

Research Report – Research Article

# Observational Study of a Wearable Sensor and Smartphone Application Supporting Unsupervised Exercises to Assess Pain and Stiffness

Caroline G.M. Perraudin<sup>a</sup> Vittorio P. Illiano<sup>a</sup> Francesc Calvo<sup>a</sup>  
Emer O'Hare<sup>a</sup> Seamas C. Donnelly<sup>b</sup> Ronan H. Mullan<sup>b</sup> Oliver Sander<sup>a</sup>  
Brian Caulfield<sup>c</sup> Jonas F. Dorn<sup>a</sup>

<sup>a</sup>Digital Development, Novartis Pharma AG, Basel, Switzerland; <sup>b</sup>Tallaght Hospital, Trinity College Dublin, Tallaght, Ireland; <sup>c</sup>University College Dublin, Belfield, Ireland

## Keywords

Non-interventional study · Actigraphy · Arthritis · Chronic pain · Remote monitoring

## Abstract

**Background:** Evaluation of pain and stiffness in patients with arthritis is largely based on participants retrospectively reporting their self-perceived pain/stiffness. This is subjective and may not accurately reflect the true impact of therapeutic interventions. We now have access to sensor-based systems to continuously capture objective information regarding movement and activity. **Objectives:** We present an observational study aimed to collect sensor data from participants monitored while performing an unsupervised version of a standard motor task, known as the Five Times Sit to Stand (5×STS) test. The first objective was to explore whether the participants would perform the test regularly in their home environment, and do so in a correct and consistent manner. The second objective was to demonstrate that the measurements collected would enable us to derive an objective signal related to morning pain and stiffness. **Methods:** We recruited a total of 45 participants, of whom 30 participants fulfilled pre-defined criteria for osteoarthritis, rheumatoid arthritis, or psoriatic arthritis and 15 participants were healthy volunteers. All participants wore accelerometers on their wrists, day and night for about 4 weeks. The participants were asked to perform the 5×STS test in their own home environment at the same time in the morning 3 times per week. We investigated the relationship between pain/stiffness and measurements collected during the 5×STS test by comparing the 5×STS test duration with the patient-reported outcome (PRO) questionnaires, filled in via a smartphone. **Results:** During the study, we successfully captured accelerometer data from each participant for a period of 4 weeks. The participants performed 56% of the prescribed 5×STS tests. We observed that different tests made by the same participants were performed with subject-specific characteristics that remained consistent throughout the

Jonas F. Dorn  
Digital Development, Novartis Campus, Novartis Pharma AG  
Fabrikstrasse 2  
CH-4056 Basel (Switzerland)  
E-Mail [jonas.dorn@novartis.com](mailto:jonas.dorn@novartis.com)

study. We showed that 5×STS test duration (the time taken to complete the 5×STS test) was significantly and robustly associated with the pain and stiffness intensity reported via the PROs, particularly the questions asked in the morning. **Conclusions:** This study demonstrates the feasibility and usefulness of regular, sensor-based, monitored, unsupervised physical tests to objectively assess the impact of disease on function in the home environment. This approach may permit remote disease monitoring in clinical trials and support the development of novel endpoints from passively collected actigraphy data.

© 2018 The Author(s)  
 Published by S. Karger AG, Basel

## Introduction

Arthritis is an all too common and disabling condition affecting the musculoskeletal system [1]. In the period between 2013 and 2015, over 54 million people in the USA alone – approximately 22% of the population – had a diagnosis of some form of arthritis, with an estimated annual direct and indirect cost of over USD 300 billion [2]. Patients with arthritis frequently report symptoms such as morning pain and stiffness, leading to limitations at a functional level and impaired quality of life [1, 3–5]. It is therefore logical that therapeutic interventions for arthritis would specifically target this period of the day and that any evaluation of their efficacy would focus on their direct impact on morning pain and stiffness [6].

To date, the primary means of evaluating morning pain and stiffness has been the use of subjective rating scales [7, 8]. There are inherent issues with the use of such scales, as they rely entirely on subjective reporting. As such, they are limited as a means of providing objective and reliable measures of the impact of therapeutic interventions [9]. It is crucial that we develop new methodologies that facilitate the accurate and objective measurement of symptoms such as morning pain and stiffness. The best way to achieve this is to directly measure the impact of morning pain and stiffness on motor function, focusing on tasks such as stiffness to stand, and gait.

Until recently, this was achieved through direct measurement by clinicians. This has significant drawbacks, as there is necessarily a time gap between the patient waking up and the visit. Moreover, such a solution involves significant costs. A potential solution lies in using wearable sensor technologies to facilitate the objective measurement of appropriate motor tasks [10]. A good compromise would be to passively capture the data on sensors worn by the patients, and to infer the degree of pain and stiffness from it. However, while some motor tasks have been standardized and their relationship to pain and stiffness benefits from broad coverage, there is no obvious connection between data collected during normal activities and pain/stiffness. The approach followed in this paper was to have patients perform standard tests in their own homes while wearing the sensors. Our objective was to investigate whether this trade-off solution kept the reliability of standard motor tasks and the efficiency of passive monitoring.

To date, wearable sensor technologies have not been widely used in this way due to concerns as to whether patients could carry out motor tasks in a highly standardized manner such that the resultant data would accurately reflect meaningful change in their status. Moreover, it is not obvious whether patients would remember and adhere to performing such tests as prescribed, or whether sensor devices could accurately capture the relevant data.

We have run a non-interventional study on a population of subjects with arthritis and healthy individuals which was monitored for 4 weeks, during which time they were prescribed the Five Times Sit to Stand (5×STS) test 3 mornings per week. Previous work has investigated instrumented 5×STS tests to (1) compare the instrumented STS test to the manually recorded STS test [11], (2) validate the use of sensors for data capture against gold-standard motion

**Table 1.** Distribution of patients and HVs

Disease type	Age, years	Females, %	Total number in category
RA	50.7, 33–75, 11.4	72	18
PA	47.5, 32–63, 15.5	100	2
OA	60.7, 55–70, 4.5	60	10
HV	48, 31–71, 13.6	66	15

Age is given as the mean, range, and STD. RA, rheumatoid arthritis; PA, psoriatic arthritis; OA, osteoarthritis; HV, healthy volunteer.

capture systems [12], (3) detect sit to stand transitions in free living conditions [13], and (4) understand how to best identify transitions between phases in the 5×STS test [14].

In contrast, our goals were to (a) evaluate whether or not study volunteers could reliably perform the 5×STS tests unsupervised and (b) investigate the relationship of the duration of the 5×STS test (extracted from sensor data) to pain and stiffness.

## Subjects and Methods

### Study Design

#### Recruited Participants

In total, 45 subjects participated in this study. A group of 30 patients with arthritis was selected according to the distribution of age and gender summarized in Table 1. Among them, 18 patients had rheumatoid arthritis, 2 patients had psoriatic arthritis, and 10 patients had osteoarthritis. For personal reasons, 2 of the 10 osteoarthritis patients chose to withdraw from the study on the 23rd and the 29th day of analysis, respectively. Since rheumatoid arthritis and psoriatic arthritis are two types of inflammatory arthritis, the patients from the two groups were analysed together. In addition to the patients with arthritis, a group of 15 healthy volunteers was selected according to the distribution of age and gender summarized in Table 1. The healthy volunteers were recruited through University College Dublin and the arthritis patient cohort was recruited through Tallaght Hospital via Trinity College Dublin.

#### Deployment Plan

Every participant was equipped with the ActiGraph GT9X Link device [15], which was configured to capture accelerometer data at 30 Hz whilst being worn on the wrist (the participants were free to choose which one) throughout the study period. Raw acceleration data were extracted at the end of the study through the USB interface. The participants were asked to charge the device's battery, which had an average lifetime of around 1 week, and were reminded if they did not. A smartphone was used with a pre-installed app (CentrosHealth, Boston, MA, USA) [16], which collected patient-reported outcomes (PROs) regarding the degree of pain and stiffness perceived. Reminders were sent if the PROs were not filled out. The data from the sensor devices were periodically uploaded to a cloud system in an aggregated format (1 sample/min), which enabled us to run preliminary analyses (e.g., calculating the percentage of wear time). However, the results presented in this paper have been obtained through offline analyses, which were run on the 30-Hz data retrieved at the end of the study.

#### Patient-Reported Outcomes

The PRO consists of 8 questions with instructions for completion: 4 items related to stiffness and 4 related to pain. The stiffness-related questions and pain-related questions can

be further divided into questions asked in the morning, i.e., immediately following waking up, and questions asked in the evening.

The full set of 8 questions includes:

1. *Stiffness suffered*: morning question (“yes” or “no”). The value is “yes” if the participant feels stiffness of any degree, and “no” if no stiffness is suffered
2. *Stiffness severity*: morning question. The value rates stiffness upon waking (0–4 numerical scale, where severity increases with the score)
3. *Stiffness severity during day*: evening question. The value rates stiffness over the course of the day (0–4 numerical scale, where severity increases with the score)
4. *Stiffness duration*: evening question. The value measures the duration of stiffness from waking up to the moment when it disappears. The answer is a multiple choice between “30 min,” “30 min to 1 h,” “1–2 h,” “2–4 h,” “4 h,” “all day,” and “none”
5. *Pain waking up*: morning question. The value rates pain severity upon waking (0–10 numerical scale, where severity increases with the score)
6. *Pain overnight*: morning question. The value rates pain severity over the course of the previous night (0–10 numerical scale, where severity increases with the score)
7. *Pain since getting up*: evening question. The value rates pain severity over the course of the day (0–10 numerical scale, where severity increases with the score)
8. *Pain last 24 h*: evening question. The value rates pain severity over the course of the last 24 h (0–10 numerical scale, where severity increases with the score)

The text of the questions that appeared on the participants’ smartphone is given in the Appendix.

#### Onsite Visits

At the start of the study, an investigation team visited the participants at home and provided them with the ActiGraph devices. This team had the following duties: (1) to set up the CentrosHealth app, train the participants to use it, and provide written instructions and (2) to show how to perform the 5×STS test, ask the participants to complete a test under their supervision, and record test date and time. At the end of the study, the patients were visited at home again. The investigator supervised the participants performing the 5×STS test again and collected the sensor devices.

#### Data Transformation

##### Data Set Creation

The data set was divided into development, test, and validation data sets. We assigned 24 participants to the development set, 11 to the test set, and 10 to the validation set. We used the development participants to select the model, the test participants to confirm model selection, and the validation participants to check whether the model generalized to other patients. Moreover, we took aside 5 days from each participant and added them to the validation set to validate the participant-specific model and ensure that it generalized for a given participant. The development data set was used to investigate the correlations between PRO, 5×STS test duration, and other covariates, i.e., age, gender, BMI, disease type, and seconds since getting up. We calculated the last through a sleep detection algorithm, applied to the actigraphy data, which uses a combination of Cole-Kripke [17] and Tudor-Locke [18] algorithms.

The choice of covariates was evaluated on the test set. A model was fitted on the development set, and its accuracy was tested on the test set, after which adjustments were permitted. Then, the adjusted model was tested on the validation set, and on the union of the test and validation sets.

### Time Reference

For more convenient visualization, the sensor data for each participant were organized into data analysis days (DADs), which are 24 h starting and ending at 18:00. The day on which recording starts is set to DAD 0, so that DAD 1 is the first day within the study with 24 h of data (assuming full adherence; all deployments were completed before 18:00). This approach using DADs instead of standard calendar days ensured that the overnight sleep session did not get split across 2 days when a participant went to sleep before midnight.

### Change of Coordinates

The raw accelerometer data were transformed from Cartesian  $(x, y, z)$  to spherical  $(r, \text{azimuth}, \text{elevation})$  coordinates [19], where  $r$  is the vector norm of the acceleration, azimuth is the rotation of the accelerometer in the plane of the device/watch face, and elevation is the tilt relative to the plane of the device/watch face (an example of accelerometry data in spherical coordinates is given in the Appendix). With this transformation, it was easier to observe movements that characterize the 5×STS test, such as arm position and chest tilt.

### Corrections to Accelerometer Signals

To ensure that acceleration was measured correctly in all directions of space, the data were autocalibrated to ensure that the sensor would report 1  $g$  of acceleration while at rest. Autocalibration was performed according to the method described by van Hees et al. [20]. Briefly, the method identifies the periods where the sensor is at rest and measures the acceleration in the given sensor orientation. A generalized ellipsoid is fitted into the distribution of points and used to normalize the acceleration values reported by the sensor. We selected data at rest from the start of recording until we had at least 50 points in each of 6 equal sectors of the sphere (2 polar and 4 equatorial), and resampled the data before regression to ensure a reasonably uniform coverage of orientations to avoid bias.

### 5×STS Tests

#### Test Instructions

The participants were asked to perform the 5×STS test without supervision 3 times per week (on Mondays, Wednesdays, and Fridays) while wearing the sensor device on their wrist. We decided not to use a daily schedule for two reasons. The first reason is that we wanted to avoid training effects, i.e., increases in performance caused by improvements in the ability to take the test, which are not connected to health improvements. The second reason is that we wanted to reduce the patients' burden, who additionally had to fill in the PROs on the smartphone, wear the sensor device, and charge the battery for both the smartphone and the sensor device.

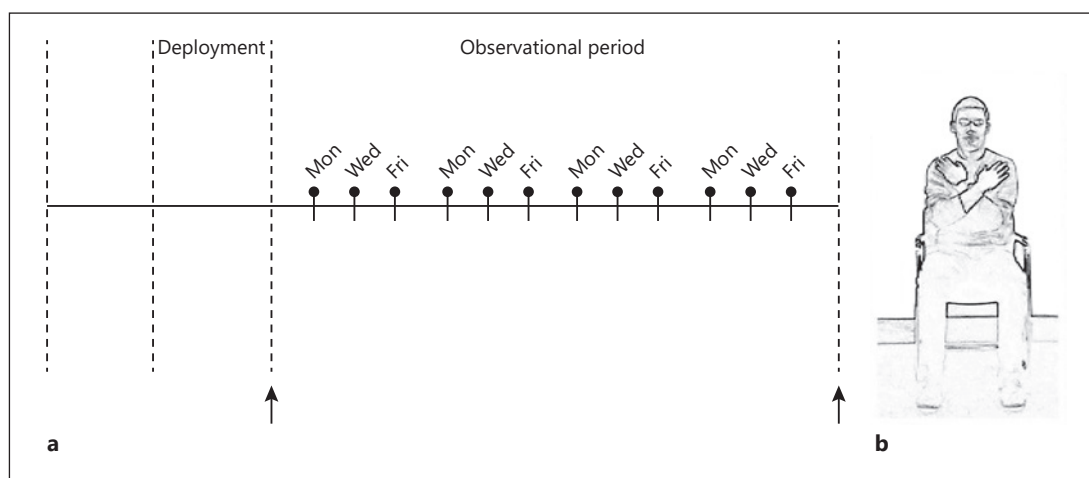
The average number of expected unsupervised tests was 12; however, this could vary based on the effective days of analysis, as well as on the days of the week on which a participant started and ended.

Figure 1a shows the 5×STS test execution plan. The 5×STS exercise consists of 5 cycles of standing and sitting, with arms crossed over the chest (Fig. 1b).

#### Adherence to 5×STS Tests

To extract the 5×STS test data and measure the participants' adherence, we built a detection algorithm to automatically detect 5×STS execution (see Appendix).

The algorithm was based on the assumptions that (1) the arms were crossed on the chest, (2) the trunk moved back and forth whilst the participant was sitting and standing, and (3) the norm of the acceleration followed a reproducible pattern, which we observed in the development set. The algorithm's parameters were calibrated on about 90 5×STS tests, which were extracted from 6 participants who belonged to the development set.



**Fig. 1. a** Overview of expected Five Times Sit to Stand tests. The participants performed two tests under the supervision of an assistant, one at the beginning and one at the end of the study. Additionally, the participants were asked to perform the tests without supervision 3 times per week, on Mondays, Wednesdays, and Fridays. **b** Arm position during the Five Times Sit to Stand test.

When expected 5×STS tests could not be found, we reverted to manual inspection, and if this failed too, we marked the test as not performed. We detected 56% of the prescribed 5×STS tests, of which 80% were automatically detected, while 20% were manually detected. All detected tests were manually reviewed, as were all the days on which a 5×STS test had been expected to occur but was not identified automatically.

#### Consistency of 5×STS Tests

To measure the consistency of the participants in performing 5×STS tests, we compared the data collected during different tests through dynamic time warping (DTW) [21]. DTW is a method of calculating the distance between time series through dilation of the shortest 5×STS test time series in such a way that the distance between the two time series is minimized. As a common choice, we considered the norm of the difference as the element-wise distance, whose sum is the quantity minimized by DTW. In addition, we normalized the result by the length of the longest time series, to make distances comparable for different lengths of the time series.

The main reason behind the choice of such a distance metric is to compare the main characteristics of how participants perform the test while abstracting from timing characteristics, e.g., speed of execution, possible pauses between any of the 5 iterations, etc. The latter characteristics are pertinent to the performance of the test, which is likely to bias the consistency metric, since the performances of the same participant on different tests are likely to be correlated.

#### *Pain and Stiffness Prediction Model*

To estimate how stiffness and pain were related to the duration of the 5×STS test and other covariates, we made a regression analysis based on a linear mixed-effects model. Such a model can cope with both fixed effects, i.e., the conventional linear regression part, and random effects, i.e., individual experimental units drawn at random from a population. The data collected for the study presented in this paper are grouped by participant, which constitutes the random experimental component. Intuitively, the data collected from each participant are correlated; however, this is not necessarily also the case for data gathered from

different participants, as there could be differences between individuals in baseline data on the reported pain and 5×STS test performance.

We built a model for the prediction of pain and stiffness where the key independent variable is the duration of the 5×STS test. Duration is one of the most widely used parameters to evaluate the test's performance [22]. Other parameters have been proposed, such as the movements' smoothness [23] or the duration of the single phases (sit to stand and stand to sit) [24]. Even though these parameters may lead to better results, they are more difficult to estimate than overall duration; hence, they are more likely to introduce a measurement error. Moreover, such parameters have been validated in studies where elderly participants are compared to younger ones. However, the effects of pain may be different – for instance, the time spent standing and sitting may be representative of the patient's need to recover before starting the next sit-to-stand cycle.

To increase the prediction accuracy, we also considered other covariates that possibly affect pain and stiffness, i.e., age, gender, BMI, seconds since getting up, and disease type.

## Results

### *Participants' Adherence*

The participants wore the sensors day and night for about 91.5% of the time on average (about 2% of the missing data are due to the device charging times). Moreover, 88.3% of the PRO questionnaires were completed (including those of the participants who withdrew consent), probably reflecting the use of reminders sent through the smartphone app (indeed, we observed peaks in adherence right after the daily reminders had been sent out at 10:00 and 11:00, and at 19:00 and 21:00, respectively). In addition, 56% of the 5×STS tests were performed, even though the reminders used were transient (i.e., a banner that would disappear as soon as the smartphone was unlocked). The per-patient and condition-aggregated adherence regarding sensor wear time and 5×STS tests is given in Tables A1 and A2 in the Appendix.

In Figure 2a, we plotted the time span between the first and the last test, normalized by the total days of analysis, versus the number of 5×STS tests performed, normalized by the total number of prescribed tests. With this visualization, we could analyse the adherence to the 5×STS tests and distinguish different adherence patterns: the participants who performed the 5×STS test regularly for a certain duration but then stopped (points on the diagonal) from those who forgot to perform some tests but overall stuck to performing the 5×STS tests as prescribed throughout the study (points spread across the top), and those who were highly adherent regarding both aspects (points in the top right). As shown in Figure 2a, there was a considerably higher number of participants with high adherence (participated in performing the tests for more than 80% of the time, completed more than 20% of the tests) than participants with low adherence (participated in performing the tests for less than 80% of the time, completed less than 20% of the tests), which is significant according to the  $\chi^2$  test. Moreover, we observed no significant difference in the number of 5×STS tests performed between healthy volunteers and patients. To check whether they followed different probability distributions, we ran the Kolmogorov-Smirnov test (at the 5% significance level), which could not reject the hypothesis that they came from the same distribution.

### *Consistency of 5×STS Execution*

Consistency of 5×STS execution was assessed by comparing the accelerometry traces within and between participants via the DTW-based distance (Fig. 2b). The DTW-based distance metric is sensitive to any difference in execution of the test, including, for example,

the position of the arms, the amount of sideways sway during the test, or the intensity of the forward/backward movement of the trunk.

On the other hand, DTW adjusts for speed differences. This choice ensures a performance-independent measure of consistency, which also prevents bias from participants who perform the test at similar speeds. Moreover, we compared the data from different 5×STS tests without compensating for possible consistency-degrading factors due to, e.g., more pain suffered, a wrong sensor position, etc.

We used the values of the DTW distances to make relative comparisons between different 5×STS tests. In particular, we wished to check whether supervision played an important role by comparing distances between two supervised tests with distances between a supervised and an unsupervised test. Moreover, we wished to compare the distance between two tests performed by the same subject with the distances between two tests performed by different subjects, to measure the degree of consistency for each participant in performing the test.

From Figure 2b we observed no evident increase in distances between supervised and unsupervised tests when compared to the distances between two unsupervised tests. In contrast, two unsupervised tests made by the same person were remarkably more similar than two unsupervised tests made by different participants. This result suggests that there was no systematic difference in performance of supervised and unsupervised tests; thus, the participants were consistent in the way they performed the 5×STS test and they complied with the supervisor's instructions, even when unsupervised. In contrast, there was an evident difference in test execution between two different participants. In conclusion, the participants performed the 5×STS test in an individually consistent manner.

We applied the two-sample Mann-Whitney U test (at the 5% significance level) to find significance differences between different groups (we chose this test as the data are not normally distributed). The test could not reject the hypothesis that the distances between two supervised tests and the distances between a supervised and an unsupervised test came from the same distribution. In contrast, the test rejected the hypothesis that the distances between two tests performed by the same participant and the distances between two tests performed by different participants came from the same distribution.

The outlying point at the top right of Figure 2b corresponds to a participant who kept a wrong arm position during unsupervised tests (the arms were not crossed with the hands on the chest), despite performing a correct supervised test during the entry visit. This participant was the only one by whom the tests were performed with noticeable inconsistency. At the bottom left of the left figure, we may observe a group of 5 points that are further from the diagonal. These points relate to those participants who kept a wrong arm position during supervised tests. For three of them, we detected a wrong supervised test during the first visit of the instructor; hence, the supervisor, who happens to have been the same for all of them, did not instruct these participants correctly. Interestingly, those participants' unsupervised tests were adherent to the instructions given in the leaflet, even if their supervised test during the entry interview was not.

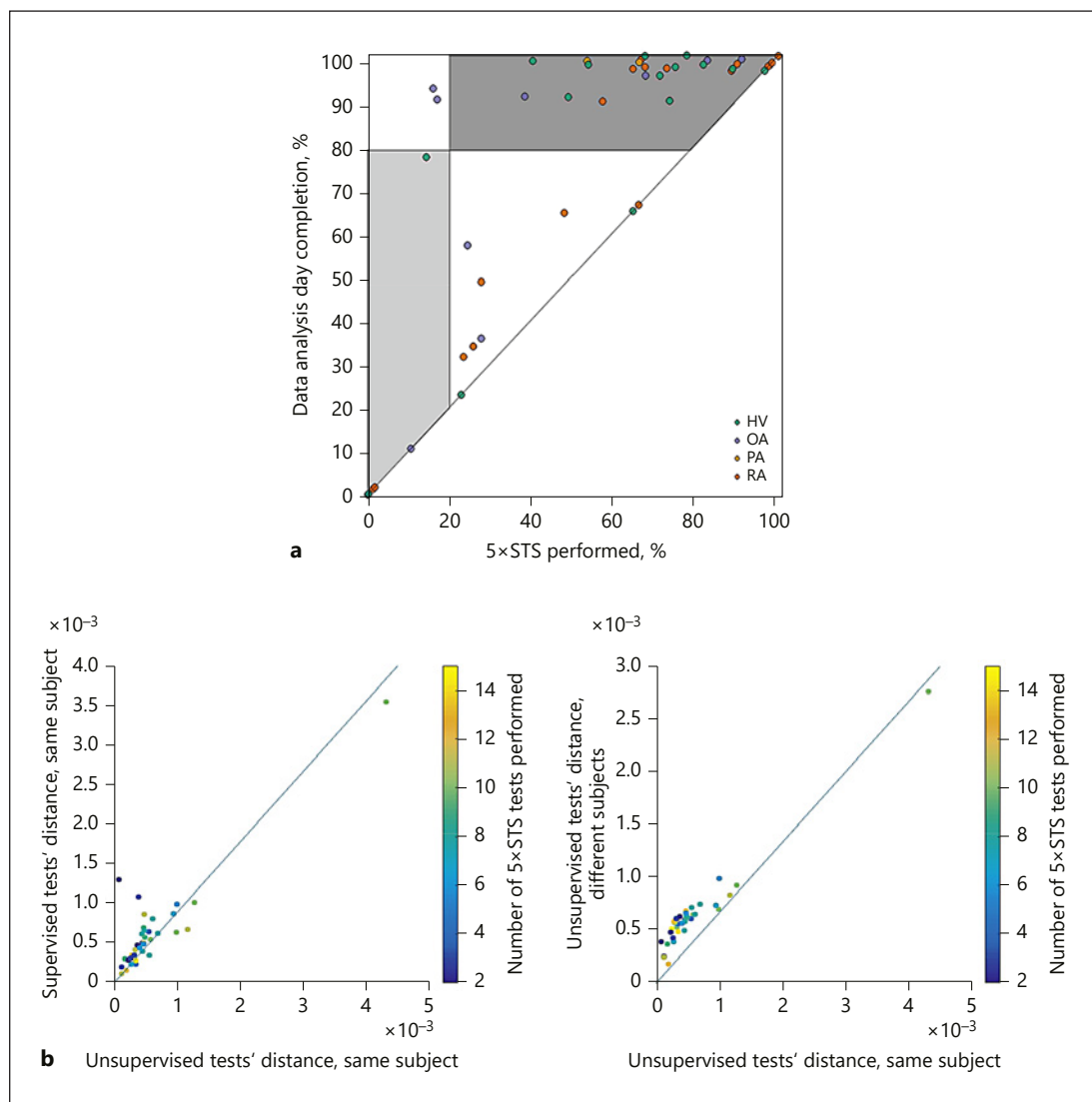
In summary, each participant performed unsupervised tests with high consistency, compared to the remarkable difference between tests performed by different participants.

#### *Pain and Stiffness Prediction Model*

##### Model Creation

We fitted a random intercept model where each participant was given an individual intercept in the regression line. An additional individual slope term was also investigated, but it was observed that the regression would not converge, possibly due to the small number of participants. Thus, one general slope term based on all the data was used for each participant. In the mixed-effects model, the duration of the 5×STS test (gathered from the sensor data)





**Fig. 2. a** Study completion percentage against percentage of Five Times Sit to Stand (5xSTS) tests performed. Light grey area: low adherence (participated in performing the tests for less than 80% of the time, completed less than 20% of the tests). Dark grey area: high adherence (participated in performing the tests for more than 80% of the time, completed more than 20% of the tests). There were more participants with high adherence (top right corner) than participants with low adherence (bottom left corner). Moreover, we observed a group of participants whose adherence was around 50%, but they took their tests until the end of the study (top centre side). The latter possibly were participants who forgot to take some tests. HV, healthy volunteers; OA, osteoarthritis; PA, psoriatic arthritis; RA, rheumatoid arthritis. **b** Comparison of different 5xSTS tests. Each point identifies the maximum distance, measured with dynamic time warping (DTW) [21], between two tests that are selected according to a specific criterion, which varies with the axes. On the left plot, the x axis characterizes supervised tests, while the y axis characterizes supervised tests performed by the same participant. On the right plot, the x axis characterizes tests performed by the same participant, while the y axis characterizes tests performed by different participants. The presence of a supervisor did not impact consistency (left). In contrast, different participants performed the 5xSTS tests differently (right).

was an independent variable, and pain waking up (i.e., morning pain severity, collected through PRO) was the dependent variable. Since each participant was given an individual intercept, the PRO prediction was not driven by the inter-subject differences (e.g., disease type) but by the longitudinal variables (i.e., 5×STS test duration and seconds since getting up). The Akaike information criterion [25] was used as a quality measure for comparing regressions with different numbers of parameters.

Table 2 shows the statistics of linear regression performed on a set of covariates, all of which had been centred and scaled to make the coefficients comparable. Note that the lowest Akaike information criterion would have been achieved with disease, seconds since getting up, and gender. However, we decided to remove seconds since getting up, as its contribution was negligible (regression coefficient = 0.01) and it was statistically insignificant ( $t = 0.14$ ). Moreover, we decided to keep gender, even though it resulted as being slightly insignificant, as its contribution was great.

From Table 2, we observed a strong consistency for the estimated coefficients of 5×STS test duration, suggesting that its relationship to pain and stiffness was stable and reliable.

### Model Accuracy

The random intercept mixed-effects model was fitted to the development data set, and evaluated by analysing the prediction accuracy on the test data set.

Figure 3a shows the regression lines projected across the 5×STS test durations, and the samples used for fitting the model for pain severity upon waking (pain waking up). In particular, we show the global regression line, which describes the overall behaviour for all participants, and the individual regression lines, which cater for individual variations in reporting pain (affecting the PROs) and physical fitness (affecting 5×STS test performance).

Figure 3b shows the forest plots for the coefficients of 5×STS test duration in the fitted model. Forest plots summarize a correlation between two variables, here 5×STS test and pain reported through PROs, along with the confidence interval of such a correlation. When the confidence interval excludes the line that represents zero correlation, the correlation is statistically significant. We observed a positive correlation between 5×STS test duration and PROs. For most PROs (5 out of 8), this correlation was significant, as the error bars were strictly positive. The mean values for the coefficients of the questions asked in the morning (stiffness suffered, stiffness severity, pain waking up, and pain overnight) were larger than the values for the questions asked in the evenings (stiffness severity during day, stiffness duration, pain since getting up, and pain in the last 24 h).

Figure 4 shows the prediction discrepancy plots, which highlight the difference between the distribution of expected residuals from the model and the actual distribution on the data samples. The latter are taken from disjoint data sets to demonstrate the model's generality. To produce the prediction discrepancy plot, we first calculated the cumulative distribution function (CDF) of the predicted pain values, based on the covariates. Then, we calculated the CDF for the actual pain values and compared the two CDFs.

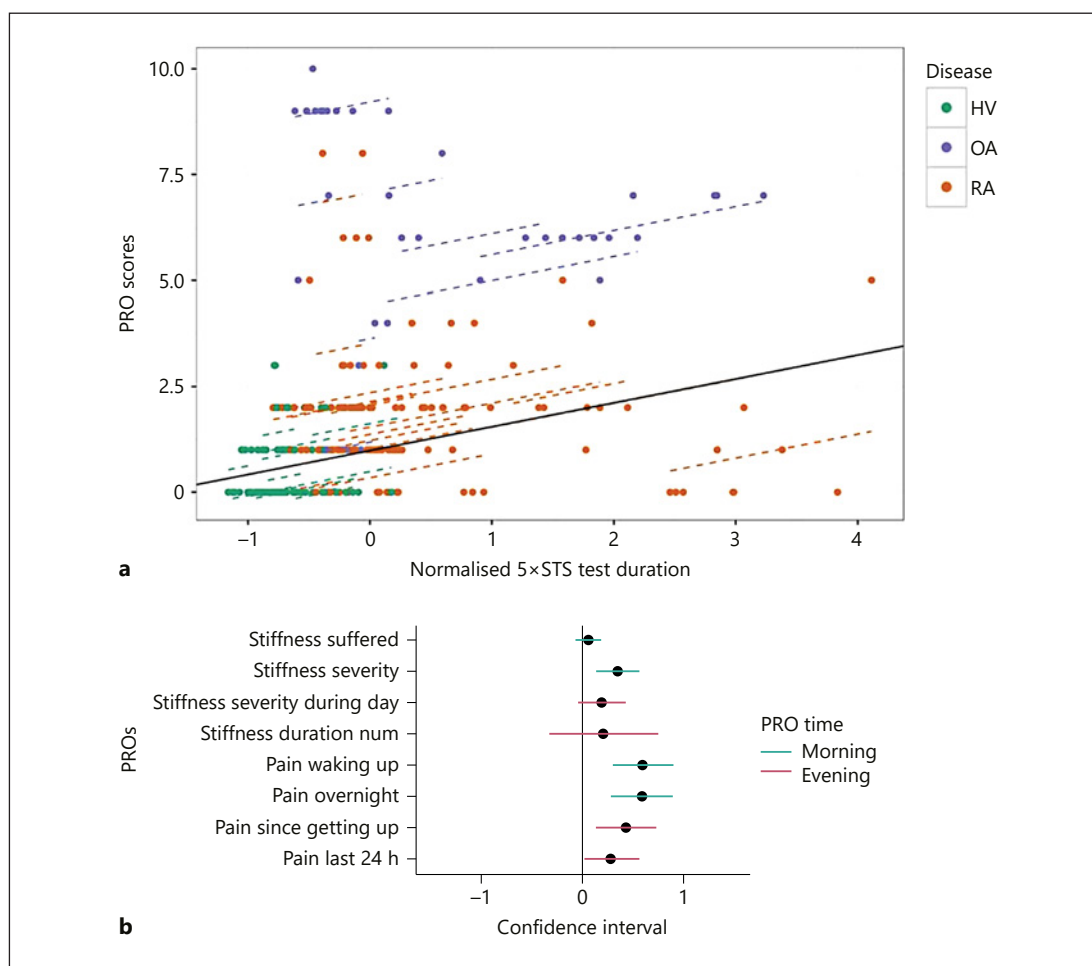
Each quantile of the expected distribution of the residuals was expected to cover that quantile's fraction of the total residuals. For example, the 0–5% quantile was expected to cover 5% of the observed residuals.

While the model overestimated responses on the test set, which resulted in more residuals than expected on the low tail of residuals, it underestimated responses on the validation set, which resulted in more residuals than expected on the high tail of residuals. In conclusion, we showed that 5×STS test duration has a positive and significant correlation with 5 out of 8 PROs. Moreover, we observed an unbiased distribution of residuals from the fit, although the frequency of very high (on the test set) and very low (on the validation set) residuals was higher than expected.

**Table 2.** Mean (standard error) for the parameters of the mixed-effects model with different combinations of covariates for model selection

Covariates, <i>n</i>	Duration in seconds (CS)	Disease OA	Disease RA	Age (CS)	BMI (CS)	Seconds since getting up (CS)	Gender	Intercept	AIC
5	0.50 (0.15)	4.7 (1.19)	0.87 (0.93)	0.12 (0.48)	0.02 (0.39)	0.01 (0.07)	-1.5 (0.86)	1.14 (0.75)	371.32
4	0.55 (0.16)	-	-	0.99 (0.57)	0.40 (0.49)	0.01 (0.07)	-1.46 (1.17)	2.67 (0.61)	383.52
4	0.49 (0.15)	4.74 (1.18)	0.89 (0.92)	0.15 (0.47)	0.02 (0.39)	-	-1.44 (0.85)	1.13 (0.74)	376.97
4	0.53 (0.15)	4.68 (1.27)	0.87 (0.99)	-0.07 (0.49)	0.06 (0.42)	0.01 (0.07)	-	0.71 (0.75)	373.77
4	0.50 (0.15)	4.82 (1.07)	0.85 (0.90)	-	0.03 (0.38)	0.01 (0.07)	-1.45 (0.81)	1.10 (0.71)	369.72
4	0.49 (0.15)	4.71 (1.09)	0.89 (0.85)	0.13 (0.46)	-	0.02 (0.07)	-1.50 (0.83)	1.13 (0.69)	369.26
3	0.56 (0.15)	-	-	1.09 (0.55)	-	0.01 (0.07)	-1.53 (1.16)	2.69 (0.61)	382.60
3	0.49 (0.15)	4.76 (1.08)	0.91 (0.84)	0.15 (0.47)	-	-	-1.44 (0.82)	1.11 (0.68)	374.89
3	0.53 (0.15)	4.73 (1.17)	0.92 (0.90)	-0.06 (0.48)	-	0.01 (0.07)	-	0.68 (0.68)	371.85
3	0.50 (0.15)	4.85 (0.96)	0.88 (0.82)	-	-	0.02 (0.07)	-1.45 (0.79)	1.08 (0.65)	367.59
2	0.58 (0.15)	-	-	-	-	0.01 (0.07)	-0.92 (1.20)	2.45 (0.64)	384.92
2	0.52 (0.15)	4.67 (1.11)	0.90 (0.95)	-	0.05 (0.4)	-	-	0.72 (0.71)	377.36
2	0.49 (0.15)	4.92 (0.95)	0.91 (0.82)	-	-	-	-1.38 (0.78)	1.05 (0.64)	373.23
2	0.53 (0.15)	4.67 (1.01)	0.93 (0.87)	-	-	0.01 (0.07)	-	0.69 (0.65)	370.19
1	0.59 (0.15)	-	-	-	-	0.01 (0.07)	-	2.19 (0.53)	385.71
1	0.52 (0.15)	4.73 (1.00)	0.95 (0.86)	-	-	-	-	0.68 (0.64)	375.61

The values on the left of the parentheses show the regression coefficients, while the values in parentheses show the standard error. The *t* value can be calculated by dividing the coefficient by the standard error. The first column shows the number of covariates (the dummy variables diseaseOA and diseaseRA are counted as one covariate). The second column shows the statistics for the regression intercept. The third column is the main independent variable, i.e., the one that varies in time and allows to predict the evolution of pain and stiffness, after centering and scaling (CS) operations (mean subtraction and division by standard deviation). The third to ninth columns show the statistics with each of the covariates. If the covariate is not present in the model, dashed lines are shown. Finally, the last column shows the Akaike information criterion (AIC). For each number of covariates, the most relevant results are shown, in terms of the AIC. RA, rheumatoid arthritis; OA, osteoarthritis.

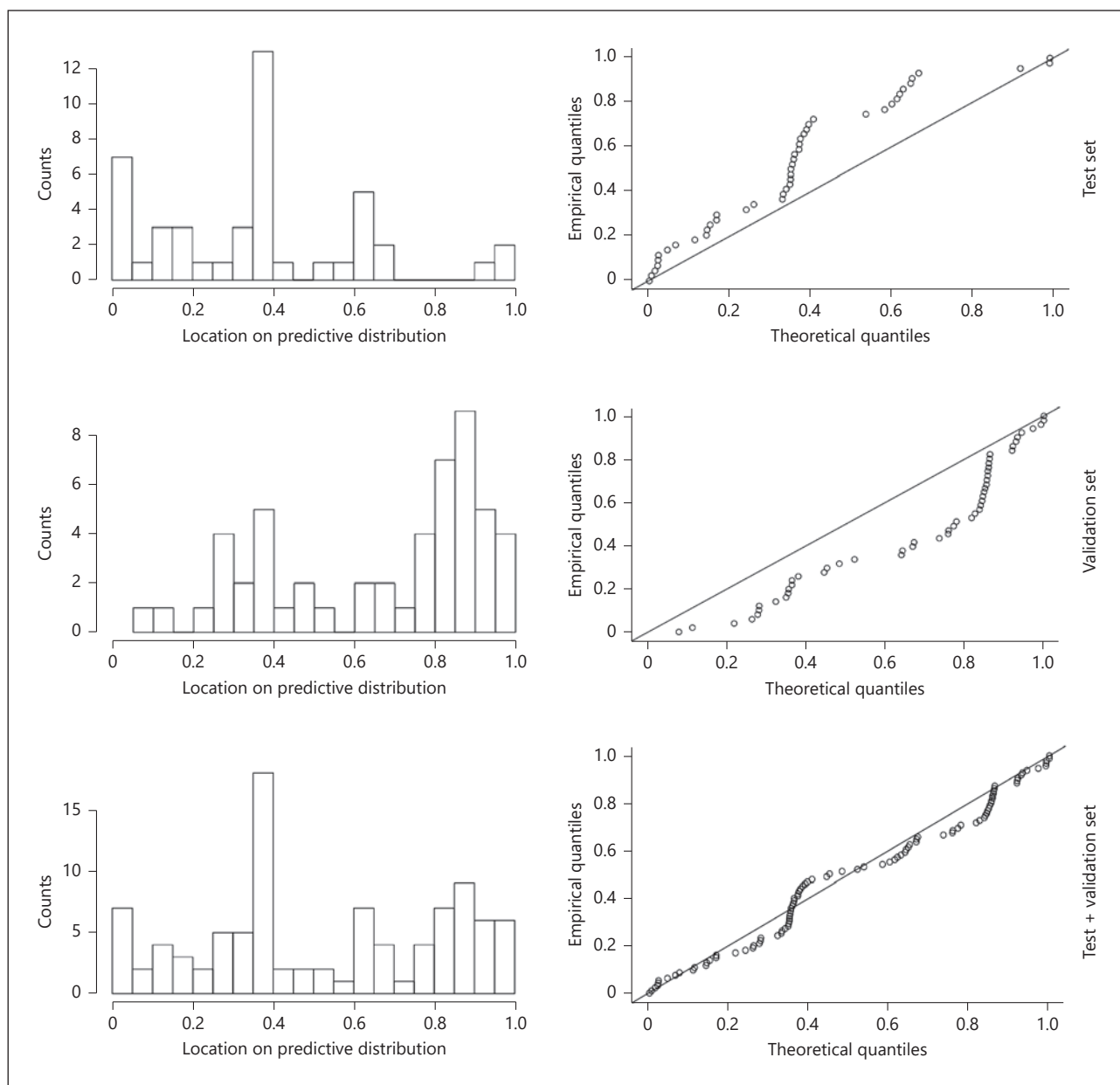


**Fig. 3. a** Prediction model of the pain waking up reported by the participants (patient-reported outcome [PRO]) as a function of Five Times Sit to Stand (5xSTS) test duration through a mixed-effects model. The solid line represents the global model, i.e., the estimates obtained over the subset of fixed-effects components that are part of the mixed-effects model. The dashed lines are obtained by adding the individual intercept to the global model. Some of the individual lines differ noticeably from the global line, which motivates our choice of an individual intercept. HV, healthy volunteers; OA, osteoarthritis; RA, rheumatoid arthritis. **b** Forest plot measuring the relationship strength between 5xSTS test duration and the pain/stiffness reported by the participants. Dots and horizontal lines show the correlation value  $\pm 2$  times the standard error. The vertical black line indicates a coefficient of zero. A high correlation between 5xSTS test duration and pain severity was demonstrated. The relationship to pain severity on waking up was strongest and highly significant. Among stiffness-related PROs, the strongest relationship was to stiffness severity rated in the morning.

### Discussion/Conclusion

The principal finding from this study is that the participants performed 56% of the prescribed 5xSTS tests, with no qualitative difference between supervised and unsupervised tests. This suggests that, with appropriate preparation, participants in clinical trials could be relied on to capture data during standardized motor task performance tests.

Many modern clinical studies on pain and stiffness rely on tests performed in the clinic [26–29], which are necessarily infrequent, as many tests are invasive and require expensive clinic times. Moreover, in the case of morning stiffness, the clinic assessment cannot be



**Fig. 4.** Prediction discrepancy plots for the predictions of pain waking up, obtained from a mixed-effects model. This was fitted on the duration of the Five Times Sit to Stand exercise, along with two covariates: gender and disease type. The graphs on the left show the histograms of the empirical cumulative distribution function (CDF) against the predicted CDF, while the graphs on the right show the predicted CDF as a function of the empirical CDF. The distribution of residuals from the fit is unbiased, although the frequency of very high and very low residuals was higher than expected.

performed shortly after waking up. An alternative approach is to use PROs [30], which enable participants to record their status at home; yet the outcome is subjective and prone to bias. Passive activity monitoring is another home assessment option [31], which is objective thanks to the use of sensors in lieu of questionnaires. However, the main problem is that not many clinical endpoints have been defined yet. The approach pursued in this paper is that of unsu-

pervised tasks, which combine the best characteristics of both approaches. The data are collected at the correct time and with high frequency and can be interpreted without the need to define new endpoints, since the data representation may change (e.g., from human driven to sensor based) but the semantics are kept unchanged.

The advantages of sensor-based monitoring during 5×STS tests have already been studied in the context of fall risk estimation [32]. However, such a study was done only under supervised settings. In contrast, we have shown that patients can perform the test unsupervised consistently, despite using a conservative consistency metric (it is likely that our consistency measure degrades if there is a disease-related change in the participant's status). We conclude that the execution of 5×STS tests could be monitored without supervision.

The adherence to the prescribed tests obtained in our study (56%) is comparable to that obtained in a related study [33] (61% adhered to the test sessions), where participants with Parkinson's disease were asked to perform prescribed tests on a daily basis while holding or wearing a smartphone. We observed that some participants performed the last test near the end of the study, yet some tests were missing, indicating that they possibly forgot to perform them. This was possibly caused by the test schedule, which was set to every Monday, Wednesday, and Friday, to avoid overburdening the participants. However, adherence could have been higher with a daily schedule, which may have been easier to remember, or with stronger reminders on the smartphones.

We evaluated the information captured by 5×STS test performance by analysing the correlation between performance impairment and morning pain and stiffness. Moreover, we built a mixed-effects model that predicted the severity of morning pain and stiffness via 5×STS duration, disease type, and gender. We observed that different participants reported pain severity in a consistent way, but the pain reported by different participants differed under similar circumstances because of individual pain thresholds and physical differences. By developing individual pain models, we could accurately predict the PRO, providing a potentially useful objective measure for evaluating pain and stiffness.

The study discussed in this paper was run at two sites in Dublin. However, a similar study can be run at more sites spread across different places. Indeed, the low number of required visits makes the study scalable. Moreover, we observed that the participants could follow the instructions after a short training, which lasted about 10 min. When running such a study in different locations around the world, we expect gender to lose its correlation with reported pain, as this may be culture related. Instead, the use of an explicit "culture" covariate may be considered.

In the future, unsupervised instrumented tests could be used not only to measure the performance of widely accepted performance assessments, but also to develop new, more accurate endpoints, which are tailored to the characteristics of body-worn sensors and exploit their full potential. This study has allowed us to collect data not only during 5×STS tests, but also during everyday activities such as morning routines, sleep, and commuting to work. Data collected during everyday activities have previously been used to estimate the risk of adverse events [34], or specific activities such as walking [35]. However, the relationship between sensor data collected during normal activities and qualitative health metrics, such as pain suffered during such activities, would benefit from a broader coverage. We plan to contribute to this aspect in the future.

### Acknowledgements

The authors thank Farid Khalfi, PhD, of Novartis, Dublin, Ireland, for providing medical writing support, in accordance with Good Publication Practice (GPP3) guidelines (<http://www.ismpp.org/gpp3>).

### Statement of Ethics

The study protocol has been approved by the research institute's committee on human research. The participants of this study gave their written informed consent prior to participation. Ethical approval was obtained for this study for all participants – for the patients through Trinity College Dublin/Tallaght Hospital and for the healthy volunteers through University College Dublin. The informed consent allows Novartis Pharma AG to share the data from this study with direct collaborators only. Collaboration proposals are welcome.

### Disclosure Statement

S.C.D., R.H.M., and B.C. have nothing to declare. C.G.M.P., V.P.I., F.C., E.O., O.S., and J.F.D. are employees of Novartis Pharma AG, Basel, Switzerland.

### Funding Sources

This study was funded by Novartis Pharma AG, Basel, Switzerland.

### Author Contributions

All authors were involved in the conception, drafting, and critical review of this article. All authors approved the final version to be published and agree to be accountable for all aspects of this work.

### Appendix

The participants were asked to respond to the questionnaire below.

1. "Were your joints stiff when you woke up today?"
2. "Please rate the activities in each category according to the following scale of difficulty: Stiffness severity upon waking up, first thing in the morning"
3. "Please rate the activities in each category according to the following scale of difficulty: Stiffness severity after sitting, lying down or resting during the day"
4. "How long did this stiffness last?"
5. "Can you rate the pain you were experiencing at the moment you woke up this morning on a scale of 0–10 where 0 represents no pain and 10 represents pain as bad as you can imagine?"
6. "Can you rate the pain you have experienced overnight on a scale of 0–10 where 0 represents no pain and 10 represents pain as bad as you can imagine?"
7. "Can you rate the pain you have experienced since you got up this morning on a scale of 0–10 where 0 represents no pain and 10 represents pain as bad as you can imagine?"
8. "Can you rate the OVERALL pain you have experienced IN THE LAST 24 HOURS on a scale of 0–10 where 0 represents no pain and 10 represents pain as bad as you can imagine?"

The 5×STS tests are detected with a semi-automated procedure, which consists of running a detection algorithm on the accelerometer data, converted into spherical coordinates. Such an algorithm is run in the mornings (6:00 to 12:00) of the days where a test was expected. When automatic detection failed, we reverted to manual inspection. The detection algorithm is described by the pseudo-code below.

```

function detect5xSTS (Azimuth, Elevation, Acceleration)
  # Initialise 5xSTS Detection
  stsDetectionTimes = { }
  elevationThresholdCrosses = lastElevationTCrossing = 0
  timeBelowTCondition = timeAboveTCondition = elevationTrigger = False
  tCurrent = 0

  for (tCurrent in 1,...,length(Elevation))
    # Get the elevation threshold with the data within the last 30 samples, i.e. 1 second
    T = 0.5 * max(Elevation(tCurrent-29 : tCurrent)) +
      0.5 * min(Elevation(tCurrent-29 : tCurrent))

    tCurrent = tCurrent+1
    elevationTimeBelowT = count(elevationMetric (lastElevationTCrossing : tCurrent) < T)

    if not timeBelowTCondition and elevationTimeBelowT > 10 and elevationTimeBelowT < 100
      timeBelowTCondition = True
      if timeAboveTCondition
        elevationTrigger = True
        timeAboveTCondition = False
      endif
    endif

    if elevationTimeBelowT >= 100
      # Maximum time below the threshold overcome, reset the condition.
      timeBelowTCondition = False
    endif

    elevationTimeAboveT = count(elevationMetric (lastElevationTCrossing : tCurrent) > T)

    if not timeAboveTCondition and elevationTimeAboveT > 10 and elevationTimeAboveT < 100
      timeAboveTCondition = True
      if timeBelowTCondition
        elevationTrigger = True
        timeBelowTCondition = False
      endif
    endif

    if elevationTimeAboveT >= 100
      # Maximum time above the threshold overcome, reset the condition.
      timeAboveTCondition = False
    endif

    if (elevationTrigger)
      lastElevationTCrossing = tCurrent
      elevationThresholdCrosses = elevationThresholdCrosses + 1
      elevationTrigger = False
    endif
    #Check if acceleration matches the pattern
    (accelerationPatternMatched,startPatternMatch,endPatternMatch) =
      accelerationPatternMatch(acceleration)
    if (Azimuth in [1.5,3] or Azimuth in [0,1.5]) and
      (elevationThresholdCrosses > 7) and accelerationPatternMatched
      stsDetectionTimes = stsDetectionTimes + (startPatternMatch,endPatternMatch)
      elevationThresholdCrosses = 0
    endif
  endfor

  return stsDetectionTimes
endfunction

```

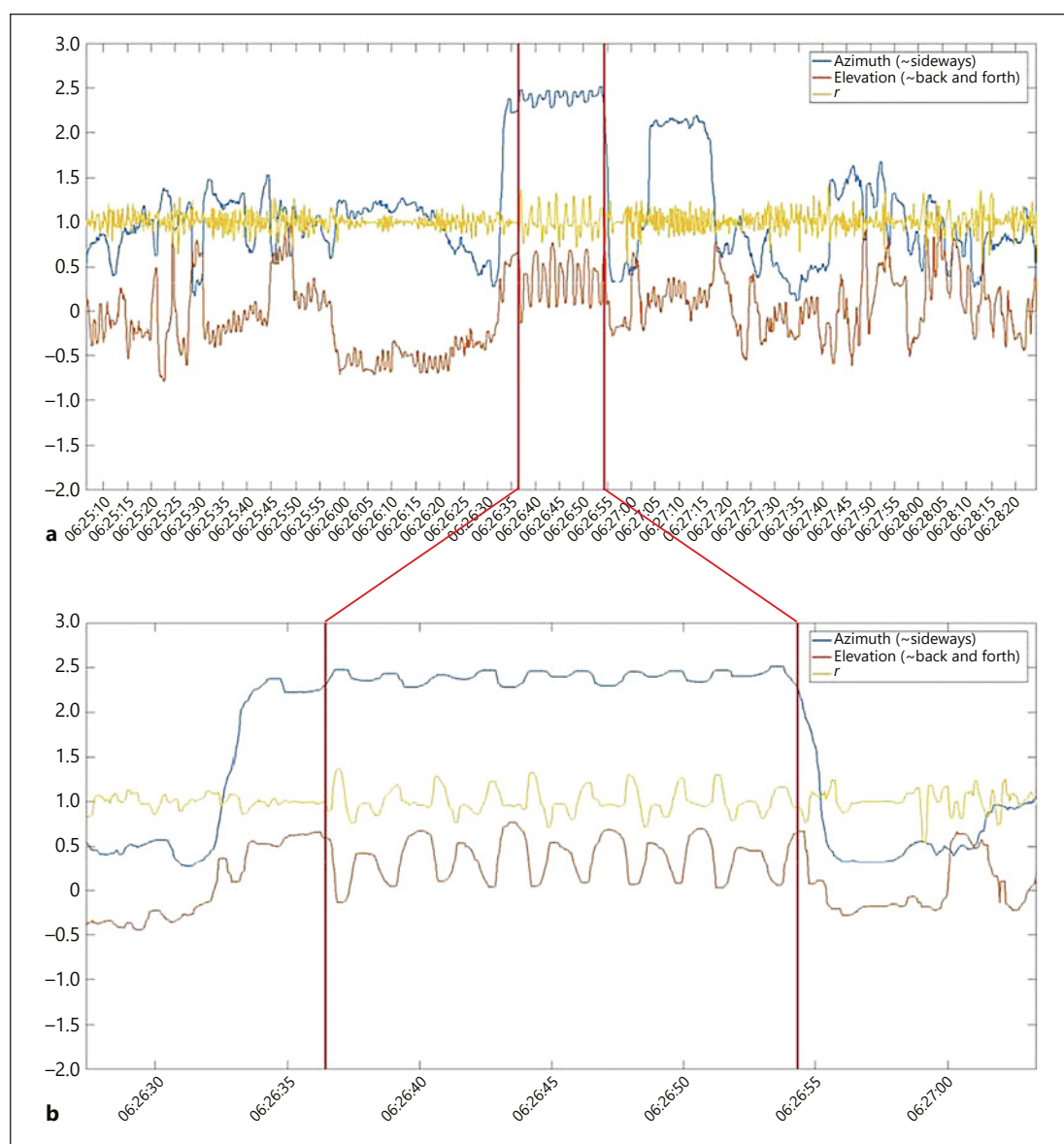


where accelerationPatternMatch is a function that transforms Acceleration into low/medium/high accelerations, which are denoted by a/b/c respectively, and then detects when the regular expression below is matched:

'b{1,5}c{1,60}+[b]+[a]+[b]\*[a]+[b]+[c]+[b]+[c]+[b]+[a]+[b]\*[a]+[b]+[c]+[b]+[c]+[b]+[a]+[b]\*[a]+[b]+[c]+[b]+[c]+[b]+[a]+[b]\*[a]+[b]+[c]+[b]+[c]+[b]+[a]+[b]\*[a]+[b]+[c]\*b{1,5}'

The transformation of Acceleration into the a/b/c levels is done thanks to two thresholds, which were calibrated using data from 6 participants, who all belonged to the development set.

With the algorithm presented above, 80% of the 5×STS tests that were performed were detected. Moreover, we validated the accuracy of detecting the test start and end by visual inspection of every test. An example of 5×STS test data in spherical coordinates is given in Figure A1.



**Fig. A1.** **a** Example of accelerometer data in spherical coordinates during a 3-min observation period. The red vertical lines identify an unsupervised Five Times Sit to Stand exercise. **b** Focus on the Five Times Sit to Stand test data.

The remaining 20% of the 5×STS tests were detected through manual inspection. When also manual inspection failed, we marked the test as not performed. In total, 56% of the prescribed tests were detected, either automatically or manually.

The algorithm described above did not generate false positives, i.e., it never happened that the algorithm found a 5×STS test where there was none. In contrast, we had some false negatives, for different possible reasons such as the following: not all 5×STS tests were performed, some 5×STS test attempts were stopped and repeated, etc.

We report the adherence to 5×STS tests and the sensor wear time for each patient in Table A1. Moreover, we report such results aggregated by condition in Table A2.

**Table A1.** Adherence in sensor wear time and 5×STS tests performed for each participant

Participant No.	Sensor wear, %	5×STS tests performed, <i>n</i>	Condition
1	70.8	10 of 15	RA
2	98.49	8 of 12	RA
3	43.35	3 of 12	RA
4	83.42	9 of 13	RA
5	94.87	15 of 15	RA
6	97.24	3 of 12	RA
7	97.63	13 of 13	RA
8	98.69	14 of 14	RA
9	97.71	9 of 12	RA
10	95.75	0 of 12	RA
11	97.71	7 of 12	RA
12	97.6	11 of 12	RA
13	98.99	4 of 14	RA
14	97.64	11 of 12	RA
15	84.19	8 of 12	RA
16	97.75	0 of 12	RA
17	97.9	12 of 12	RA
18	98.39	6 of 12	RA
19	95.62	8 of 15	RA
20	99.76	8 of 12	RA
21	95.49	5 of 13	OA
22	97.84	10 of 12	OA
23	94.06	10 of 14	OA
24	98.28	3 of 12	OA
25	96.23	11 of 12	OA
26	59.99	2 of 12	OA
27	97.39	2 of 14	OA
28	72.33	1 of 9	OA
29	98.62	0 of 14	OA
30	69.74	3 of 12	OA
31	94.18	11 of 12	HV
32	97.36	10 of 13	HV
33	97.66	8 of 12	HV
34	88.39	3 of 12	HV
35	97.79	6 of 15	HV
36	97.22	9 of 12	HV
37	96.92	9 of 12	HV
38	98.45	0 of 14	HV
39	95.3	10 of 12	HV
40	97.22	2 of 14	HV
41	99.21	12 of 12	HV
42	97.72	7 of 14	HV
43	80.55	10 of 13	HV
44	77.32	8 of 12	HV
45	78.95	6 of 11	HV

RA, rheumatoid arthritis; OA, osteoarthritis; HV, healthy volunteer.

**Table A2.** Adherence in sensor wear time and 5×STS tests performed aggregated by clinical condition

Disease	Sensor wear, %	5×STS tests performed, %
RA	92.18	61.80
OA	88.00	37.60
HV	92.95	59.80
Total	91.50	55.76

RA, rheumatoid arthritis; OA, osteoarthritis; HV, healthy volunteer.

## References

- 1 Branco JC, Rodrigues AM, Gouveia N, et al: Prevalence of rheumatic and musculoskeletal diseases and their impact on health-related quality of life, physical function and mental health in Portugal: results from EpiReumaPt – a national health survey. *RMD Open* 2016;2:e000166.
- 2 Murphy LB, Cisternas MG, Pasta DJ, et al: Medical expenditures and earnings losses among US adults with arthritis in 2013. *Arthritis Care Res (Hoboken)* 2018;70:869–876.
- 3 Anyfanti P, Triantafyllou A, Panagopoulos P, et al: Predictors of impaired quality of life in patients with rheumatic diseases. *Clin Rheumatol* 2016;35:1705–1711.
- 4 da Silva JA, Phillips S, Buttgerit F: Impact of impaired morning function on the lives and well-being of patients with rheumatoid arthritis. *Scand J Rheumatol Suppl* 2011;125:6–11.
- 5 Kłak A, Raciborski F, Samel-Kowalik P: Social implications of rheumatic diseases. *Reumatologia* 2016;54:73–78.
- 6 Halls S, Dures E, Kirwan J, et al: Stiffness is more than just duration and severity: a qualitative exploration in people with rheumatoid arthritis. *Rheumatology (Oxford)* 2015;54:615–622.
- 7 Deodhar A, Braun J, Inman RD, et al: Golimumab reduces sleep disturbance in patients with active ankylosing spondylitis: results from a randomized, placebo-controlled trial. *Arthritis Care Res (Hoboken)* 2010;62:1266–1271.
- 8 Minnock P, Veale DJ, Bresnihan B, et al: Factors that influence fatigue status in patients with severe rheumatoid arthritis (RA) and good disease outcome following 6 months of TNF inhibitor therapy: a comparative analysis. *Clin Rheumatol* 2015;34:1857–1865.
- 9 Barsky AJ, Orav EJ, Ahern DK, et al: Somatic style and symptom reporting in rheumatoid arthritis. *Psychosomatics* 1999;40:396–403.
- 10 Martin JL, Hakim AD: Wrist actigraphy. *Chest* 2011;139:1514–1527.
- 11 Van Lummel RC, Walgaard S, Maier AB, et al: The Instrumented Sit-to-Stand Test (iSTS) has greater clinical relevance than the manually recorded Sit-to-Stand Test in older adults. *PLoS One* 2016;11:e0157968.
- 12 Papi E, Osei-Kuffour D, Chen YM, McGregor AH: Use of wearable technology for performance assessment: a validation study. *Med Eng Phys* 2015;37:698–704.
- 13 Ganea R, Paraschiv-Ionescu A, Aminian K: Detection and classification of postural transitions in real-world conditions. *IEEE Trans Neural Syst Rehabil Eng* 2012;20:688–696.
- 14 Doulah A, Shen X, Sazonov E: Early detection of the initiation of sit-to-stand posture transitions using orthosis-mounted sensors. *Sensors (Basel)* 2017;17:E2712.
- 15 Pavord ID, Mathieson N, Scowcroft A, et al: The impact of poor asthma control among asthma patients treated with inhaled corticosteroids plus long-acting  $\beta_2$ -agonists in the United Kingdom: a cross-sectional analysis. *NPJ Prim Care Respir Med* 2017;27:17.
- 16 Pugliese L, Woodriff M, Crowley O, et al: 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:e535.
- 17 Jean-Louis G, Kripke DF, Cole RJ, et al: Sleep detection with an accelerometer actigraph: comparisons with polysomnography. *Physiol Behav* 2001;72:21–28.
- 18 Tudor-Locke C, Barreira TV, Schuna JM Jr, et al: Fully automated waist-worn accelerometer algorithm for detecting children’s sleep-period time separate from 24-h physical activity or sedentary behaviors. *Appl Physiol Nutr Metab* 2014;39:53–57.
- 19 Mathworks: Transform Cartesian coordinates to spherical. <https://ch.mathworks.com/help/matlab/ref/cart2sph.html> (accessed 12 July 2018).
- 20 van Hees VT, Fang Z, Langford J, et al: Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents. *J Appl Physiol* (1985) 2014;117:738–744.

- 21 Berndt DJ, Clifford J: Using dynamic time warping to find patterns in time series. AAAI Technical Report WS-94-03 1994;10:359–370.
- 22 Paul SS, Canning CG: Five-repetition sit-to-stand. *J Physiother* 2014;60:168.
- 23 Ganea R, Paraschiv-Ionescu A, Büla C, et al: Multi-parametric evaluation of sit-to-stand and stand-to-sit transitions in elderly people. *Med Eng Phys* 2011;33:1086–1093.
- 24 Van Lummel RC, Ainsworth E, Lindemann U, et al: Automated approach for quantifying the repeated sit-to-stand using one body fixed sensor in young and older adults. *Gait Posture* 2013;38:153–156.
- 25 Lubke GH, Campbell I, McArtor D, et al: Assessing model selection uncertainty using a bootstrap approach: an update. *Struct Equ Modeling* 2017;24:230–245.
- 26 Waehrens EE, Amris K, Fisher AG: Performance-based assessment of activities of daily living (ADL) ability among women with chronic widespread pain. *Pain* 2010;150:535–541.
- 27 Amris K, Waehrens EE, Christensen R, et al: Interdisciplinary rehabilitation of patients with chronic widespread pain: primary endpoint of the randomized, nonblinded, parallel-group IMPROVe trial. *Pain* 2014;155:1356–1364.
- 28 Dobson F, Hinman RS, Roos EM, et al: OARSI recommended performance-based tests to assess physical function in people diagnosed with hip or knee osteoarthritis. *Osteoarthritis Cartilage* 2013;21:1042–1052.
- 29 Lucey P, Cohn JF, Prkachin KM, et al: Painful monitoring: automatic pain monitoring using the UNBC-McMaster shoulder pain expression archive database. *Image Vis Comput* 2012;30:197–205.
- 30 Turk DC, Dworkin RH, Burke LB, et al: Developing patient-reported outcome measures for pain clinical trials: IMMPACT recommendations. *Pain* 2006;125:208–215.
- 31 Sit AJ: Continuous monitoring of intraocular pressure: rationale and progress toward a clinical device. *J Glaucoma* 2009;18:272–279.
- 32 Doheny EP, Fan CW, Foran T, et al: An instrumented sit-to-stand test used to examine differences between older fallers and non-fallers. *Conf Proc IEEE Eng Med Biol Soc* 2011;2011:3063–3066.
- 33 Lipsmeier F, Taylor KI, Kilchenmann T, et al: Evaluation of smartphone-based testing to generate exploratory outcome measures in a phase 1 Parkinson's disease clinical trial. *Mov Disord* 2018;33:1287–1297.
- 34 Stack E, Agarwal V, King R, et al: Identifying balance impairments in people with Parkinson's disease using video and wearable sensors. *Gait Posture* 2018;62:321–326.
- 35 Hickey A, Del Din S, Rochester L, Godfrey A: Detecting free-living steps and walking bouts: validating an algorithm for macro gait analysis. *Physiol Meas* 2017;38:N1–N15.