Northern Periphery and Arctic Programme
2014–2020

EUROPEAN UNION
Investing in your future
European Regional Development Fund

WEARABLE SENSOR TECHNOLOGIES
SENDoc

Alamäki A., Barton J., Condell J., Crowe C., Gillespie J., Kelly, D., Muñoz Esquivel K., Nevala E., Nordström A., Sica, M., Tedesco S., Åkerlund Larsson M.

Deliverable D.T2.3.1
Final Technology Report on the Installation and Testing of wearable sensor systems for the support of independent living for aging populations

Ulster University

region västerbotten

Tyndall National Institute
Institiúid Náisiúnta

Karelia
UNIVERSITY OF APPLIED SCIENCES

Contents

1. General Introduction 1
2. The Transnational Trial 2
2.1 Introduction 2
3. Progress to Date 3
3.1 Ireland 3
3.2 Finland 5
3.3 Northern Ireland 6
3.4 Sweden 9
4. Analysis 10
4.1 Pre-trial questionnaire results 10
4.2 System Usability Scale 11
4.2.1 Analysing the SUS score by cohort region 12
4.2.2 Analysing the SUS score by gender 13
4.2.3 Analysing the SUS score by number of wearables 13
4.2.4 Analysing the SUS score by age 13
4.3 Bespoke Usability Questionnaire 14
4.3.1 A predictive model for continued device usage 16
5. General conclusions 18
References 20

1. General Introduction

Nowadays, population aging is growing due to increased life expectancy and decreasing childbirth. The percentage of people aged over 65 years in the year 2016, was 20.5 % in Finland, 19.8 % in Sweden, 17.9 % in UK, 13.2 % in Ireland and 19.2 % in EU [1]. This issue, together with the limitations that rural residents have in order to access necessary and appropriate health services, causes severe socio-economic challenges to countries. For instance, about 1.2 million of people aged 65 years or older lived in the UK in 2016 and most of them in rural areas [2]. This number is projected to increase by an average of 20% by 2024. Consequently, the development of a remote health system for elderly subjects - especially in rural areas - is needed. Future development in remote health care technologies have the capacity not only to speed up the role of home health care within the overall health care system, but also to encourage foster community-based autonomy for individuals.

In order to enhance a home health care system, studies on “off the shelf” devices for monitoring and rehabilitation are growing. However, these investigations are confined in most of the cases to a healthy young or adult population [3,4,5]. The applications of commercial wearable technologies on elderly adults is fundamental, taking into account the numerous applications and their potential benefits in healthcare [6]. Even though elderly people accept commercial wearables and consider them useful [7,8], specific devices for older adults are still insufficient [6].
In Deliverable D.T2.1.1, we presented a first evaluation of available commercial wearable systems and fitness trackers in real rehabilitation situations and processes. In particular, the aim of the research is to determine whether there is some kind of “added value” for elderly patients, care givers and physiotherapists. A number of those devices were selected as part of each country specific trial to be tested in real life clinics and community volunteers.

In addition to the Regional Trials, the SENDoc PEC decided to start an additional Transnational trial involving wearable activity trackers, mobile phones and usability questionnaires. The symptoms of neurological diseases which affect elderly people, such as Parkinson’s, and other common chronic conditions related to the ageing process are often only evaluated in a clinical setting, once or twice a year. Physical activity trackers enable the continuous collection of physiological data which could facilitate the remote monitoring of disease progression. It has already been demonstrated that some “off the shelf” sensors can give reliable measurements for certain physiological parameters and can “add value” to end users (elderly patients, physiotherapists, and occupational therapists). In particular, fitness trackers have been found to estimate steps, energy expenditure, and some sleep parameters (e.g. Total sleep time) with a good level of accuracy in a healthy older adult population [9]. This study seeks to increase our understanding of how older adults feel about using these wireless sensor monitoring devices to discover information about their health in daily life. The feedback obtained through the utilization of these technologies could benefit patients by encouraging practices that lead to a healthier life as evidence shows that patients who have a more highly engaged role in managing their own health have better treatment outcomes [10].

2. The Transnational Trial

2.1 Introduction

Improvements made in wireless sensing technology in recent years has enabled the use of wearable devices in commercial applications for personal health monitoring. These technologies hold great potential to enhance home health care monitoring and rehabilitation systems for which there is a growing demand due to the ageing population trend which can be seen in developed countries. This issue, in conjunction with the current limitations faced by residents of rural areas in accessing the necessary and appropriate health services, highlights a need for the development of remote healthcare systems for elderly patients.

The Transnational Trial aims to test one such system. Following the decision of the PEC, it was determined that Tyndall-UCC would lead this trial as part of WP3. The Ulster team has taken a supportive and collaborative role in conjunction with the Tyndall-UCC team. Both teams worked together to define a protocol and technology to employ which were tentatively defined at our Technical meeting in Cork on March 2018. It was decided to deliver technology kits comprised of a mobile phone, smart watch, and accelerometer data logging device. It was agreed that the study would focus mainly on the usability of the devices from an elderly person's perspective. Additionally, it was decided that the study would be an investigative one with the aim to research whether or not the level of activity of participants was related to their level of actual frailty. All of the teams involved engaged in extensive conversations to finalise the experimental protocol at the SENDoC 4th Partner's Meeting in Derry, Ulster in May 2019. Data collection trials subsequently began with each participant taking a kit home with them and wearing the set of devices for 7 consecutive days. Participants also underwent a short physical and mental assessment and filled out standardized usability questionnaires. A population of individuals over the age of 65 was considered for the data capture.

In summary, during the transnational demonstration, participants tested a system consisting of an off-the-shelf fitness tracker in conjunction with its associated mobile phone application for a period of one week. They also wore an accelerometer data logger during this time. The addition of this research grade accelerometer, while not necessary for the usability aspects would allow comparison with gold standard equipment. At the end of the trial, the usability of the system was reported on by the cohort of elderly subjects. All of the devices and clinical assessments used in the trial including experimental protocols were described in detail in D.T2.2.1.

3. Progress to Date

3.1 Ireland

The Ireland team collaborated in conjunction with the other teams to define an experimental procedure that would be carried out by the researcher over the course of the study. Tyndall-UCC then purchased 20 of the trial kits and distributed 5 kits to each of the other regions keeping 5 kits for use in Cork. An instruction manual was also created for participants in order to guide them on how to use the devices during the trial. After the finalization of the trial protocol, this participant training leaflet detailing the use of the devices was updated to reflect the changes incurred during the latest Mi Fit app update. A pilot trial was undertaken wherein we received positive feedback in relation to the experimental set-up but experienced problems when retrieving the trial data from the off-the-shelf fitness tracker. Therefore, the method for automatically exporting data from the device was specified in an instruction manual for researchers and manually logging the data from the Mi Band 3 was proposed as an alternative if this method should fail.

In September 2018, an ethics application form for a new study was created and submitted to the Clinical Research Ethics Committee of the Cork Teaching Hospitals (CREC) by Dr Suzanne Timmons. This was approved in Jan 2019. However, following the SENDoc Partners' Meeting in Derry/Londonderry in May, an amendment to the ethics application had to be made in order to include the Geriatric Depression Scale as an assessment in the protocol, which has delayed the process and the recruitment of participants for the actual trial. There have been ten participants including the pilot. 50% have been male and 50% female with an age range of 68-80 years old. The participants wore the devices for a period of one week while completing the trial protocol. The participant trials were suspended in March 2020 in advance of government restrictions and did not continue due to feasibility issues from a post COVID-19 perspective.

Figure 1: Transnational Trial equipment and screenshot

3.2 Finland

We have participated in the development of the research protocol and included measurements. Karelia has taken care of all the translations from English to Finnish, for example, the information sheet and consent form. Karelia UAS has recruited participants from Karelia´s learning environments Fysiotikka and Voimala and also from elderly persons' unions (Joensuun Eläkeläiset Ry, Joensuun eläkkeensaajat Ry, Eläkeliitto Joensuu, Joensuun seniorit) in the Joensuu region. So far, 23 participants have completed the evaluations and measurements of activity tracker usability and accuracy research with elderly patients. Data from the bands and results from the questionnaires have been saved to the Dropbox. Half of the participants have used both the Xiaomi Mi Band and the Axivity AX3 and half of them have only worn the Xiaomi Mi Band during the trial.

Figure 2: Finnish participants with colour coded bands

Some early phase observations that might impact on results have been noted. For example, the protocol should have included a section regarding instructing participants on how to add exercises to the Mi Fit App. Additionally, the use of two devices (wearing bands on both wrists) has had an impact on usability results (see usability questionnaire - other comments). Having the Axivity on the dominant and Xiaomi Mi Band on the non-dominant wrist has also impacted accuracy results because they are worn on different wrists. Some of the elderly participants have found the Axivity uncomfortable to wear. Quite a lot of the elderly subjects have said that the “locking system” of the Xiaomi Mi Band is too tight and difficult to use.

3.3 Northern Ireland

The Ulster team supported and collaborated with Tyndall-UCC from the start of the project on defining the trial protocol and the technology to employ. Additionally, Ulster assisted in writing all the required supportive material for the trial including the researchers guide and the user guides for the selected devices. These were distributed to each partner before being given to the participants in the trial to aid in device training / for functionality referral during the trial.
Before the trial could begin at Ulster, ethical approval had to be gained. However, due to the potential vulnerable nature of the participants and the ethical procedures in place at Ulster University for vulnerable participants, the process to apply for ethics was highly complex and time consuming.

Figure 3 illustrates the process followed to define the protocol and the technology employed on the transnational demonstration by all SENDoc teams. In parallel, the timeline for the ethics application at Ulster can be seen.

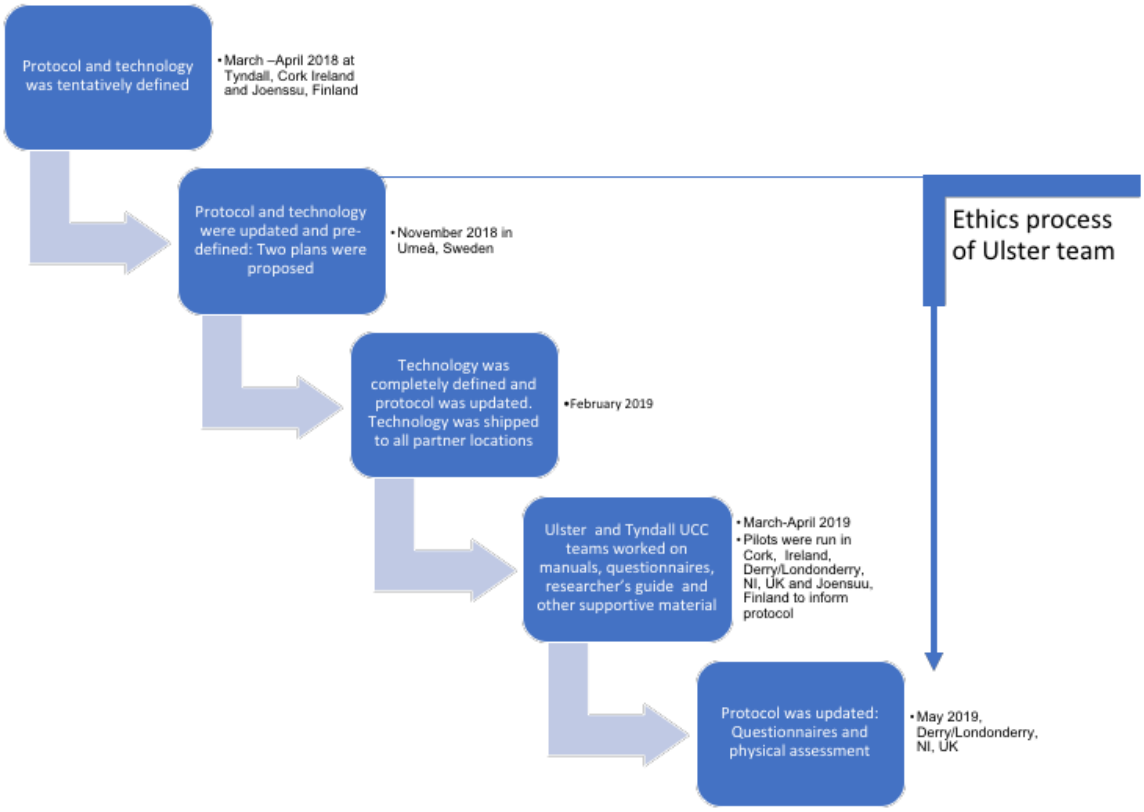


Figure 3. Timeline for Transnational demonstration definition and the ethics process of Ulster team

The original discussed plan in March 2018 was to engage with patients at clinics in Northern Ireland from the Western Health and Social Care Trust. In September 2018, we started to draft the ethics application and in December 2018 a completed version of the ethics was put online in the Integrated Research Application System (IRAS) system (The Health Research Authority, 2019). It was initially planned that patients would be recruited through the Western Health and Social Care Trust (WHSCT). However after numerous meetings and discussions with different staff within WHSCT it became clear that in order to get ethical approval from the trust, we would have to hire a research nurse and contract WHSCT staff to help in getting the ethics application through the different IRAS processes. The quote obtained was approx. £40,000, which was considered to be an extremely high cost to the project. As a result, we decided to recruit healthy elderly volunteers, and not patients. Therefore, instead of applying for ethics approval though the Health Research Authority in Northern Ireland, the Ulster team moved to obtain ethical approval from the Ulster University ethics committee. Figure 4 illustrates the complete ethics process and timeline followed by the Ulster team.

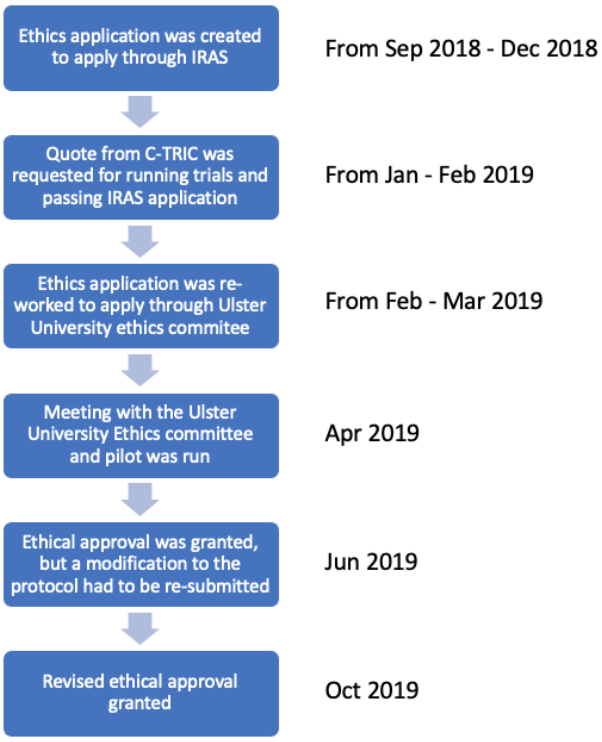


Figure 4. Ethics process and timeline followed by Ulster team

The Ulster team then worked on creating and submitting the ethics application for the Ulster University ethics committee. This was submitted in March 2019 first to internal reviewers, then to the Faculty ethics committee and finally to the Ulster University Ethics committee. A face to face meeting was attended at the end of April 2019. The ethics application was granted at the beginning of June 2019, but a modification had to be re-submitted owed to changes in the transnational demonstration protocol at the SENDoc physical meeting in May 2019. In late October 2019, University ethical approval for a study involving healthy elderly volunteers was approved.

After designing posters and leaflets to advertise the trial, the Ulster team distributed the materials to local community groups which held activities and classes for older adults. These groups included U3A, Eglinton Community Centre, The Old Library Trust, Creggan Community Centre and the DIAL centre. Additionally, word of mouth was used to advertise the trial. Significant interest in the study started building in January 2020 as volunteers were keen to test a device which provided extra exercise incentive to them after the Christmas period. Since Karelia had completed their study in June 2019, they were able to send their equipment to Ulster so that up to 10 trials could take place in parallel.

By the 11th March 2020, 14 participants had taken part in the study and a further 8 had been scheduled to be completed by the 1st April, which would have brought the Ulster total to 22. However, due to government restrictions surrounding the COVID-19 pandemic the trial had to be temporarily halted. It was hoped that it could be restarted again before the end of the project but due to ongoing government restrictions, various rules surrounding social distancing, and university legislation this was not possible. The Ulster team therefore finished with 14 complete trials, of which 3 were male and 11 were female. The ages ranged from 65-95 with a mean of 71.6 and a standard deviation of 7.8. 50% of the participants wore both the MiBand and Axivity and 50%

wore only the MiBand activity tracker. Figure 5 shows a selection of the healthy older adults who took part in the study at Ulster.

Figure 5: A sample of the participants who took part in the transnational trial from Ulster University

After the trial finished, elders who agreed provided a video testimony which can be found here:
https://www.youtube.com/watch?v=i6qlLxSqkxE

The general consensus among the participants in Northern Ireland was that the elders enjoyed using the technology and found that having an activity tracker encouraged exercise – in fact one gentlemen compared having an activity tracker to having a personal trainer. The average score on the system usability scale within the Northern Irish cohort was 67.85 with a standard deviation of 23.1. Some participants found the donning and doffing of the equipment difficult due to the fastening clip of the Mi Band being particularly awkward to use.

3.4 Sweden

The Swedish team participated in the development of the test protocol and included measurements. After the finalization of the protocol, work began on translating the procedure from English to Swedish. The translations have been completed for all necessary documents and forms. Afterwards, the protocol was given a trial run by testing it on health care staff as well as elderly subjects in order to fully comprehend all aspects of the procedure and to ensure that nothing vital had been lost in translation. The trial was successfully carried out with 20 healthy elderly volunteers.

4. Analysis

In total, 65 elders participated in the study from the four locations. The mean age of the participants was 70.52 years old with a standard deviation (SD) of 5.65. The mean height of the population was 169.43 cm (SD = 9.05) and the mean weight was 73.45 kg (SD = 13.09). The cohort was comprised of 37 females (56.9%) and 28 males (43.1%). 59 (90.8%) of the participants were right-handed and 6 (9.2%) were left-handed.

One of the key aims of the analysis was to evaluate the relationship between recorded features and key usability measures. More specifically, the aim was to investigate possible predictors (or markers) of good usability and continued device usage among older adults.

Based on the extracted features, two types of models were developed: 1) A multinomial logistic regression model was implemented to assess the statistical power of specific features and 2) A Random Forest machine learning model was implemented to evaluate the potential of predicting continued device usage. The multinomial logistic regression model is interpretable, allowing evaluations to be conducted on the contributions of different features and categories. The Random Forest model is not interpretable, but is more likely to perform better in modelling relationships between features and usability.

4.1 Pre-trial questionnaire results

Before the trial, participants completed four questionnaires: 36-Item Short Form Health Survey (SF-36), Mini-Mental State Examination (MMSE), Geriatric Depression Scale (GDS) and Mobile Device Proficiency Questionnaire (MDPQ). These were explored as part of our initial analysis to help form a baseline understanding of the population in our cohort. Summary statistics for each of these are presented in Table 1 with the general health variable selected to represent the SF-36 questionnaire and the MDPQ overall variable selected to represent the MDPQ questionnaire.

	Mean	**SD**	**Variance**
SF-36 General Health	72.54	18.960	359.471
MMSE	28.49	1.552	2.410
GDS	1.43	2.114	4.468
MDPQ Overall	3.53	1.256	1.577

Table 1: Pre-trial questionnaire results

It is important to note that our cohort was comprised of mostly healthy volunteers as defined by SF-36 results (72.54 out of 100, sd =18.96). Results showed 6 participants scored below 50 on the SF-36 general health component inferring that a small number of elders in the study perceived that they were struggling with health issues.

We can observe that the average MMSE value is 28.49 (SD =1.552). A score below 25 on the MMSE indicates some degree of dementia. Only one participant scored below 25 on the MMSE with a score of 24. GDS scoring was on average 1.43, where a score above 4 indicates some degree of depression. Only 6 participants reported scores above 6 in the GDS. The MDPQ Overall score was calculated by averaging the score from each of the 8 MDPQ categories. The overall value averaged 3.53, which is between 3 “not very easily” and 4 “somewhat easily”, indicating that our participants are between states when it comes to overall mobile phone device proficiency.

4.2 System Usability Scale

After wearing the device for 1 week, participants were asked to answer the System Usability Scale (SUS) questionnaire, a reliable and standardised tool consisting of 10 questions that can be used to evaluate the usability of a piece of software, hardware, or wearable device [11]. SUS questions use a 5 point likert scale and the questionnaire is scored out of 100 points. In our study, participants were asked to answer the SUS questionnaire to evaluate the Xiaomi Mi Band 3 activity tracker after an average device usage of 7.12 days (SD 1.526 days). Only the usability of the wearable device was to be considered by the participants. The final results showed the average the SUS score (n=65) was 67.15 with a standard deviation (SD) of 18.269. To contextualise this result with other activity trackers currently on the market, we compared our findings with the analysis conducted by Steinert, Haesner, and Steinhagen-Thiessen [12]. In this work, 20 elders were asked to evaluate 5 different activity trackers over a 2-hour period. The results from their study were presented in Table 2. Although a different cohort, review time, and a number of devices for context have been used, it is interesting to note that the Mi Band 3 in our study, evaluated across 4 independent locations, is ranked similar to the Fit-Bit Flex and Nike FuelBand. SUS scores of the Mi Band 3 in this work were significantly higher than the other 3 sensors assessed by Steinert et al. [12].

Activity Tracker	**Mean**
Fit-Bit Flex	66.25
Garmin VivoFit	52.5
Jawbone UP24	60.8
Nike FuelBand	65
Sony SmartBand	37.5
Average	56.*38*

Table 2: Summary statistics from market competitors

Summary Statistic		**n**	**mean**	**SD**	**test statistic (df)**	**p-value**
Region	Karelia	23	68.3	11.95	0.091 (2)	0.913
	Ireland	22	65.9	19.34		
	Umea	20	67.3	23.28		
Gender	Male	28	65.98	18.25	0.447 (63)	0.656
	Female	37	68.04	18.5		
Number of Wearables	MiBand	28	69.4	19.3	0.851 (63)	0.398
	MiBand + Axivity	37	65.5	17.5		
Age	Below 70	23	71.3	14.6	0.411 (2)	0.814
	70-74	32	67.6	19.3		
	Above 74	20	67.3	23.3		

Table 3: Summary statistics on participants

Further analysis was conducted to explore whether there were usability differences with respect to the region our participants came from, the gender of our participants, and the number of wearables worn in the trial. The results of these analyses are found in Table 3.

4.2.1 Analysing the SUS score by cohort region

The cohort of 65 persons was comprised of elderly volunteers from Northern Ireland (14), Republic of Ireland (8), Finland (23), and Sweden (20). SUS scores were analysed to investigate if geographical location affected the usability rating. The elders who participated in the trial from Northern Ireland and the Republic of Ireland were grouped into one cohort comprising of 22 participants for this analysis.

SUS percentiles were converted into 3 ordinal values representing not-acceptable (0≤ SUS <50), marginal (50≤ SUS <70) and acceptable (70≤ SUS ≤100) [13]. For Karelia, the largest percentage 44% (11 people) in the SUS score was observed in the categories corresponding to marginal usability scores of the Mi Band 3. For Ireland, 40.9% (9 people) were observed in the marginal and acceptable categories. For Umea, the largest percentage 45% (9 people) is in the acceptable category. In total, 28 people thought the device had a score larger than 70 (43.1%).

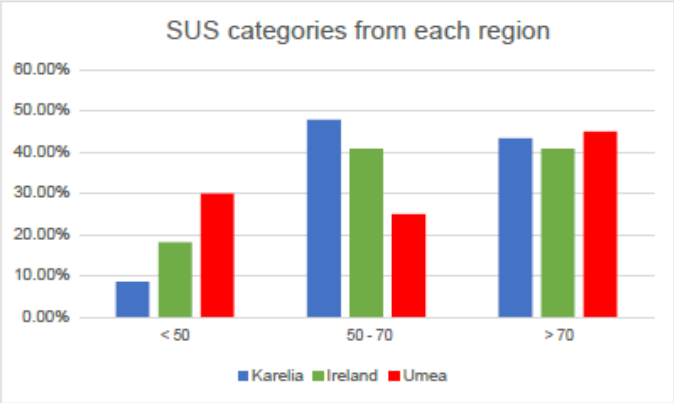
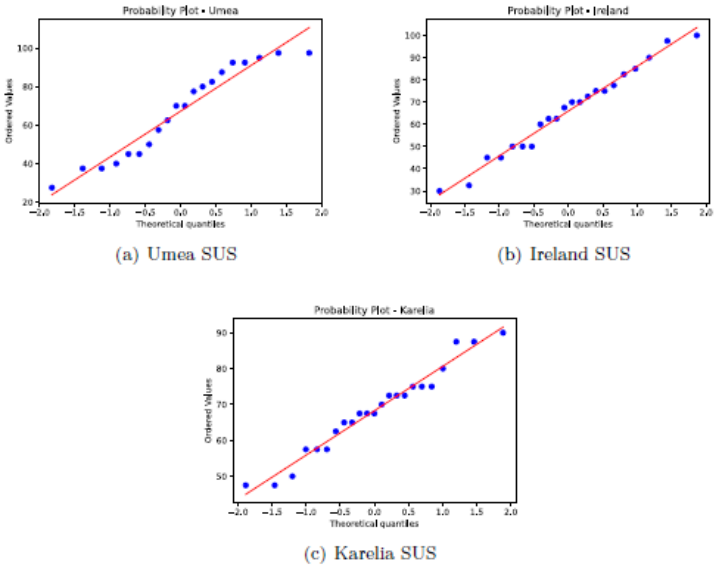


Figure 6. Histogram of SUS categories from each region

Observing Figure 6, the distributions are not identical therefore a one-way ANOVA test was applied to compare mean ranks. Results presented in Table 3, show the differences between the mean rank of the SUS score for each region are not statistically significant.

Group means were compared using a one-way ANOVA. A Levene's test was performed resulting in p=0.85, therefore the variances can be assumed as homogenous and equal variances are assumed. Observing the normal quantile-quantile plots for each region in Figure 7 the quantiles mainly lie on or close to the red line suggesting a normal distribution. From the results in Table 3, the means of the SUS score for each region are not statistically significant.



(a) Umea SUS (b) Ireland SUS

(c) Karelia SUS

Figure 7. Q-Q plots of the SUS score from each region

4.2.2 Analysing the SUS score by gender

An independent-samples t-test was conducted to compare the System Usability Score between the genders. The results (p=0.656) suggest that there was no difference in perceived system usability whether the participant was male or female.

4.2.3 Analysing the SUS score by number of wearables

Although instructed to only consider the Mi Band 3 while completing the SUS, it is possible that the participants' opinion has been subconsciously influenced while wearing a multi-sensor system for a week. This is an interesting research question to consider as many wearable sensors systems currently on the market rely on multi-sensor setups. An independent-samples t-test was therefore conducted to compare the SUS of participants wearing only the Mi Band compared with participant wearing the Mi Band and the Axivity ax3. The results from Table 3 suggest that there was no difference in perceived system usability whether the participant wore one activity tracker or both activity trackers (p=0.398).

4.2.4 Analysing the SUS score by age

The participants were divided into three categories of elders according to their age (below 70, between 70 to 74 and above 74). In order to understand the significance of these means, a Kruskal-Wallis H-test was conducted to compare the SUS between the three age categories. Levene's test p-value=0.00016 therefore the variances can be assumed as not homogeneous. The results from the statistical test suggest that the means of the SUS score for each age category are not statistically significant (p=0.814).

All of the SUS analysis shows that regardless of comparing region, gender, wearables, or age; there was no difference in perceived system usability.

4.3 Bespoke Usability Questionnaire

The SUS is a standardised and validated short 10 question survey to help validate the usability of a piece of hardware, software, or wearable device. However, in order to better understand the participants specific opinions of the wearable device usability, a bespoke usability questionnaire was designed entitled “Accuracy, feasibility and acceptability of wireless monitoring in older people and in people with Parkinson’s Disease”. The questionnaire first collected dichotomous data on the participants’ familiarity with wearable devices and if they liked the appearance. Then, a series of questions relating to usability, accuracy, and acceptability were asked using an ordinal 5 category scale, ranging from strongly agree to strongly disagree. The questionnaire ended with 4 general questions which gathered data on length of time worn and usage at night. The final question and most important question with respect to this study, asked the participants if they would continue to use the device after the trial had finished. We will perform various analyses on this final question in order to gain further insights into what factors influence continued device usage.

Doing so will help inform future wearable device design, thus ensuring there is the highest possible chance of uptake. This will, in turn, allow for the successful rollout of telehealth services in the future such as remote rehabilitation.

In total, 42 participants said they would like to continue using the wearable device and phone app, while 23 participants said they would not like to continue using the wearable device and phone app. Correlation between the continued device usage question and the SUS questionnaire was first assessed. The results are displayed in Table 4. The participants who indicated they would like to continue using the device consider its usability to be within the “acceptable” category (range above 70), while those who indicated they would not be interested in continuing to use the device rank its usability within the “marginally low acceptability” category (range above 50 and below 65). Results from an independent t-test show that the SUS scores for the two ’continued use’ groups were statistically significant.

Continue to use wearable device?	**n**	**Mean**	**SD**	**t-Value (df)**	**p-Value**
Yes, I would like to	42	71.8	17.08	-2.92(63)	0.005
No, I am not interested	23	58.7	17.64		

Table 4: Summary statistics of the SUS score by Question 21

In order to evaluate what effect previous activity tracker experience has on continued device usage, analysis was performed on Usability Question 2 (“Have you previously used a wrist-worn activity tracker before the project?”).

In total, 13 participants said they had previously used a wrist-worn activity tracker, while 52 participants said they had never used a wrist-worn activity tracker before the trial. The correlation between the previous experience question and the SUS questionnaire was calculated. The results are displayed in Table 5. An independent t-test was performed to check the statistical significance of these results. The results from this test show that there is no correlation between the users system usability scale score and whether they had previous experience of a wrist-worn activity tracker (p=0.277). This finding shows that previous experience has no effect on a user’s acceptance of a wearable device. Each participant in the study received a brief training lasting 10-15 minutes and a manual therefore this might have helped to enhance the user’s ability.

Previously worn an activity tracker?	n	Mean	SD	t-Value (df)	p-Value
Yes	13	72.12	15.61	-1.10(63)	0.277
No	52	65.91	18.81		

Table 5: Summary statistics of the SUS score by Question 2

Further analysis was performed on the continued device use question to evaluate how continued use was linked with other elements of usability. Correlations between Q21 and the other bespoke usability questions were therefore analysed. Kendall's tau b rank correlation coefficient was used to measure the strength of the ordinal association between the usability questions and the continued device usage question. The results of the analysis are presented in Table 7. Results revealed 5 usability questions which have a strong correlation with the continued device usage question. Question 10 and Question 17 were the top-ranking features, each with a p-value=0.003, highlighting that both comfort at night and becoming more active are key early indicators as to whether a user would continue using and wearing the device.

Correlation	Rank
±0.10	Very weak
±0.10 to 0.19	Weak
± 0.20 to 0.29	Moderate
±0.30	Strong

Table 6: Kendall's Tau b Correlation rank

Continued device usage Kendalls Tau Correlation			
Q.Num	Question	Rank	p-value
10	The activity tracker was comfortable to wear at night	0.348	0.003
17	Using the activity tracker helped me be more active	0.340	0.003
5	The activity tracker accurately tracked my physical activity	0.317	0.005
6	I was able to wear the device easily without help from another person	0.308	0.009
9	The activity tracker was comfortable to wear during the day	0.306	0.01
4	I think that monitoring my health 24 hours a day, 7 days a week is a good thing	0.264	0.02
5	I am comfortable with my health data being stored on the internet	0.264	0.02
13	I had no concerns about my privacy while wearing the device	0.253	0.035
2	Have you previously used a wrist worn activity tracker before the project?	0.209	0.094
14	I was happy to wear the sensor in public	0.206	0.08
8	I was able to perform my daily tasks as usual while wearing the device	0.202	0.093
18	Over the last week, how many days did you wear the device	0.187	0.112
19	Did you wear it at night-time?	0.169	0.177
16	I was happy to wear the sensor around the house	0.164	0.180

12	I was able to put on the device in a reasonable amount of time	0.119	0.305
1	Had you heard of wearable smart devices before the project	0.115	0.360
7	I was able to remove the device easily without help from another person	0.083	0.501
20	Did you remove the device during the day for reasons other than getting the device wet?	-0.078	0.531
3	Did you like the appearance of the wrist worn activity tracker	0.054	0.664
11	I was concerned that the device was not securely attached to me	-0.019	0.873

Table 7: Kendall's Tau b Correlation for each usability question

Further analysis was performed to evaluate the feasibility of features automatically capture by the activity tracker as predictors of continued usage in the future. Walking activity features (WT10M1, WT10M2, WS10M1, WS10M2, STS5) were compared with continued device usage using Kendall's rank correlation coefficient. Based on Question 17 (Using the activity tracker helped me be more active) being highly correlated with continued device usage we hypothesize that walking activity features will also correlate with continued device usage. Analysis was performed to test this hypothesis. The results of this analysis are presented in Table 8. Both of the 10-meter walk step count features (WS10M1 and WS10M2) had a strong correlation with the continued device usage question.

Continued device usage (Q21) Kendalls Tau Correlation		
Question	**Rank**	**p-value**
WS10M2	-0.255	0.018
WS10M1	-0.239	0.026
WT10M1	-0.194	0.059
WT10M2	-0.083	0.422
STS5	0.057	0.578

Table 8: Kendall's Tau b Correlation for each walking activity feature

4.3.1 A predictive model for continued device usage

An important criterion for wearable technologies is user acceptance. This increases the likelihood that individuals will continue to use the device long term and outside of periods when they are being actively monitored. Factors potentially influencing a user's acceptance of a wearable device include comfort, simplicity and device intrusiveness. For example, if a device requires frequent interaction, then it could become a nuisance. The metrics displayed in Table 7 are quite specific and therefore it would be worthwhile exploring these features in more detail. A predictive model is frequently used in statistics and machine learning techniques to model the current data and predict future outcomes. Within this section, we evaluate models that predict if a user will continue to use a device after the monitoring period. These predictions will be based on the usability questionnaire where the answer to Question 21 will be predicted based on the answers to the previous questions.

Feature selection. Prior to parameter tuning and classification, feature selection

was performed to find features that have potential to discriminate between continued device usage or not. This is of particular importance in this study as every additional feature required for accurate classification are additional data which needs to be captured from users. Feature selection focused on identifying the features which provide the best information for classifying usability Question 21 “Would you continue using the [wearable] device and app after the trial”. The features were chosen from the results of the Kendall’s Tau b Correlation from Table 7. A total of 3 feature subsets were chosen. The first subset is based on the 2 highest correlated features (Question 10 and Question 17) such that the selected feature had p=0.003. A second subset was selected to include features with p<=0.01 (Questions 6, 9, 10, 15, 17). Finally, a third subset was selected to include features with p<=0.1 (Questions 4, 5, 6, 9, 10, 13, 15, 17). For the remainder of the analysis, models developed using the three feature subsets will be known as the 2-feature model, 5-feature model, and 8-feature model respectively.

Predicting Continued Device Usage. Initial experimentation was performed using multiple classifiers, ranging from lazy classifiers, such as k-NN to ensemble learners such as Random Forest. From this experimentation we found that Random Forest provided the highest predictive performance on classifying if users would continue to use the device after the trial had ended. For comparison with the Random Forest models, regression multinomial models were also performed. These multinomial models are helpful for simplicity and interpretability of the study findings providing a high and quick overview.

In order to ensure that the Random Forest provides the best possible classification results, parameter tuning must be completed. To reduce the incidence of bias in parameter tuning, tuning was only performed on the training data (70% of the data). Table 9 displays the ranges tested during tuning for each hyperparameter. Parameter tuning was performed as it can get close to the performance of a grid search while implementing parameters which can reduce the performance impact of iteratively trying each combination. Regarding the multinomial models, all data was included to observe and assess the statistical/discrimination power of the model at once. For the Random Forest models, validation of final classification was achieved using a train/test split validation of 70/30 to check model accuracy.

	Range tested
Number of estimators	200 – 2000 (increments of 100)
Maximum features	Auto / sqrt
Maximum depth	10 – 110 (increments of 10)
Minimum samples per split	2, 5, 10
Minimum samples per leaf	1, 2, 4
Bootstrap	True / False

Table 9: Classification parameter tuning

Each Random Forest model individually had its parameters tuned using a randomised search, the range of which is detailed in Table 9. The two feature model parameters were as follows. Number of estimators=300, maximum features=auto, maximum depth=50, minimum samples per split=10, minimum samples per leaf=2 and bootstrap=false. The five feature model conversely used the following parameters. Number of estimators=300, maximum features=auto, maximum depth=50, minimum samples per split=10, minimum samples per leaf=2 and bootstrap=false. The eight feature model parameters were as follows. Number of estimators=600, maximum

features=auto, maximum depth=50, minimum samples per split=2, minimum samples per leaf=2 and bootstrap=false. The results from each of the 2, 5 and 8 feature models for both the Multinomial model and Random Forest model are displayed in Table 10.
The findings from the multinomial 2 feature model display an overall accuracy of 80% and the findings from the Random Forest 2 feature model correlate with an average accuracy of 80%, an average precision of 0.80 and an average recall value of 0.80. Both set of results show that a reasonably accurate prediction can be made of the usability Question 21.
By increasing the models features to 5, the multinomial model displays an overall percentage of 83.1% and the Random Forest model correlates with an average accuracy of 80%, an average precision of 0.80 and an average recall value of 0.80. The multinomial results have improved slightly compared to the findings from the 2-feature model whereas the Random Forest results remained the same.
The final model employs 8 parameters. The findings from this multinomial model display an overall percentage of 84.6% while the Random Forest model's average accuracy increases to 85%, displaying an average precision of 0.88 and an average recall value of 0.85. Both set of results improved from the 2-feature model and 5-feature model findings. Nonetheless, the improvement from the 2-feature model to the 8-feature model was 5%.

			Multinomial model Predicted class		Random Forest model Predicted class	
Features			No	Yes	No	Yes
2	Actual Class	No	15	8	5	1
		Yes	5	37	3	11
5	Actual Class	No	16	7	5	1
		Yes	4	38	3	11
8	Actual Class	No	17	6	5	0
		Yes	4	38	3	12

Table 10: Classification confusion matrix for 2, 5 and 8 feature models

5. General conclusions

The present study investigates the usability of commercial fitness trackers (and phones) in a population aged 65 and over across 4 different regions of the NPA. The interim deliverable reported on the setting up of the trial, the protocols , details of the kit and some preliminary reports from each of the regions.
One tangential finding that has been highlighted is the time delay that is now incurred with submitting ethical applications and receiving approval in universities, particularly after the implementation of GDPR. The university systems have become ultra-cautious and it behoves researchers to factor these potential delays into the planning of their projects and trials. Clinical partners do not seem to be subject to these delays due to the nature of their organisations.
The study was successfully completed across the 4 regions of the NPA at differing rates because of these delays and the clinical partners had finished before the pandemic struck whereas the academic partners were in the middle of their studies. Both Irish academic partners had plans to match the targeted numbers from Finland and Sweden but eventually had to concede defeat to Covid-19.

The design of the trial and choice of technology highlighted issues which were not thought of when conceiving of the Transnational Trial including:

- Questionnaires common across regions, available in their native language
- The added burden of translating the protocols/user guides for the Swedish and Finnish partners
- Technology which could be used across all regions – the initial choice of mobile phone for the project did not have a Finnish language option

This work used a combination of validated questionnaires to gather 65 elders opinions on the usability of an off-the-shelf wearable sensor system, the Xiaomi Mi Band 3. To gain further insights into the factors which may influence an elder wanting to use a wearable device, we also designed a bespoke usability questionnaire for this study. Various analyses were performed examining the statistics from the pre-trial questionnaires, summary statistics of the SUS score with respect to region/ gender/ wearables/ age, and findings that focused specifically on the final question from the bespoke usability questionnaire to determine what factors influence continued device usage.

The results from the SUS show that there is no notable difference in perceived system usability regardless of region, gender, or age, eliminating the notion that usability perception differs based on geographical location, sex, or deviation in elders age. It was also noted that there was no statistical difference in the SUS score depending on whether the participant was asked to wear one or two wearables. This is likely because two wrist worn sensors are still deemed to be unobtrusive in everyday life. Further research is required to observe whether these results would scale based on anatomical location or additional wearable sensors. Additionally, there is no statistically significant difference in the usability scores for a wearable sensor system regardless of whether the elderly participant had previously owned a wearable device or not. This is likely because of the training sessions that were provided to each participant at the start of the trial.

Previous works have suggested that usability and ease of use are as important as device accuracy when it comes to technology acceptance and device uptake. Using the bespoke questionnaire, we were able to determine that the most important factor that influenced continued device usage within an elderly cohort was device comfort. Feeling that the device was fit-for-purpose (i.e. it helped them achieve the task it claimed it would) was the second most important factor. We presented a Random Forest model with 80% accuracy using these two features which could be used as an early identifier of continued device use – for example if the user is asked these two questions after the first day of the study their response would be a clear sign whether the person is interested in using a wearable sensor system long-term. By including the top 8 ranked questions from the bespoke questionnaire as features to our model, the accuracy increased to 88%.

This study, although based on an activity tracker, suggests that the results are transferable to other wearable sensor systems. Future work will aim to test this hypothesis by using the same usability questionnaire on a cohort sampling other wearable technology, for example a wearable smart glove, smart insoles, or other wearable devices.

References

1. Ec.eu, 2018 https://ec.europa.eu/eurostat/cache/infographs/elderly/index.html [Accessed 6th December 2018].
2. Gov.uk, 2018 https://www.gov.uk/ [Accessed 6th December 2018].
3. Rosenberger ME, Buman MP, Haskell WL, McConnell MV, Carstensen LL. 24 Hours of sleep, sedentary behavior, and physical activity with nine wearable devices. Med Sci Sports Exerc 2016; 48, 3, p. 457-465.
4. Tedesco S, Sica M, Garbay T, Barton J, O'Flynn B. A comprehensive comparison of commercial wrist-worn trackers in a young cohort in a lab-environment. IEEE PERCOM Workshops 2018, Athens, GR.
5. Xie J, Wen D, Liang L, Jia Y, Gao L, Lei J. Evaluating the validity of current mainstream wearable devices in fitness tracking under various physical activities: Comparative study. JMIR mHealth uHealth 2018; 6, 4, e94.
6. 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, p. 1277.
7. Ehn M, Eriksson LC, Akerberg N, Johansson AC. Activity monitors as support for older persons' physical activity in daily life: Qualitative study of the users' experiences. JMIR Mhealth Uhealth 2018; 6, 2, e34.
8. Puri A, Kim B, Nguyen O, Stolee P, Tung J, Lee J. User acceptance of wrist-worn activity trackers among community-dwellers older adults: Mixed method study. JMIR Mhealth Uhealth 2017; 5, 11, e173.
9. Tedesco S, Sica M, Ancillao A, Timmons S, Barton J, O'Flynn B. Validity Evaluation of the Fitbit Charge2 and the Garmin vivosmart HR+ in Free-Living Environments in an Older Adult Cohort. JMIR Mhealth Uhealth 2019;7(6):e13084
10. Hibbard JH, Greene J. What the evidence shows about patient activation: better health outcomes and care experiences; fewer data on costs. Health affairs (Project Hope). 2013;32:207–14.
11. Brooke, J. SUS: A quick and dirty usability scale. Usability Eval. Ind. 1995, 189.
12. Steinert, A.; Haesner, M.; Steinhagen-Thiessen, E. Activity-tracking devices for older adults:comparison and preferences. Universal Access in the Information Society 2018, 17, 411–419.doi:10.1007/s10209-017-0539-7.
13. Bangor, A.; Kortum, P.T.; Miller, J.T. An Empirical Evaluation of the System Usability Scale. International Journal of Human Computer Interaction 2008, 24, 574 594, [https://doi.org/10.1080/10447310802205776]. doi:10.1080/10447310802205776.