hours, which may not solely indicate inactivity. On days without such prolonged gaps, both methods yield identical step estimates. Despite these identical estimates, a one-sided paired t-test comparing Method 1 and Method 3 yielded a p value of 3.618e-06, indicating a significant difference between the two methods. The identification of gap types in Method 3 allows for estimating the amount of data collection time for each day, thereby influencing the quality of the data.

The commonly used threshold of 10 h of waking wear time was used to differentiate between sufficient and insufficient measurement days [35]. Figures 1 and 2 illustrate the joint distribution of estimated daily steps and unmeasured time for each participant on sufficient and insufficient days. Figure 1 shows the box plot of daily steps on sufficient and insufficient days, indicating more steps on sufficient days for most patients. Figure 2 displays the box plot of daily unmeasured time on sufficient and insufficient days, with a clear separation at the chosen cutoff of 14 h. Given the observed differences in different unmeasured time between sufficient and insufficient days, the definition of 10 h per day is appropriate. Of the total measured days, 63% had sufficient measurement. The percentage of days with sufficient measurement varied by participants (range: 9%–100%).

Clinical Events Analysis

Using watch data processed with Method 3, we investigated whether the eight digital biomarkers of physical activity predicted time to death, first clinical event (Cox

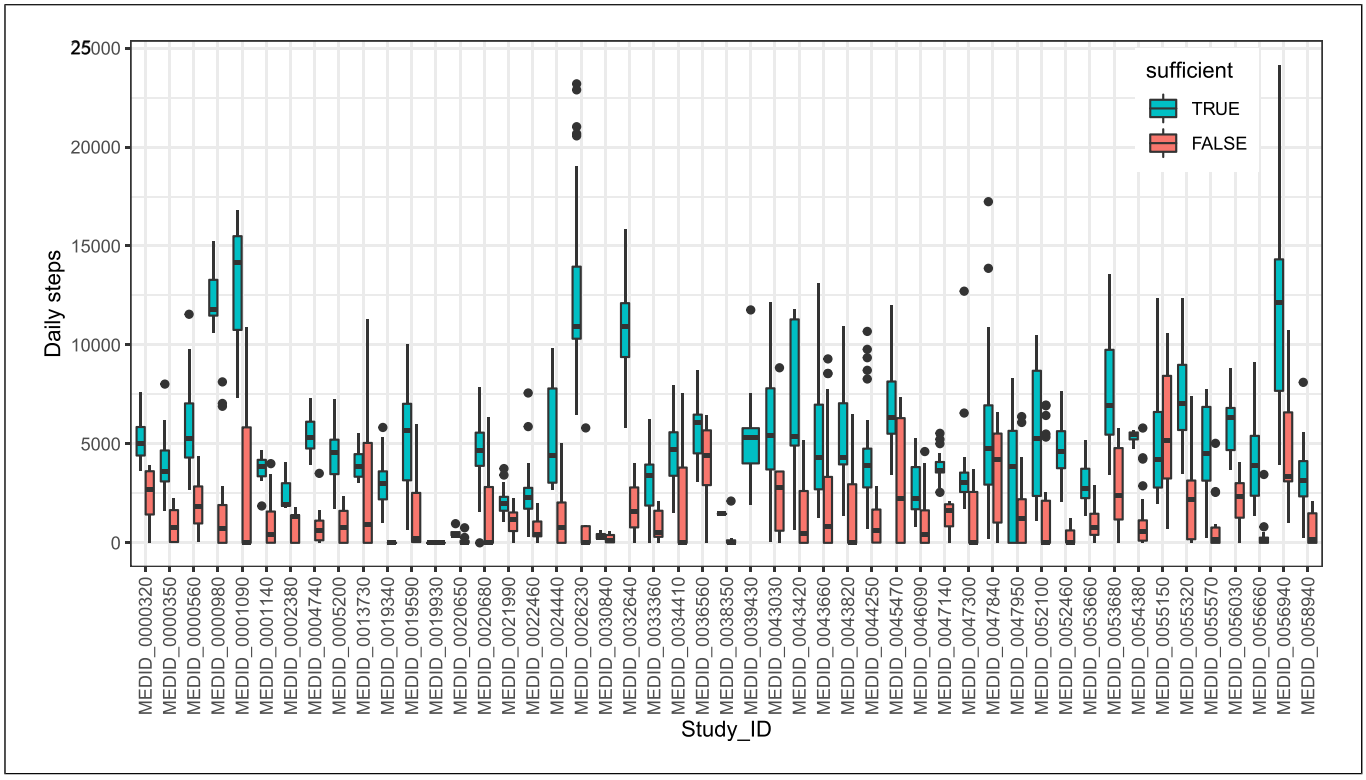

Fig. 1. Box plot of daily steps on days with sufficient measurements ("sufficient days") and days with insufficient measurements ("insufficient days"). There are more steps on sufficient days for most patients.

regression models), or incidence of any clinical event (decision tree analysis). Of the 50 participants, 49 provided both complete EHR and Apple Watch data; thus, 49 participants were included in the clinical events' analysis.

Predicting Time to Death or Time to First Clinical Event

While none of the digital biomarkers of physical activity predicted time to death (data shown in online suppl. Table S1), the daily step count on sufficient days tended to predict the time to first clinical event with a hazard ratio of 0.79, suggesting the hazard decreases by a factor of 0.79 for every 1,000 increase in daily steps on sufficient days (p = 0.068, Table 1). The power analysis shows at least 14 participants with events (57 total participants) will be needed to achieve 80% power at a significance level of 0.05 with a hazard ratio of 0.79, based on the current event occurrence rate and variance of the variable.

Predicting Incidence of Any Clinical Events

Decision tree analyses were performed to identify the digital biomarker that differentiated participants with clinical events. One participant was excluded from this analysis because daily unmeasured time on insufficient days for this participant was missing; therefore, 48 participants were included in the decision tree analysis. Daily median step count on sufficient days predicted clinical event occurrence with an accuracy of 83.3%. Participants with greater than 2,510 median daily steps on sufficient days were predicted to have no clinical events, and those with less than 2,510 daily steps on sufficient days were predicted to have at least one clinical event.

Of the 48 participants included in the analysis, 12 had at least one clinical event. A naïve classifier would predict that no one would have a clinical event, but with the rule of median daily step count <2,510 on sufficient days, 8 participants were predicted to have at least one clinical event. Of the 8 participants classified by median daily steps, 6 were correctly classified and had at least one clinical event. A graphical illustration of the decision tree classification and Gini impurity results are depicted in online supplementary Figure S3.

Participant Clusters and the Association between Participant Clusters and Clinical Events

A hierarchical clustering was also performed to examine patterns of daily steps among participants. Both steps on sufficient days and insufficient days were

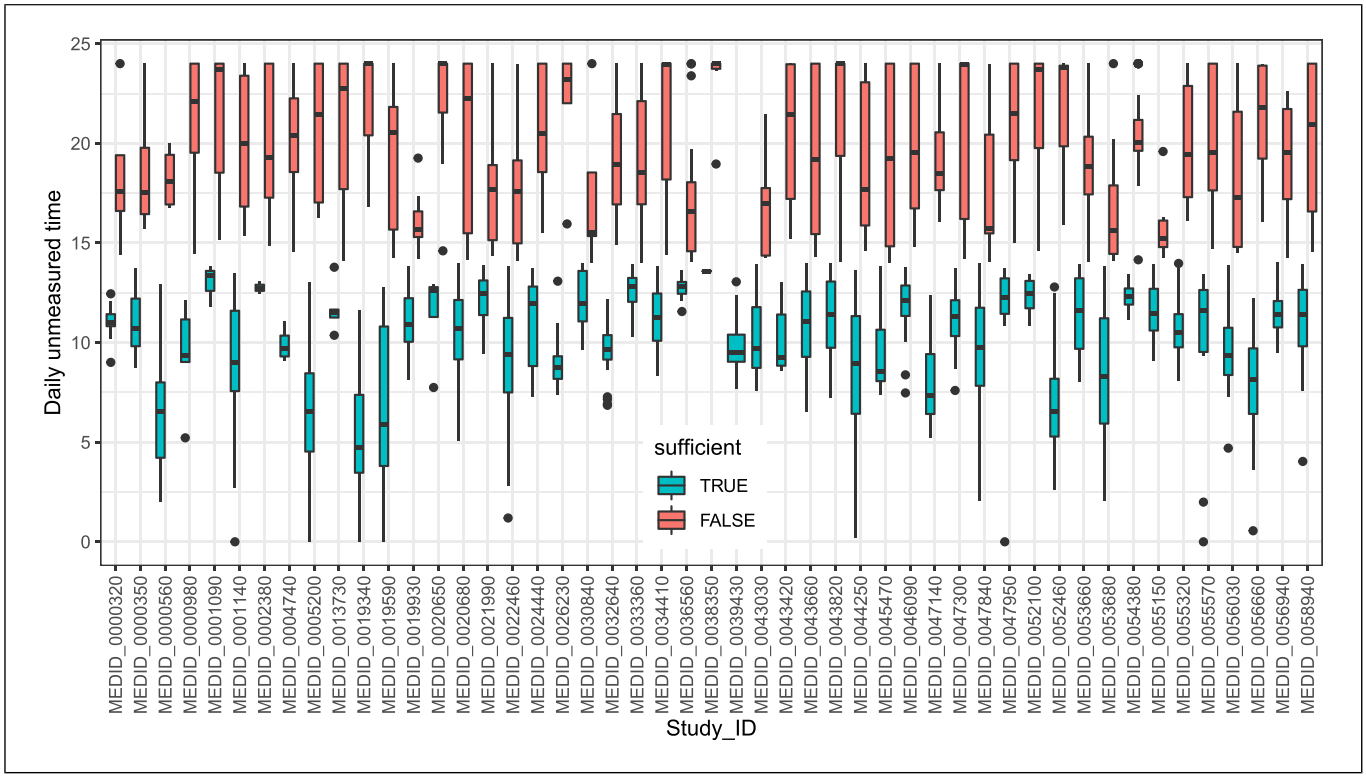

Fig. 2. Box plot of daily unmeasured time on sufficient days and insufficient days. There is a clear separation between unmeasured time at the chosen cutoff of 14 h. The populations have quite different unmeasured time.

Table 1. Digital biomarkers and *p* values to predict time to first clinical event using univariate Cox regression model

Variable	*p* value	Coef	95% CI	Hazard ratio
Number of sufficient days	0.678	9.49E−03	(−3.53E−02, 5.43E−02)	1.01
Number of insufficient days	0.396	−0.027	(−0.088, 0.035)	0.97
Steps on sufficient days	0.068	−0.235	(−0.486, 1.72E−02)	0.79
Steps on insufficient days	0.803	−7.57E−02	(−0.671, 0.520)	0.93
Unmeasured time on sufficient days	0.701	−0.056	(−0.345, 0.232)	0.95
Unmeasured time on insufficient days	0.328	−0.119	(−0.359, 0.120)	0.89
Number of days with no measurement	0.306	0.033	(−0.030, 0.096)	1.03
Ratio of insufficient days to sufficient days	0.728	−0.041	(−0.272, 0.190)	0.96

A total of 49 participants were included in this analysis. Of those included, there were 13 participants that experienced at least one clinical event (see online supplementary material for details on clinical events by participants). The two variables concerning daily steps have been adjusted to reflect changes per thousand steps per day.

considered in the development of the clusters. Five clusters, depicted in different colors in Figure 3a and online supplementary Figure S4, were identified.

The daily steps on sufficient days and insufficient days for each participant cluster are shown in Figure 3b. Survival analysis was performed to understand the death hazard and time to first clinical event with different clusters, and results are shown in Figures 3c and 3d, respectively. Cluster 4 (blue) had the highest hazard in both analyses and had the lowest

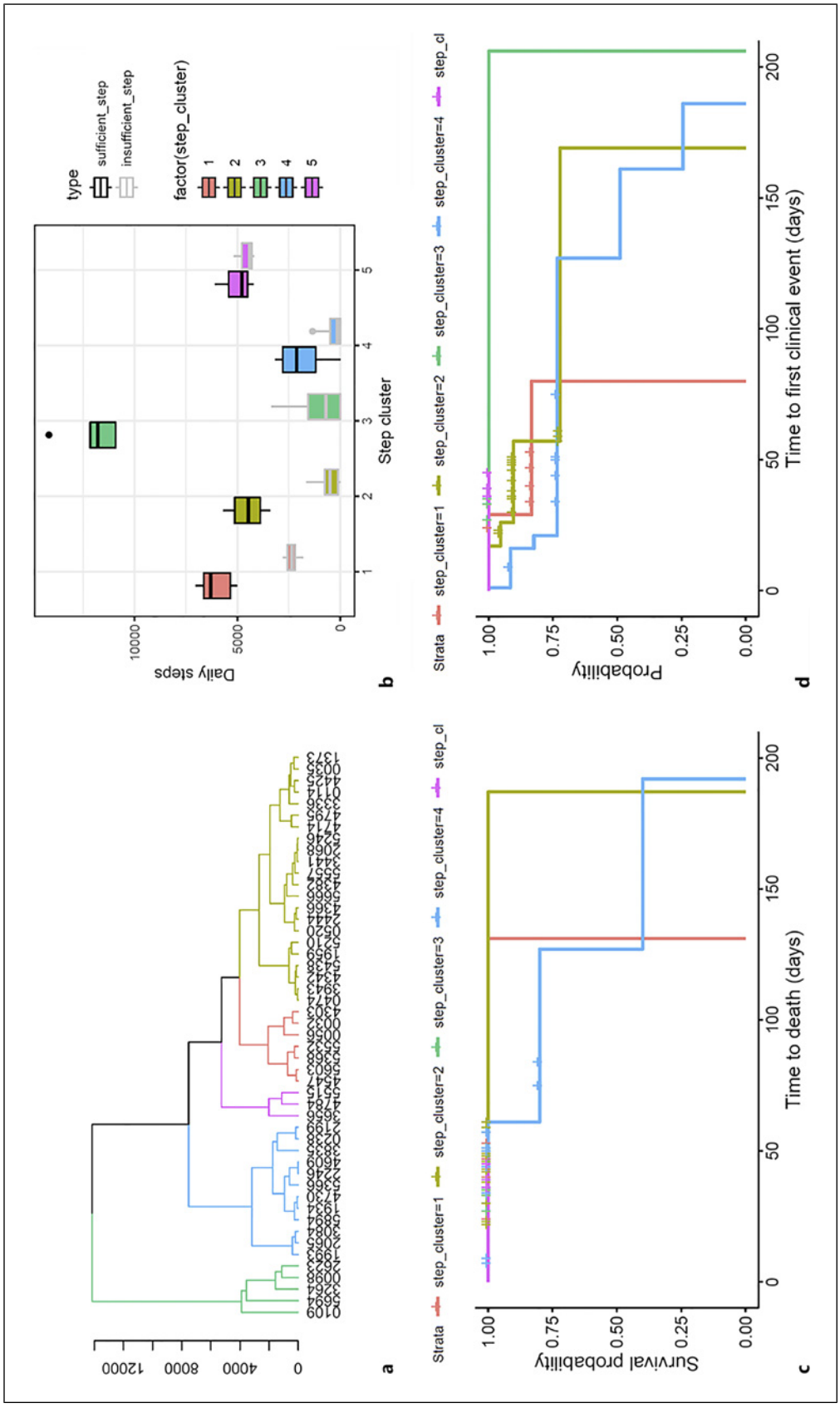

Fig. 3. Participant clustering by daily steps and the prediction of time to clinical events. **a** Dendrogram illustrating the 5 primary clusters of participants based on steps. **b** Steps on sufficient days and insufficient days for the five clusters. **c** Kaplan-Meier plot for the time to death analysis with different participant clusters. **d** Kaplan-Meier plot for the time to first clinical event analysis with different participant clusters.

daily steps on sufficient days. In contrast, Cluster 3 (green) had the lowest hazard in both analyses and the highest steps on sufficient days.

Comparison to Analyses without Preprocessing

To demonstrate the importance of the data preprocessing that was performed using Method 3, we repeated the aforementioned analyses using raw recorded steps from the Apple Watch (i.e., Method 1). This comparison evaluated whether the data preprocessing procedures contribute additional insights into clinical event prediction compared to the use of raw recorded steps.

Raw recorded steps (Method 1) did not predict time to death ($p = 0.513$) nor time to first clinical event ($p = 0.259$). The p value quantifies how likely it is to observe data at least as extreme as what was actually observed, assuming the null hypothesis is true. Both of these p values substantially exceed the ones resulting from our analysis using Method 3 indicating less substantial or non-significant results when unprocessed daily steps are utilized (data not shown).

In terms of predicting clinical event occurrence, a decision tree model with a depth of 1 yielded an accuracy of 75.5%. This accuracy was similar to a rudimentary model that predicts no occurrence of clinical events (yielding an accuracy of 36/49 = 73.5%), indicating that raw recorded steps are not predictive of clinical events.

Comparing Compliance between Participant-Provided and Study-Provided Devices

A t-test was conducted to compare the daily unmeasured time between participants (with complete watch data) using their self-owned Apple Watch ($n = 12$) and those using provisioned watches ($n = 37$). No statistically significant distinction in compliance was observed ($p = 0.215$), suggesting similar levels of compliance between the two groups.

Discussion

Principal Findings

Method for Estimating Daily Steps from Consumer-Grade Wearable Device Data in Clinical Trials

Because real-world data acquired from consumer wearable devices often exhibits gaps in step sampling, we evaluated three methods to determine step counts. We describe these methods and present the best of the three to address these data gaps. Specifically, we devised a robust preprocessing procedure to distinguish between gaps resulting from inactivity and those arising from non-wear. This method (Method 3) synthesizes a priori knowledge of human behavior with heuristic rules, effectively differentiating and categorizing gap types. The method considers data availability from other channels at the same time, further enhancing the precision of daily step estimation. Furthermore, this method mitigates the risk of incorporating days with exceptionally low step counts due to inadequate device wear. By systematically addressing data gaps in consumer-grade wearable device data, this method emerges as an asset in ensuring the accuracy and interpretability of consumer wearable devices data.

Importantly, our method, #3, in comparison to the first two methods, yields estimations of daily step counts from a consumer-grade wearable sensor that closely aligns with findings from previous studies that have deployed research-grade sensors. Our method estimated an average of 5,011 ± 3,354 daily steps for oncology patients, a pattern in line with expected human activity levels in oncology patients. This daily step count is similar to the reported 4,935 ± 2,884 daily average for metastatic breast cancer patients prior to a step-increasing intervention [42], as well as the 4,966 steps recorded for rectal cancer patients before chemoradiotherapy [43]. Comparable investigations have documented diverse step counts, spanning from an average of 6,689 ± 3,261 inpatient steps following lung resection surgery [44], to 4,099 steps for children and adolescents prior to chemotherapy [45], and 3,883 steps for head and neck cancer patients at the early stages of chemoradiotherapy [46].

Analysis of Real-World Daily Step Count, Clinical Events, and Clinical Implications

This study highlights the potential of daily stepping behavior to predict both the timing and occurrence of clinical events based on an analysis of data obtained from cancer patients undergoing intravenous therapy. This prediction is evident across Cox regression, decision tree and clustering models. In our research, a lower count of real-world steps aligns with clinical events, including emergency department and hospital visits, palliative care and hospice referrals, and mortality. This association also finds some support in scientific rationale and existing literature that reveals correlations between decreased steps and heightened fatigue, cumulative symptom count [45], higher number of chemotherapy-related symptoms [17], and higher likelihood of hospitalization during radiation therapy [19]. It is essential to note, however, that in the current study, the positive signal for daily step counts was discernible only when utilizing preprocessed data. Our method reduced the risk of incorporating days

with exceedingly low step counts during periods when individuals are not wearing the device, thereby offering a more precise evaluation of waking stepping behaviors that accounts for data quality and satisfactory sampling. While results from this study are preliminary, they support the continued investigation into the predictive potential of daily stepping behavior through well-controlled and sufficiently powered studies.

Within the context of patients undergoing anti-cancer treatment, data missingness (as measured by insufficiently measured days) serves as a meaningful feature extracted from real-world data. This finding could be interpreted to suggest that patients who are doing more poorly are less able to adhere to data collection protocols.

Furthermore, we observed that compliance was consistent regardless of whether participants use provisioned devices or their own (BYOD). This implies that both provisioning and BYOD options are viable choices [10–13] as compliance levels in this research did not significantly differ between the two approaches. Given the relatively small number of participants that used their own devices (n = 12), additional studies are needed to confirm this finding.

Strengths and Limitations

This study includes oncology patients across diverse cancer types and stages, supported by a data collection period that captures real-world behaviors over 1 month. There is evidence that consumer wearable sensors have a low error rate, particularly for step counting [47]. A case in point is the Apple Watch, which has been shown to be extremely accurate for measuring step counts for adults [48]. While the scope of this study did not include evaluation of all types of consumer wearables that capture similar data, the utilization of a common consumer wearable like the Apple Watch adds value to our study as it offers evidence that participants can use devices with which they are familiar or quickly learn to use provisioned ones.

A key finding of our work is that the method we developed can effectively address the challenge of potential gaps in consumer-grade device data, simultaneously estimate data missingness and yield reliable, high-quality step count estimates that correlate with clinical events across multiple analyses.

It is important to acknowledge certain limitations associated with consumer-grade devices, such as expense of devices, the potential need for technology skills to link watches to phones and maintain devices, and the device-charging needs. Battery life is of particular relevance for the Apple Watch, as charging is required daily. When a device is being charged, behaviors cannot be measured, and thus, this is an important factor to consider when selecting a device to estimate physical activity and other behaviors over 24 h. These devices may also lack transparency from manufacturers regarding the processing of raw accelerometer data, inaccessible embedded algorithms, and potential system updates performed without researchers' or clinicians' awareness. While our method and usability research describes a reasonable estimate of daily steps in oncology patients and demonstrates lower variance compared to the other two methods, the absence of an independent validation hinders comprehensive performance assessment. Additionally, the accuracy and reliability of the Apple Watch and its underlying algorithms for measuring steps in oncology patients at home remains, to our knowledge, untested.

Finally, it is important to note that this study aimed to evaluate the usability of real-world physical activity data for patients with oncology. The sample size was relatively small and overall there were few deaths and clinical events, limiting our ability to detect significant predictors. The model simplified its approach by omitting demographic factors such as age, cancer type, and stage to ensure broad applicability across diverse populations, reducing the necessity for recalibration across varying groups. Physical activity and its measurements are also impacted by environmental biases, such as climatic factors, living environment (rural or urban), and the doses of chemotherapy treatment in this oncology population. Despite these limitations, this study provides proof-of-concept evidence that digital biomarkers of physical activity have clinical relevance in this patient population. Still, larger prospective studies are needed to confirm our findings.

Conclusion

We developed a comprehensive data processing and analysis method to effectively address challenges associated with data collected using consumer devices, resulting in reasonable estimates of steps and data missingness. Daily step counts on days with sufficient sampling is a predictor of the timing and occurrence of clinical events, with individuals exhibiting the highest daily step counts having a reduced hazard of death or clinical events. These findings hold potential implications for informed clinical decision-making and the integration of patient-focused drug development strategies. Further investigation is warranted to establish real-world stepping behavior as a potential biomarker for disease progression in patients undergoing anti-cancer therapy. To ensure rigorous study outcomes, the

incorporation of a research-grade wearable device, coupled with ongoing data quality assessment, is recommended throughout a longer study duration. Enhanced statistical power and refined measure development can be achieved through the inclusion of a larger dataset encompassing a higher frequency of clinical events.

Acknowledgments

We thank members of the Medable Patient Caregiver Network for their valuable input and critical insights during early development of the DigiBioMarC app that was conducted with Duke and Stanford Universities prior to the research with KPNC. This study could not have been done without the able assistance of Elaine M. Kurtovich, KPNC Research Project Manager, and Maya E. Ramsey, KNPC Research Associate. At the time of this work, R.Y., I.O.G., and S.W.D. were employed by Medable Inc., which developed the DigiBioMarC app with funding from the NCI. A.K. and S.A. were subcontracted to work with Medable to enroll KPNC participants and conduct all study participant focused efforts. Y.Z., K.L., and J.B. were subcontracted to work with Medable to process and analyze the data. R.L. and E.N. did not receive any funding as part of this effort.

Statement of Ethics

The KPNC Institutional Review Board (IRB) approved this study (KPNC IRB #1576055). We obtained written informed consent within the app from participants before any data collection. The app was housed on the patients' iPhones, which were designed by Apple to be password-protected. In addition, the app itself was password-protected and could be opened by a passcode, facial recognition, or fingerprint, depending on how the user preferred to set it up for controlled access. These biometrics were not available to KPNC or Medable but were stored on the participants' own phones as part of the Apple iPhone operating system. The Medable platform use was accessed by Kaiser information technology and technology teams; security and privacy for all data captured by the iPhone and the study as a whole were reviewed and approved by the KPNC IRB. All data collected using the app was stored in a secure Health Insurance Portability and Accountability Act (HIPAA)-compliant cloud that was accessed and controlled by KPNC project staff. The data were deidentified for analysis. Participants who completed the surveys in their app and 2 semi-structured video conferencing interviews each received a USD 100 Amazon gift certificate in appreciation for the time and effort they spent.

Conflict of Interest Statement

At the time of this work, R.Y., I.O.G., and S.W.D. were employed by Medable Inc., which developed the DigiBioMarC app with funding from the National Institutes of Health. Y.Z., K.L., and J.B. were subcontracted to work with Medable to process and analyze the study data, and A.K. and S.A. were subcontracted to work with Medable to enroll KPNC participants and conduct all study participant focused efforts. R.L. and E.N. have no competing interests as they were not paid for their roles on the project.

Funding Sources

Research reported in this publication was partially supported by the National Cancer Institute of the National Institutes of Health under contract number HHSN261201800010C. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author Contributions

I.O.G. and S.W.D. researched the literature, conceived the study, and obtained funding from NIH for the study. I.O.G. was the P.I. for the NIH fast track study award. I.O.G. and R.Y. developed the protocol. E.N., R.L., and A.K. were involved in protocol development, IRB approval, and patient recruitment. S.A. prepared the initial study data file from KPNC. Y.Z. prepared the analysis data files and conducted the analysis with K.L. and J.B. in collaboration with I.O.G. and R.L. I.O.G., Y.Z., K.L., and S.W.D. wrote the first draft of the manuscript. All authors reviewed and edited the manuscript and approved the final version of the manuscript.

Data Availability Statement

This work was conducted, in part, with NIH Small Business Innovation Research (SBIR) funding. Due to the sensitive nature of the data captured, the data release requirements of the KPNC subcontract, and the intellectual property associated with this work, all requests for data should be directed to the Chair of the Publication Committee, Dr. Oakley-Girvan (oakley@stanford.edu). Requests for data will be considered by the Publication Committee under a data use agreement and will follow strict guidelines, including HIPAA regulations, regarding data use and the purpose of the analysis.

References

1 Khozin S, Coravos A. Decentralized trials in the age of real-world evidence and inclusivity in clinical investigations. Clin Pharmacol Ther. 2019;106(1):25–7. https://doi.org/10.1002/cpt.1441

2 Dorsey ER, Kluger B, Lipset CH. The new normal in clinical trials: decentralized studies. Ann Neurol. 2020;88(5):863–6. https://doi.org/10.1002/ana.25892

3 Mowlem FD, Tenaerts P, Gwaltney C, Oakley-Girvan I. Regulatory acceptance of patient-reported outcome (PRO) data from bring-your-own-device (BYOD) solutions to support medical product labeling claims: let's share the success stories to move the industry forward. Ther Innov Regul Sci. 2022;56(4):531–5. https://doi.org/10.1007/s43441-022-00412-1

4 Fda 2022. Pageshttps://www.fda.gov/about-fda/oncology-center-excellence/advancing-oncology-decentralized-trialson 01/03/2023 2023

5 DiMasi JA, Smith Z, Oakley-Girvan I, Mackinnon A, Costello M, Tenaerts P, et al. Assessing the financial value of decentralized clinical trials. Ther Innov Regul Sci. 2022;57(2):209–19. https://doi.org/10.1007/s43441-022-00454-5
6 Rowe-Rendleman CL. Ophthalmic drug discovery and development: regulatory aspects of patient focused drug development in ophthalmology. Pharm Res. 2019;36(4):54. https://doi.org/10.1007/s11095-019-2577-8
7 Azodo I, Williams R, Sheikh A, Cresswell K. Opportunities and challenges surrounding the use of data from wearable sensor devices in health care: qualitative interview study. J Med Internet Res. 2020;22(10):e19542. https://doi.org/10.2196/19542
8 Harmon DM, Noseworthy PA, Yao X. The digitization and decentralization of clinical trials. Mayo Clin Proc. 2023;98(10):1568–78. https://doi.org/10.1016/j.mayocp.2022.10.001
9 Crossnohere NL, Fischer R, Crossley E, Vroom E, Bridges JF. The evolution of patient-focused drug development and Duchenne muscular dystrophy. Expert Rev Pharmacoecon Outcomes Res. 2020;20(1):57–68. https://doi.org/10.1080/14737167.2020.1734454
10 Byrom B, Doll H, Muehlhausen W, Flood E, Cassedy C, McDowell B, et al. Measurement equivalence of patient-reported outcome measure response scale types collected using bring your own device compared to paper and a provisioned device: results of a randomized equivalence trial. Value Health. 2018;21(5):581–9. https://doi.org/10.1016/j.jval.2017.10.008
11 Hudgens S, Newton L, Eremenco S, Crescioni M, Symonds T, Griffiths PCG, et al. Comparability of a provisioned device versus bring your own device for completion of patient-reported outcome measures by participants with chronic obstructive pulmonary disease: quantitative study findings. J Patient Rep Outcomes. 2022;6(1):119. https://doi.org/10.1186/s41687-022-00521-3
12 Newton L, Knight-West O, Eremenco S, Hudgens S, Crescioni M, Symonds T, et al. Comparability of a provisioned device versus bring your own device for completion of patient-reported outcome measures by participants with chronic obstructive pulmonary disease: qualitative interview findings. J Patient Rep Outcomes. 2022;6(1):86. https://doi.org/10.1186/s41687-022-00492-5
13 Pugliese L, Woodriff M, Crowley O, Lam V, Sohn J, Bradley S. Feasibility of the "bring your own device" model in clinical research: results from a randomized controlled pilot study of a mobile patient engagement tool. Cureus. 2016;8(3):e535. https://doi.org/10.7759/cureus.535
14 Basch E, Deal AM, Dueck AC, Scher HI, Kris MG, Hudis C, et al. Overall survival results of a trial assessing patient-reported outcomes for symptom monitoring during routine cancer treatment. Jama. 2017;318(2):197–8. https://doi.org/10.1001/jama.2017.7156
15 Basch E, Schrag D, Henson S, Jansen J, Ginos B, Stover AM, et al. Effect of electronic symptom monitoring on patient-reported outcomes among patients with metastatic cancer: a randomized clinical trial. Jama. 2022;327(24):2413–22. https://doi.org/10.1001/jama.2022.9265
16 Beauchamp UL, Pappot H, Holländer-Mieritz C. The use of wearables in clinical trials during cancer treatment: systematic Review. JMIR Mhealth Uhealth. 2020;8(11):e22006. https://doi.org/10.2196/22006
17 Nyrop KA, Deal AM, Choi SK, Wagoner CW, Lee JT, Wood WA, et al. Measuring and understanding adherence in a home-based exercise intervention during chemotherapy for early breast cancer. Breast Cancer Res Treat. 2018;168(1):43–55. https://doi.org/10.1007/s10549-017-4565-1
18 Low CA, Dey AK, Ferreira D, Kamarck T, Sun W, Bae S, et al. Estimation of symptom severity during chemotherapy from passively sensed data: exploratory study. J Med Internet Res. 2017;19(12):e420. https://doi.org/10.2196/jmir.9046
19 Ohri N, Halmos B, Bodner WR, Cheng H, Guha C, Kalnicki S, et al. Daily step counts: a new prognostic factor in locally advanced non-small cell lung cancer? Int J Radiat Oncol Biol Phys. 2019;105(4):745–51. https://doi.org/10.1016/j.ijrobp.2019.07.055
20 Ohri N, Kabarriti R, Bodner WR, Mehta KJ, Shankar V, Halmos B, et al. Continuous activity monitoring during concurrent chemoradiotherapy. Int J Radiat Oncol Biol Phys. 2017;97(5):1061–5. https://doi.org/10.1016/j.ijrobp.2016.12.030
21 Phillips SM, Awick EA, Conroy DE, Pellegrini CA, Mailey EL, McAuley E. Objectively measured physical activity and sedentary behavior and quality of life indicators in survivors of breast cancer. Cancer. 2015;121(22):4044–52. https://doi.org/10.1002/cncr.29620
22 Jonker LT, Hendriks S, Lahr MM, van Munster BC, de Bock GH, van Leeuwen BL. Postoperative recovery of accelerometer-based physical activity in older cancer patients. Eur J Surg Oncol. 2020;46(11):2083–90. https://doi.org/10.1016/j.ejso.2020.06.012
23 Peddle-McIntyre CJ, Cavalheri V, Boyle T, McVeigh JA, Jeffery E, Lynch BM, et al. A Review of accelerometer-based activity monitoring in cancer survivorship research. Med Sci Sports Exerc. 2018;50(9):1790–801. https://doi.org/10.1249/MSS.0000000000001644
24 Low CA. Harnessing consumer smartphone and wearable sensors for clinical cancer research. NPJ Digit Med. 2020;3:140. https://doi.org/10.1038/s41746-020-00351-x
25 Dunn J, Kidzinski L, Runge R, Witt D, Hicks JL, Schüssler-Fiorenza Rose SM, et al. Wearable sensors enable personalized predictions of clinical laboratory measurements. Nat Med. 2021;27(6):1105–12. https://doi.org/10.1038/s41591-021-01339-0
26 Picard RW, Boyer EW. Smartwatch biomarkers and the path to clinical use. Med. 2021;2(7):797–9. https://doi.org/10.1016/j.medj.2021.06.005
27 Galarnyk M, Quer G, McLaughlin K, Ariniello L, Steinhubl SR. Usability of a wrist-worn smartwatch in a direct-to-participant randomized pragmatic clinical trial. Digit Biomark. 2019;3(3):176–84. https://doi.org/10.1159/000504838
28 Tedesco S, Barton J, O'Flynn B. A Review of activity trackers for senior citizens: research perspectives, commercial landscape and the role of the insurance industry. Sensors. 2017;17(6):1277. https://doi.org/10.3390/s17061277
29 Demanuele C, Lokker C, Jhaveri K, Georgiev P, Sezgin E, Geoghegan C, et al. Considerations for conducting bring your own "device" (BYOD) clinical studies. Digit Biomark. 2022;6(2):47–60. https://doi.org/10.1159/000525080
30 Ratitch B, Rodriguez-Chavez IR, Dabral A, Fontanari A, Vega J, Onorati F, et al. Considerations for analyzing and interpreting data from biometric monitoring technologies in clinical trials. Digit Biomark. 2022;6(3):83–97. https://doi.org/10.1159/000525897
31 Mâsse LC, Fuemmeler BF, Anderson CB, Matthews CE, Trost SG, Catellier DJ, et al. Accelerometer data reduction: a comparison of four reduction algorithms on select outcome variables. Med Sci Sports Exerc. 2005;37(11 Suppl l):S544–54. https://doi.org/10.1249/01.mss.0000185674.09066.8a
32 Oakley-Girvan I, Yunis R, Fonda SJ, Neeman E, Liu R, Aghaee S, et al. Usability evaluation of mobile phone technologies for capturing cancer patient-reported outcomes and physical functions. Digit Health. 2023;9:20552076231186515. https://doi.org/10.1177/20552076231186515
33 Swartz AM, Cho CC, Welch WA, Widlansky ME, Maeda H, Strath SJ. Pattern analysis of sedentary behavior change after a walking intervention. Am J Health Behav. 2018;42(3):90–101. https://doi.org/10.5993/AJHB.42.3.9
34 Diaz KM, Howard VJ, Hutto B, Colabianchi N, Vena JE, Blair SN, et al. Patterns of sedentary behavior in US middle-age and older adults: the REGARDS study. Med Sci Sports Exerc. 2016;48(3):430–8. https://doi.org/10.1249/MSS.0000000000000792
35 Migueles JH, Cadenas-Sanchez C, Ekelund U, Delisle Nyström C, Mora-Gonzalez J, Löf M, et al. Accelerometer data collection and processing criteria to assess physical activity and other outcomes: a systematic Review and practical considerations. Sports Med. 2017;47(9):1821–45. https://doi.org/10.1007/s40279-017-0716-0
36 Therneau T. A package for survival analysis in R. R Package Version. 2023;3:5–7. https://CRAN.R-project.org/package=survival

37 Therneau TM, Grambsch PM. Modeling survival data: extending the Cox model. New York: Springer; 2000.
38 Kassambara A, Kosinski M, Biecek P. Survminer: drawing survival curves using "Ggplot2,". 2021. https://CRAN.R-project.org/package=survminer
39 Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: machine learning in Python. J Machine Learn Res. 2011;12(85):2825–30.
40 Suthaharan S. Decision tree learning. In: Machine learning models and algorithms for big data classification: thinking with examples for effective learning. 2016; p. 237–69.
41 R foundation for statistical computing. R core team, R: a language and environment for statistical computing. Vienna, Austria; 2021. https://www.R-project.org/
42 Caru M, Abdullah S, Qiu L, Kanski B, Gordon B, Truica CI, et al. Women with metastatic breast cancer don't just follow step-count trends, they exceed them: an exploratory study. Breast Cancer Res Treat. 2023;200(2):265–70. https://doi.org/10.1007/s10549-023-06980-6
43 Loughney L, West MA, Dimitrov BD, Kemp GJ, Grocott MP, Jack S. Physical activity levels in locally advanced rectal cancer patients following neoadjuvant chemoradiotherapy and an exercise training programme before surgery: a pilot study. Perioper Med. 2017;6:3. https://doi.org/10.1186/s13741-017-0058-3
44 Rodríguez-Torres J, López-López L, Cabrera-Martos I, Quero-Valenzuela F, Cahalin LP, Cebrià-I-Iranzo MDÀ, et al. Inpatient step counts, symptom severity, and perceived health status after lung resection surgery. Cancer Nurs. 2021;44(5):361–8. https://doi.org/10.1097/NCC.0000000000000810
45 Withycombe JS, McFatrich M, Hinds PS, Bennett A, Lin L, Maurer SH, et al. Can steps per day reflect symptoms in children and adolescents undergoing cancer treatment? Cancer Nurs. 2022;45(5):345–53. https://doi.org/10.1097/NCC.0000000000001062
46 Gururaj R, Samuel SR, Vijaya Kumar K, Hegde A, Prakash Saxena PU, Nagaraja R, et al. Relationship between physical activity, objective sleep parameters, and circadian rhythm in patients with head and neck cancer receiving chemoradiotherapy: a longitudinal study. Laryngoscope Investig Otolaryngol. 2021;6(6):1455–60. https://doi.org/10.1002/lio2.664
47 Henriksen A, Johansson J, Hartvigsen G, Grimsgaard S, Hopstock L. Measuring physical activity using triaxial wrist worn polar activity trackers: a systematic Review. Int J Exerc Sci. 2020;13(4):438–54. https://doi.org/10.70252/RPTD5267
48 Veerabhadrappa P, Moran MD, Renninger MD, Rhudy MB, Dreisbach SB, Gift KM. Tracking steps on Apple watch at different walking speeds. J Gen Intern Med. 2018;33(6):795–6. https://doi.org/10.1007/s11606-018-4332-y