Crowdsourced Data Collection of Physical Activity and Health Status: An App Solution

Daniel Kelly[1(✉)], Brian Caulfield[2], and Kevin Curran[1]

[1] Computer Science Research Institute, Ulster University, Coleraine, Northern Ireland
d.kelly@ulster.ac.uk
[2] INSIGHT Center, University College Dublin, Dublin, Ireland

Abstract. Health status measurements are vital in understanding a patient's health. However, current means of measuring health status, such as questionnaires, are limited. Research has shown that there is a need for more objective and accurate methods of measuring health status. We postulate that novel sensor solutions could be used to make observations about a patients' behaviour and make predictions relating to their health status. In order to achieve this overall goal, the problem of building a dataset comprising behaviour observations, from sensors, and health status measure must be addressed. In this work, we propose a crowd-sourced solution to this dataset problem where a Smartphone App is developed in order to facilitate in the collection of behaviour data, via sensors, and health status information. Results show that, after just 4 months, 1311 people have downloaded the App and 541 participants have completed a health status questionnaire (SF-36). Preliminary analysis of the data also shows a statistically significant correlation between the amount of time a participant is active and the health status of the participant.

1 Introduction

Chronic diseases are the most common causes of death and disability throughout the world [1]. In the UK, for example, 70% of all healthcare costs are chronic disease related [2]. Health Status measurements, such as Health Related Quality of Life (HRQOL), are used as a means of quantifying the impact of chronic disease on a patients' daily life [3]. These measures are vital in understanding a patients health and their response to particular treatments and have become a central feature in many chronic disease studies [4]. Questionnaires are used to evaluate health status. However, evidence suggests questionnaire results are only useful in large groups and should not be relied upon on an individual basis [5]. It is only worth continuing to prescribe symptomatic treatments if the patient can report benefit, but due to the limited reliability of health status questionnaires for individual patients there is currently no way of accurately accessing that benefit.

There is therefore a need for accurate and individualized methods for clinicians to assess functional aspects of a patient's life. An innovative solution to

P. Perego et al. (Eds.): MobiHealth 2016, LNICST 192, pp. 151–159, 2017.
DOI: 10.1007/978-3-319-58877-3_20

this need is to utilize remote sensing technologies in the community, rather than questionnaires, to compute accurate, objective and individualized QOL measurements.

Our overall research goal is therefore to develop an unobtrusive sensing system which can objectively measure a persons' longitudinal behaviour and make accurate predictions about their health status based on their behaviour. However, there exists an initial problem which must be solved prior to solving this overall problem health status prediction. This initial problem relates the collection of appropriate data. In order to accurately model the mapping between sensor data and health status, a data-set comprising mobile sensor data and corresponding health status information must be acquired. The data-set must include participants with a broad spectrum of health measurements. Recording patient data alone would represent only a small window in the health status spectrum.

Modern smartphones are equipped with multiple sensors. The combination of these sensors, built within the common and non-invasive form factor of a mobile phone, have the potential of tracing human activities at scales that were previously unattainable. Smartphones can therefore enable a new type of data collection by harnessing the power of the crowd. Crowdsourced data collection, using smart-phones, presents a major opportunity to collect sensor data from a large, and varied, set of participants. The aim of this work is therefore to develop a smartphone App to facilitate the crowdsourced based data collection of motion sensor data and health status information. Additionally, we will discuss preliminary observations made from data which have been collected.

2 Methods

In this section, we will describe the development of an Android App aimed at recording motion sensor and health status information.

2.1 Motion Sensors

The Accelerometer, built into a participants' Smartphone, is used by the App to measure physical activity. Sensor data capture and recording is performed in the background, and data is recorded constantly while the App is enabled.

We postulate that features, extracted from motion signals, relating to the duration a person was stationary and active could potentially be used as a health status indicator. For example, using the total time a person was active as a feature. In order to investigate this, we propose two duration based measures: (1) Total Movement Duration (TMD) and (2) Average Stationary Period (ASP). TMD specifies the total amount of time in which the phone was detected as moving during a given day. The phone was deemed to be moving if the variance of the accelerometer magnitude was greater than a predefined threshold. For each 2 s window where the phone was deemed to be moving, 2 s were added to the overall TMD measure for that day. ASP was calculated as the average of a set of stationary period durations for a given day. The set of stationary period durations store the set of times between when the phone stopped moving and when the phone started to move again (i.e. the amount of time the phone was stationary). ASP therefore stores the average period of time a participant's phone was stationary during a given day.

In order to reduce the size of data being uploaded by a participant, we implement a system whereby each hour of motion sensor data is processed on the Smartphone, and extracted features for each hour are then uploaded to the server. These features include TMD and ASP features, as well as additional Accelerometer and Gyroscope statistical feature to be utilized in future works. Additional data processing and feature extraction can then be performed on the server using the hourly data.

2.2 User Interface

The App features two main User Interface (UI) sections. The first provides users with activity feedback and the second provides a means for users to take a health status questionnaire.

Activity Feedback. A study carried out in 2012 showed that 1 in 5 smartphone users had a health tracking App installed on their phone [6]. Health tracking is therefore a genre of App which the general public actively install on their smartphone. In order to get potential participants interested in and contributing to our data collection, we postulated that the App should be designed and marketed as a Health tracking App. Based on a review of the top health tracking Apps on the Google Play store, we found that a common feature of all health tracking Apps was that some level of quantitative feedback was provided to the user on their activity levels.

To increase App downloads, and improve user retention within the experiment, the App was therefore designed to provide users with visual feedback on the duration and intensity of their activities over time using graphs and statistics calculated from motion sensors (see Fig. 1). The App has received 14 reviews on Google Play, with an average rating of 4.6/5.

Health Status Interface. A key aim of this study is to record health status for a set of participants with a broad spectrum of health measurements. We utilize the Short-Form 36 (SF-36) survey in order to measure participant health status. The SF-36 is a non-illness specific health status measure which has been validated in a general adult population [7] and in a chronic illness patient population [8,9].

The SF-36 is a general health instrument that measures eight health related concepts: physical functioning (PF-10 items), role limitations due to physical problems (RP-4 items), bodily pain (BP-2 items), general health perceptions (GH-5 items), vitality (VT-4 items), social functioning (SF-2 items), role limitations due to emotional problems (RE-3 items), and perceived mental health

Fig. 1. "Health-U" App - (Left) Visual feedback showing current activity, (Middle) Activity history showing daily activity, (Right) Health status questionnaire.

(MH-5 items). These eight scales can be aggregated into two summary component measures: the Physical (PCS) and Mental (MCS) Component Summary Scores [10]. A questionnaire UI screen was integrated into the App to allow users to answer the SF-36 questions (See Fig. 1 (Right)). Questions are multiple choice and radio buttons are used to select answers to individual questions.

3 Results

The key aims of this study is to (1) develop an App based crowdsourced data collection platform which can record motion sensor and health status information and (2) investigate the feasibility of using such an App to build a dataset which can be later used to develop models linking activity to health status.

The App, named "Health-U", was published on Google Play and anyone with an Android phone could download and install the App and participate in the study. Upon launching the App for the first time, participants are shown a participant consent screen where details about the study, and data collected during the study, are explained. Participants are then given the choice to consent via a button labelled "I Consent" or to reject via a button labelled "Do not participate". Ethical approval for this study was granted by Ulster University Ethics committee and the contents of the participant consent screen was reviewed by the Ethics Committee.

The App was downloaded by a total of 1311 users in the first four months that the App was live. Of the 1311 downloads, 541 participants completed the SF-36 questionnaire. An average of 114 h of sensor data was uploaded by each participant. Of the 541 participants who completed the questionnaire, 263 participants (48.6%) uploaded at least 1 h of sensor data. This statistic shows approximately half of all users which downloaded the App and completed the questionnaire, disabled the sensing, or uninstalled the App, within an hour of installing the App. Figure 2 details the number of participants that uploaded a minimum number of hours. For example, it can be seen that 115 participants uploaded at least 72 h of data.

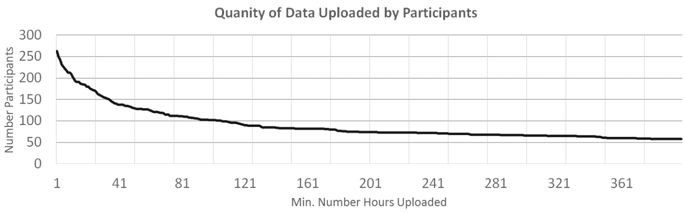

Fig. 2. Quantity of data uploaded by participants

Based on the SF-36 data uploaded by the 541 participants, Table 1 details the mean and standard deviation SF-36 scores, for the 8 different concepts and the 2 summary measures, of participants based on categories of gender, age and country. It can be seen, for example, that PF is generally higher in younger participants. Conversely, MH is generally higher for older participants.

It can be seen that the majority of the average scores are lower when compared to the study conducted in 2007 by Burholt et al. [11]. A possible explanation for this is that it has been shown that lower SF-36 scores are obtained when the questionnaire is self-administered when compared with scores which were obtained when the questionnaire was interviewer-administration [12].

3.1 Activity Duration

An initial statistical analysis of sensor information was carried out by computing the overall mean and standard deviation (SD) for the TMD and ASP measures. Results showed that, on average, a participants' phone moves for an average of 1 h and 33 min per day (SD = 4 min 50 s). Additionally, a participants' phone stays stationary for an average period of 22 min (SD = 14 min).

A qualitative analysis of movement data was performed in order to investigate potential links between movement and health status. Figure 3 shows movement duration data (TMD), and individual SF-36 scores, for two female participants (both aged 40–50). As detailed in the Fig. 3, Participant A has significantly lower SF-36 measures than Participant B. Interestingly, it can be seen that a large portion of Participant A's time is stationary, while data for Participant B shows that movement occurs regularly between 10 am and 11 pm. This does give an indication of the potential merit of using activity duration as an indicator of health status.

Further to the qualitative analysis above, we perform a quantitative evaluation to further investigate potential links between SF-36 scores and movement durations. Table 2 details correlations between the 10 SF-36 scores (8 SF-36 concepts and 2 summary measures) and the two duration measures TMD and ASP.

Table 1. SF-36 scores for participant demographics (Gender, Age, Country).

		PCS	MCS	PF	RP	BP	GH	VT	SF	RE	MH
	N	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)	Mean (SD)
Gender											
Female	306	49.0 (9.0)	48.6 (9.5)	71.8 (26.6)	75.1 (30.7)	66.3 (28.3)	52.5 (23.3)	45.3 (21.5)	61.9 (28.5)	60.8 (33.3)	53.7 (23.8)
Male	234	51.2 (7.8)	51.7 (7.6)	75.2 (27.2)	78.2 (30.0)	72.9 (23.9)	58.0 (20.9)	52.2 (17.8)	67.4 (28.4)	67.0 (31.6)	61.2 (21.2)
Age											
18–21	105	50.5 (8.0)	49.4 (9.0)	77.0 (25.3)	76.8 (30.1)	72.0 (26.5)	55.5 (21.2)	47.5 (19.0)	63.0 (29.0)	61.5 (34.0)	55.1 (23.6)
22–25	77	51.2 (7.6)	49.3 (9.2)	79.2 (22.8)	82.3 (23.8)	73.1 (21.1)	53.3 (21.8)	46.9 (19.9)	65.1 (27.9)	61.2 (31.8)	54.5 (23.0)
26–30	62	49.0 (8.0)	48.7 (9.7)	72.2 (25.9)	73.0 (32.2)	70.1 (26.2)	52.2 (21.6)	44.9 (21.0)	59.5 (26.8)	59.3 (33.3)	55.6 (25.1)
31–35	61	49.2 (8.0)	48.6 (8.3)	72.3 (25.3)	75.9 (26.0)	67.8 (25.9)	52.0 (21.0)	48.3 (21.4)	57.9 (28.7)	60.1 (33.3)	52.6 (20.6)
36–40	55	50.0 (8.8)	50.9 (8.4)	68.9 (27.5)	73.4 (31.6)	72.4 (28.2)	57.9 (22.5)	50.7 (19.4)	65.0 (29.2)	60.5 (32.4)	60.7 (22.3)
41–50	90	48.8 (10.4)	50.1 (8.9)	70.1 (32.7)	74.0 (36.9)	65.7 (28.8)	52.5 (25.2)	46.8 (21.7)	63.1 (29.4)	69.0 (30.7)	58.5 (22.9)
51–60	49	49.9 (8.9)	51.9 (9.0)	68.3 (26.9)	75.5 (31.4)	62.8 (30.8)	59.5 (23.5)	50.8 (20.5)	73.0 (28.1)	69.6 (33.3)	61.5 (21.5)
60+	38	51.1 (8.1)	52.5 (6.9)	73.1 (23.9)	80.3 (24.8)	65.4 (24.1)	59.6 (20.0)	55.0 (16.0)	72.8 (24.4)	69.0 (28.8)	61.8 (21.8)
Country											
UK	145	50.1 (9.3)	49.7 (9.4)	75.1 (29.0)	79.1 (31.2)	69.4 (28.1)	52.5 (23.0)	46.8 (20.6)	63.3 (29.6)	65.6 (34.1)	56.4 (24.4)
USA	82	49.5 (8.6)	51.0 (9.2)	70.2 (28.2)	73.0 (32.1)	65.6 (25.0)	57.1 (23.6)	48.6 (22.1)	69.7 (29.3)	67.8 (33.3)	60.1 (24.7)
Ireland	75	50.6 (9.0)	49.7 (9.1)	76.0 (25.9)	78.7 (30.2)	71.9 (29.0)	55.7 (23.8)	47.2 (20.4)	64.0 (28.9)	64.7 (32.3)	55.9 (24.1)
Canada	57	49.6 (7.3)	48.4 (8.5)	77.5 (22.4)	79.6 (26.1)	68.0 (23.9)	50.2 (20.2)	43.4 (20.8)	60.9 (29.1)	61.1 (32.4)	55.0 (20.7)
Spain	27	52.7 (7.1)	53.3 (8.3)	78.1 (20.9)	80.4 (26.1)	76.5 (28.2)	61.5 (16.7)	57.0 (19.1)	69.4 (28.3)	70.7 (32.6)	63.7 (20.6)
Australia	48	48.4 (8.6)	49.1 (7.8)	67.4 (27.8)	72.2 (30.6)	64.0 (24.4)	53.6 (22.0)	48.7 (17.8)	63.8 (23.7)	55.1 (31.5)	55.4 (20.4)
New Zealand	33	50.9 (6.7)	49.4 (8.2)	78.0 (23.2)	82.2 (27.2)	69.6 (24.4)	55.9 (19.7)	46.0 (17.7)	62.2 (29.3)	63.9 (28.4)	55.5 (20.6)
Other	74	49.3 (8.3)	50.2 (8.3)	67.4 (26.4)	68.7 (30.6)	71.4 (26.0)	56.9 (23.0)	53.2 (17.9)	62.0 (26.7)	58.5 (30.9)	56.3 (21.1)
Overall	541	49.9 (8.6)	49.9 (8.9)	73.3 (26.9)	76.4 (30.4)	69.2 (26.7)	54.7 (22.5)	48.3 (20.2)	64.2 (28.6)	63.6 (32.7)	56.9 (23.0)
Burholt [11]	*13917*	*N/A*	*N/A*	*77.8 (30.0)*	*78.3 (32.3)*	*70.1 (28.9)*	*66.2 (24.0)*	*57.3 (22.3)*	*80.2 (28.1)*	*87.0 (26.0)*	*74.0 (19.9)*

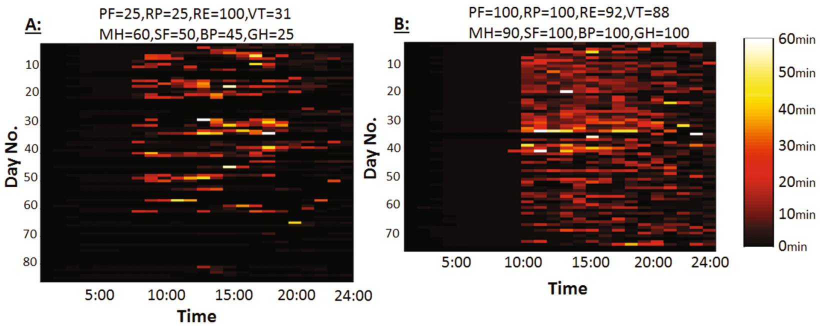

Fig. 3. Sample movement durations (TMD) for each day and hour, for 2 participants. Blank (black) areas of the graph denote no motion recorded for that hour. This can be due to the sensor being turned off, or because the phone remained stationary for the entire hour

Results showed that correlations between ASP and the SF-36 measures were not significant. However, results do show that the was a statistically significant correlation between TMD and the PCS, MCS, PF, RP, RE and SF concepts of the SF-36 scale. In particular, a correlation of $r = 0.221$ was shown between TMD and the RP component. The largest correlation between SF-36 and ASP was for the BP component, with $r = 0.042$. Table 2 also details the average TMD and ASP for 5 different ranges of scores for the different SF-36 concepts.

Table 2. Correlation between duration measures and SF-36 scores (* = Statistically significant correlation, where p-value < 0.05, calculated using two-tailed test)

			SF-36 Bands				
Measure SF-36 Activity		**Correlation**	**0-20** Mean (SD)	**21-40** Mean (SD)	**41-60** Mean (SD)	**61-80** Mean (SD)	**81-100** Mean (SD)
PCS			*N=0*	*N=76*	*N=419*	*N=46*	*N=0*
	(TMD)	0.195*		54m (2m:23s)	100m (4m:54s)	101m (5m:39s)	
	(ASP)	0.017		30m (12m:44s)	21m (13m:10s)	17m (16m:35s)	
MCS			*N=0*	*N=77*	*N=393*	*N=71*	*N=0*
	(TMD)	0.150*		63m (3m:3s)	100m (5m:3s)	96m (4m:51s)	
	(ASP)	0.038		30m (13m:10s)	21m (13m:8s)	18m (15m:18s)	
PF			*N=28*	*N=47*	*N=102*	*N=86*	*N=278*
	(TMD)	0.206*	66m (2m:56s)	68m (3m:19s)	70m (4m:32s)	99m (5m:31s)	105m (4m:48s)
	(ASP)	-0.021	32m (12m:53s)	20m (13m:15s)	28m (13m:53s)	21m (12m:9s)	20m (13m:47s)
RP			*N=34*	*N=53*	*N=55*	*N=86*	*N=312*
	(TMD)	0.221*	75m (3m:2s)	66m (5m:19s)	64m (3m:21s)	74m (3m:23s)	109m (5m:9s)
	(ASP)	-0.043	23m (13m:46s)	27m (14m:19s)	24m (13m:5s)	25m (10m:52s)	20m (13m:54s)
RE			*N=61*	*N=57*	*N=131*	*N=77*	*N=211*
	(TMD)	0.182*	56m (2m:36s)	89m (4m:20s)	77m (3m:44s)	127m (6m:56s)	104m (4m:45s)
	(ASP)	0.051	37m (12m:17s)	16m (14m:34s)	19m (14m:23s)	21m (12m:47s)	22m (13m:15s)
VT			*N=66*	*N=98*	*N=230*	*N=110*	*N=37*
	(TMD)	0.114	69m (4m:22s)	99m (3m:45s)	89m (4m:55s)	119m (5m:32s)	83m (4m:12s)
	(ASP)	0.031	24m (12m:42s)	21m (11m:36s)	23m (14m:9s)	19m (12m:56s)	20m (14m:30s)
MH			*N=41*	*N=95*	*N=206*	*N=111*	*N=88*
	(TMD)	0.105	55m (2m:51s)	95m (4m:34s)	92m (5m:3s)	116m (5m:1s)	90m (4m:36s)
	(ASP)	0.022	33m (14m:15s)	17m (11m:5s)	21m (14m:20s)	20m (12m:43s)	20m (13m:54s)
SF			*N=27*	*N=100*	*N=79*	*N=143*	*N=186*
	(TMD)	0.138*	46m (2m:10s)	89m (4m:20s)	77m (4m:59s)	115m (5m:7s)	96m (4m:47s)
	(ASP)	0.068	24m (11m:43s)	23m (13m:13s)	20m (13m:55s)	20m (11m:44s)	22m (14m:27s)
BP			*N=24*	*N=56*	*N=111*	*N=148*	*N=201*
	(TMD)	0.057	102m (3m:34s)	80m (3m:36s)	83m (4m:11s)	115m (6m:1s)	89m (4m:35s)
	(ASP)	0.096	20m (11m:29s)	24m (11m:9s)	24m (12m:27s)	17m (14m:30s)	22m (14m:6s)
GH			*N=48*	*N=105*	*N=174*	*N=152*	*N=60*
	(TMD)	0.062	72m (4m:50s)	92m (5m:17s)	97m (4m:49s)	95m (3m:54s)	100m (5m:52s)
	(ASP)	-0.019	25m (10m:46s)	27m (14m:2s)	21m (12m:52s)	20m (12m:27s)	22m (16m:46s)

While results have shown a statistically significant correlation between duration of activity and health status, these correlations are not strong enough to make accurate predictions about a persons' health status. Based on these results, we conclude that due to the real world and inherent uncontrolled nature of this study, where participants use the sensing modality without researcher supervision, duration of activity on its own cannot be used to consistently infer the health status of a participant. It is possible that periods of inactivity relate to periods where the phone was simply not being used/carried by the participant. During these periods, the sensor would infer that the person in being sedentary when it is possible that the person was in fact being active. In particular, the ASP measures showed no correlation with the health status measures. However, results did show that while correlation between TMD and different SF-36 components were negligible, the correlations were statistically significant.

The preliminary investigates discussed in this work have therefore indicated that additional features should be investigated to compliment the duration based features. Additional features could relate to activity intensity, type and frequency computed from accelerometer and gyroscope sensor data.

4 Conclusion

We postulate that smartphone sensors could be used to make automatic predictions about patient health status. In order to move towards this overall goal, we must first propose solutions to the problem of recording a large dataset of behaviour observations and health status information from a broad spectrum of participants. In this work, we propose a crowd-sourced solution to this problem, where a smartphone App is developed to record behaviour observations, via the recording of sensor data, and health status information, via a built-in SF-36 questionnaire.

Preliminary analysis of data obtained from our proposed crowd-sourced system demonstrates the feasibility of our solution. In just 4 months, 1311 people downloaded the App and, of these downloads, 541 participants completed the SF-36 questionnaire. Initial examination of the sensor data showed a correlation between the total amount of movement performed per day, by a participant, and the health status of the participant. While this correlation was statistically significant, the correlation was not strong. We therefore conclude that while these results demonstrate the potential of our proposed solution, further work is needed in terms of developing more discriminant features and evaluating regression models to make health status predictions.

References

1. Viswanathan, M., Golin, C.E., Jones, C.D., Ashok, M., Blalock, S.J., Wines, R.C.M., Coker-Schwimmer, E.J.L., Rosen, D.L., Sista, P., Lohr, K.N.: Interventions to improve adherence to self-administered medications for chronic diseases in the United States: a systematic review, pp. 785–795, December 2012

2. UK Department of Health: Long Term Conditions Compendium of Information, 3rd edn. (2012)
3. Kocks, J.W.H., Tuinenga, M.G., Uil, S.M., van den Berg, J.W.K., Ståhl, E., van der Molen, T.: Health status measurement in COPD: the minimal clinically important difference of the clinical COPD questionnaire. Respir. Res. **7**(i), 62 (2006)
4. Jones, P.W.: Health status measurement in chronic obstructive pulmonary disease. Thorax **56**(11), 880–887 (2001)
5. Pitta, F., Troosters, T., Probst, V.S., Spruit, M.A., Decramer, M., Gosselink, R.: Quantifying physical activity in daily life with questionnaires and motion sensors in COPD. Eur. Respir. J.: Official J. Eur. Soc. Clin. Respir. Physiol. **27**(5), 1040–1055 (2006)
6. Fox, S., Duggan, M.: Mobile Health 2012, Pew Research Center. Technical report (2012)
7. Bize, R., Johnson, J.A., Plotnikoff, R.C.: Physical activity level and health-related quality of life in the general adult population: a systematic review, pp. 401–415 (2007)
8. Boueri, F.M., Bucher-Bartelson, B.L., Glenn, K.A., Make, B.J.: Quality of life measured with a generic instrument (Short Form-36) improves following pulmonary rehabilitation in patients with COPD. Chest **119**(1), 77–84 (2001). http://www.ncbi.nlm.nih.gov/pubmed/11157587
9. Ståhl, E., Lindberg, A., Jansson, S.-A., Rönmark, E., Svensson, K., Andersson, F., Löfdahl, C.-G., Lundbäck, B.: Health-related quality of life is related to COPD disease severity. Health Qual. Outcomes **3**, 56 (2005)
10. Farivar, S.S., Cunningham, W.E., Hays, R.D.: Correlated physical and mental health summary scores for the SF-36 and SF-12 Health Survey, V. 1. Health Qual. Life Outcomes **5**(1), 54 (2007)
11. Burholt, V., Nash, P.: Short Form 36 (SF-36) health survey questionnaire: normative data for wales. J. Publ. Health (Oxf. Engl.) **33**(4), 587–603 (2011)
12. Lyons, R.A., Wareham, K., Lucas, M., Price, D., Williams, J., Hutchings, H.A.: SF-36 scores vary by method of administration: implications for study design. J. Publ. Health Med. **21**(1), 41–45 (1999)