Methods

STUDY GROUP

Patients had a history compatible with COPD, at least 10 pack years of smoking history and evidence of chronic airflow limitation (post bronchodilator $FEV_1/FVC < 0.7$) [1]. Exclusion criteria included: 1) Orthopedic, neurological, other complaints that impair normal movement patterns, 2) respiratory diseases other than COPD (e.g. asthma), 3) COPD is not the primary source of activity limitation, 4) Cognitive impairment that precludes participation, 5) Patients on variable doses of diuretics that can interfere with the doubly labeled water method; and, 6) Any hospital admission or COPD exacerbations within the previous 4 weeks.

All participants were informed of any risks and discomfort associated with the study, and written informed consent was obtained. The study was approved by the local Ethics Committee's (The Medical Ethical Board of the University Hospitals Leuven [Leuven, Belgium]; NRES Committee London – Bloomsbury [London, United Kingdom]; Sotiria Hospital Scientific and Ethics Committee [Athens, Greece] and Lothian Regional Ethics Committee [Edinburgh, United Kingdom]) in each of the participating centres as well as the independent ethical board of the PROactive project.

Study Design

This was a multi-centre crossectional validation study with 14 days of continuous assessment. No interventions were applied during the study period. Patients visited the respective centre once (V1) before the ingestion of the DLW for the assessment of: a) Lung function, b) anthropometric and demographic characteristics, c) exercise tolerance; and, d) health status. On a subsequent visit (V2) resting energy

expenditure (REE) (also measured in visit 5) was measured and patients were asked to wear simultaneously up to 4 activity monitors. Patients were instructed to ware the monitors during the day from the moment they were awake until going to bed in the evening, except for bathing and showering or aquatic activities. The monitors were used according to the manufacturers' guidelines and this was standardized across centers by strict adherence to standard operating procedures agreed during a training work shop prior to starting the study. The data corresponding to not worn periods were not included in the analysis. All remaining data were considered activity data and included in the analysis.

Data corresponding to the first week of assessment were retrieved from the different monitors on a subsequent visit V4 (one week later) and the monitors were returned to the patients for a second week of assessment. Data corresponding to the second week of assessment were retrieved in the last visit (V5). This resulted in 2 weeks of data collection.

Assessments

All patients had the following baseline assessments:

- **Lung function and dyspnoea score**

Spirometry and single breath transfer factor for carbon dioxide was measured according to the American Thoracic Society/European Respiratory Society standards [2] after administering 400µg salbutamol. Post bronchodilator static lung volumes were assessed using a body-box (whole body plethysmography). Modified Medical Research Council scale was used to evaluate dyspnoea.

- **Anthropometric measurements**

Body weight was measured to the nearest 0.1 kg with a digital scale and height was measured to the nearest 0.5 cm with a stadiometer. Body mass index (BMI) was calculated as the ratio of weight (in kilograms) to height (in metres) squared.

- **Exercise tolerance**

Functional exercise capacity was performed. A six minute walking test (6MWT) was performed according to American Thoracic Society (ATS) guidelines[3]. A symptom-limited incremental cycle-ergometer exercise test was conducted to assess the peak exercise tolerance and peak oxygen consumption (maximal exercise capacity). A minimum of 2 minutes of baseline assessment and three minutes of unloaded cycling were used preceding the incremental loading. Increments of 10-20Watt per minute were applied. During the test a 12-lead ECG and regular measures of blood pressure were obtained to assess cardiovascular safety of exercise. Patients were encouraged to conduct a maximum effort. Unloaded cycling oxygen upake (VO_2), peak work rate (W_{peak}), peak VO_2 (VO_{2peak}) and peak carbon dioxide production (VCO_{2peak}), peak heart rate (HR), peak ventilation (VE), blood pressure, oxygen saturation and ratings of perceived exertion (Borg) are reported. Peak values were those obtained in the last 30 seconds of the incremental exercise test.

- **Health status**

Questionnaires employed were the St. George's Respiratory Questionnaire (SGRQ) [4] and the COPD Assessment Test (CAT)[5].

Usability of physical activity monitors

The usability of the six monitors was assessed directly from patients using a twelve question questionnaire specifically designed for this purpose. The questionnaire is displayed in full below. Briefly the questions did examine patient's perception on the acceptance of the monitor, whether the monitors were felt to be intrusive, or obstructing daily life activity and whether they were considered user friendly. Patients were also asked how long they would be willing to wear the monitors in the context of clinical studies.

Results

Every centre used three to four monitors resulting in differences in number of patients assessed with the different monitors. The characteristics of these patients according to the monitor used are depicted in **Table S1**. Although compliance (assessed as days with more than 10 h of use over days intended to assess) range from 79% to 91% (**Table S1**) it should be highlighted that the SenseWear, a monitor worn by most of the subjects participating in the study that sense the contact of the monitor with the skin, is the monitor providing the most accurate estimates of wearing time.

Results of physical activity outcomes for each monitor are depicted in **Table S2A (Weekdays) and S2B (Weekend)**. Correlations between physical activity monitors outputs and DLW-measured energy expenditure variables are depicted in **Table S3**.

Correlation with physiological variables

Correlations of all six monitor ***measured variables*** against physiological variables of exercise capacity such as peak exercise oxygen uptake (VO_{2peak}) and power (W_{peak}) or 6MWD (m) are shown in **Table S4**. Except the Lifecorder plus which showed no statistically significant correlation with the VO_{2peak} and the 6MWD, the rest of the monitors showed statistically significant correlations with functional (6MWD) and maximal exercise capacity (VO_{2peak} or W_{peak}). The Lifecorder plus monitor showed the poorest correlation with exercise tolerance variables (average r=0.29), while the Actigraph GT3X showed the best average correlation (average r=0.82).

Inter-device reliability

Correlations between different outcomes ***measured*** by the six PAM assessed are shown in **Table S5A**. Correlations between different outcomes ***estimated*** by the six

PAM assessed are shown in **Table S5B.** The average Pearson correlation coefficient of all the Pearson correlations for the ***measured*** variables (**Table S5A**) was statistically different to the average of all the Pearson correlation coefficients for the ***estimated*** variables (**Table S5B**) (0.65±0.03 vs 0.37±0.09 measured and estimated variables respectively, $p<0.001$).

As an exploratory analysis, agreement between the DLW-measured TEE (kCal) and estimation of TEE by PAM (for the monitors reporting TEE [kCal] as outcomes) (SenseWear Armband, RT3 and Lifecorder plus) is shown as Bland and Altman plots in **Figure S1**. The mean difference between TEE measure by DLW and the estimated from the SenseWear Armband, the RT3 and the Lifecorder plus were -367.2, -273.8 and -335.2 kCal respectively. Similarly, correlations between TEE measured by DLW and these estimated parameters were inferior to those shown in **Figure 1** of the main manuscript with measured parameters: (Lifecorder $r=0.33$, $p<0.05$; RT3 $r=0.42$, $p<0.05$; SenseWear $r=0.21$, p=ns).

Repeatability of the activity measurement in two consecutive weeks

PA was assessed over two consecutive weeks for all PAMs. No statistically significant differences between the first and the second week were found for any of the variables for each monitor.

A subset of ***measured*** outputs: Lifecorder plus (AS), ActiWatch Spectrum (AC), RT3 (VMU), DynaPort MoveMonitor (Steps), Actigraph GT3X (Steps), SenseWear Armband (Steps) showing repeatability of PA assessment in two consecutive weeks in this population of COPD patients is shown in **Figure S2 and S3**. To evaluate whether including a second week of measurements added any power to the assessment of PA, we compared the standard deviation of three of the most accurate

monitors DynaPort MoveMonitor (Steps), Actigraph GT3X (Steps), SenseWear Armband (Steps) between one week (first week) of measurement and the whole period of 14 days (two weeks). The levels of variability were similar regardless of whether it was based on one or two weeks of measurement. Moreover, we have estimated repeatability of the data according to the British Standard Institution [6]. Accordingly, we have expected 95 % of the differences between week I and week II to less than two standard deviations (of the difference) and this was the case for all six monitors.

Table S1. Number of patients assessed with each of the monitors and their characteristics.

	Lifecorder	Actiwatch	RT3	DynaPort	GT3X	SenseWear
n	40	40	39	40	39	73
days	471	453	412	443	463	
Compliance	0.90	0.87	0.81	0.91	0.85	0.79
M/F	28/12	29/11	27/12	33/7	32/7	58/15
	mean ± SD	mean ± SD	mean ± SD	mean ± SD	mean ± SD	mean ± SD
Age (Years)	69 ± 5.8	69 ± 6.1	70 ± 5.9	67.3 ± 6.2	69 ± 6.6	68 ± 6.4
BMI ($Kg.m^2$)	27.2 ± 4.7	26.8 ± 4.8	25.9 ± 5.18	27.2 ± 401	26.8 ± 4.5	26.4 ± 4.5
mMRC	2.4 ± 0.8	2.3 ± 0.8	2.6 ± 0.9	2.2 ± 0.7	2.4 ± 0.9	2.4 ± 0.9
FEV_1 (L)	1.5 ± 0.5	1.6 ± 0.6	1.4 ± 0.6	1.6 ± 0.4	1.5 ± 0.5	1.5 ± 0.5
FEV_1 (% pred)	59.5 ± 19.5	61.4 ± 20.4	55.2 ± 23.1	57.9 ± 14.6	53.2 ± 18.3	56.0 ± 18.6
FVC (L)	2.9 ± 0.8	3.3 ± 0.8	3.3 ± 0.9	3.2 ± 0.8	3.5 ± 0.8	3.3 ± 0.8
FVC (% pred)	90.4 ± 20.4	98.4 ± 17.6	99.8 ± 3.4	89.7 ± 17.5	98.7 ± 19.3	94.1 ± 19.2
FEV_1/FVC	0.5 ± 0.1	0.5 ± 0.1	0.4 ± 21.4	0.5 ± 0.1	0.4 ± 0.1	0.5 ± 0.1
BODE	5.6 ± 1.7	5.4 ± 1.7	6.2 ± 1.9	5.3 ± 1.4	5.9 ± 1.9	5.7 ± 1.7
SGRQt	45 ± 19.7	43 ± 23.1	45 ± 23.1	40 ± 16.6	39 ± 20.2	42 ± 19.9
SGRQa	57.6 ± 24.1	54.1 ± 27.5	57.7 ± 31.4	55.1 ± 19.6	55.3 ± 27.4	55.2 ± 25.3
CAT	15.5 ± 4.5	17.3 ± 7.6	17.7 ± 7.8	13.4 ± 6.6	15.8 ± 7.6	15.6 ± 7.4
AEE (kCal/kg)	10.8 ± 4.8	11.7 ± 4.8	10.5 ± 4.4	12.3 ± 4.5	11.9 ± 4.3	11.8 ± 4.3
TBW (L)	34.7 ± 6.8	35.5 ± 7.1	34.6 ± 7.1	36.6 ± 6.1	36.2 ± 6.2	35.9 ± 6.3

Compliance was calculated as % days with more than 10h of data in reference to days of assessment, BMI=body mass index, mMRC=modified Medical Research Council dyspnoea score, FEV_1=forced expiratory volume in the first second, FVC=forced vital capacity, BODE=BODE index, SGRQt=St George Respiratory Questionnaire total score, SGRQa= St George Respiratory Questionnaire activity score, CAT=COPD assessment test, AEE=activity energy expenditure, TBW=total body water. No statistically significant differences were observed in any of the variables between the different groups wearing different monitors.

Table S2A. Average daily physical activity outcomes per monitor during weekdays.

	Lifecorder N=393			ActiWatch N=402			RT3 N=379			DynaPort N=358			GT3X N=412			SenseWear N=713		
	mean	±	SEM	mean	±	SEM	mean	±	SEM	mean	±	SEM	mean	±	SEM	mean	±	SEM
AS	492	±	12		±			±			±			±			±	
AC		±		201736	±	4445		±			±			±			±	
VMU		±			±		128584	±	4491		±		358821	±	10801		±	
AT (min/day)													248	±	7			
Steps (per day)		±			±			±		5697	±	181	3920	±	167	4280	±	120
TPA (min/day)		±			±			±			±			±		170	±	5
Mint		±			±			±		35.7	±	1		±			±	
WT (min/day)		±			±			±		71	±	2		±			±	
TEE (kCal)	1881	±	17		±		1813	±	29		±			±		1897	±	30
AEE (kCal)		±			±		362	±	13		±			±			±	
MET's		±			±			±			±			±		1.2	±	0.01

N=number of measurements averaged (days x patients). AS=activity score, AC=activity counts, VMU=vector magnitude units, AT=activity time, TPA=time on physical activity, Mint=movement intensity, WT=walking time, TEE=total energy expenditure, AEE=activity energy expenditure, MET's=metabolic equivalents. Note that the population assessed with each monitor is not constituted by the same subjects.

Table S2B. Average daily physical activity outcomes per monitor during weekend.

	Lifecorder N=78			ActiWatch N=79			RT3 N=73			DynaPort N=71			GT3X N=78			SenseWear N=136		
	mean	±	SEM	mean	±	SEM	mean	±	SEM	mean	±	SEM	mean	±	SEM	mean	±	SEM
AS	428	±	24		±			±			±			±			±	
AC		±		177006	±	8575		±			±			±			±	
VMU		±			±		121288	±	8683		±		292385	±	17845		±	
AT (min/day)													216	±	12			
Steps (per day)		±			±			±		4384	±	331	2769	±	288	3426	±	223
TPA (min/day)		±			±			±			±			±		152	±	11
Mint		±			±			±		29	±	2		±			±	
WT (min/day)		±			±			±		56	±	4		±			±	
TEE (kCal)	1869	±	33		±		1897	±	47		±			±		1845	±	65
AEE (kCal)		±			±		341	±	27		±			±			±	
MET's		±			±			±			±			±		1.2	±	0.02

N=number of measurements averaged (days x patients). AS=activity score, AC=activity counts, VMU=vector magnitude units, AT=activity time, TPA=time on physical activity, Mint=movement intensity, WT=walking time, TEE=total energy expenditure, AEE=activity energy expenditure, MET's= metabolic equivalents. Note that the population assessed with each monitor is not constituted by the same subjects.

Table S3. Correlations between physical activity monitors outputs and DLW-measured energy expenditure variables

			TEE	AEE
Lifecorder	**AS**	r		0.45
		p		<0.005
	TEE	r	0.76	
		p	<0.0001	
ActiWatch	**AC**	r		0.68
		p		<0.0001
RT3	**VMU**	r		0.48
		p		<0.01
	TEE	r	0.83	
		p	<0.0001	
	AEE	r		0.56
		p		<0.005
DynaPort	**Steps**	r		0.48
		p		<0.005
	Mint	r		0.70
		p		<0.0001
	WT	r		0.52
		p		<0.005
GT3X	**VMU**	r		0.68
		p		<0.0001
	Steps	r		0.71
		p		<0.001
SenseWear	**Steps**	r		0.38
		p		<0.005
	TEE	r	0.35	
		p	<0.005	
	MET's	r		0.39
		p		<0.005

Correlations (r and p values) between activity variables measured with the six different monitors and DLW measured energy variables. TEE=total energy expenditure, AEE=activity energy expenditure, AS=activity score, AC=activity counts, VMU=vector magnitude units, Mint=movement intensity, WT=walking time, MET's= metabolic equivalents.

Table S4. CORRELATIONS BETWEEN ACTIVITY MONITOR VARIABLES AND EXERCISE TOLERANCE

			Lifecorder	ActiWatch	RT3	DynaPort	GT3X	SenseWear
			AS	AC	VMU	Steps	Steps	Steps
Wpeak	**(W)**	**r**	**0.25**	**0.64**	**0.84**	**0.47**	**0.80**	**0.71**
		p	**ns**	**<0.0001**	**<0.0001**	**<0.005**	**<0.0001**	**<0.0001**
VO_2peak	**(ml/min/kg)**	**r**	**0.35**	**0.74**	**0.76**	**0.36**	**0.85**	**0.65**
		p	**<0.05**	**<0.0001**	**<0.0001**	**<0.05**	**<0.0001**	**<0.0001**
6MWD	**(m)**	**r**	**0.28**	**0.57**	**0.58**	**0.50**	**0.80**	**0.60**
		p	**ns**	**<0.0001**	**<0.0001**	**<0.001**	**<0.0001**	**<0.0001**

Best correlations (r and p values) between activity variables measured with the six different monitors and exercise tolerance variables. Wpeak=peak power achieved in the incremental cardiopulmonary exercise test on a cycloergometer in Watts, VO_2peak= peak oxygen consumption achieved in the incremental cardiopulmonary exercise test on a cycloergometer in ml per minute per kg of weight, 6MWD=distance walked in the six minutes walking test in meters, TEE= total energy expenditure, kCal=Kilocalories, AS=activity score, AC=activity counts, VMU=vector magnitude units, Walking MI=walking movement intensity.

Table S5. INTER-DEVICE RELIABILITY.

A

		ActiWatch	RT3	Dynaport			GT3X			SenseWear	
		AC	VMU	Steps	MInt	WT	VMU	AT	Steps	TPA	Steps
Lifecorder	AS	**0.6**	**0.56**	0.43	**0.84**	**0.54**				0.2	**0.47**
ActiWatch	AC		**0.7**	**0.61**	**0.79**	**0.6**	**0.72**	**0.73**	**0.63**	**0.38**	**0.51**
RT3	VMU						**0.66**	**0.58**	**0.89**	0.25	**0.52**
DynaPort	Steps						**0.75**	**0.53**	**0.91**	**0.59**	**0.84**
	Mint						**0.81**	**0.78**	**0.92**	**0.56**	**0.80**
	WT						**0.72**	**0.57**	**0.91**	**0.63**	**0.80**

B

		RT3		DynaPort	SenseWear	
		TEE	AEE	TEE	TEE	METs
Lifecorder	TEE	**0.87**	**0.63**	**0.94**	0.24	0.14
RT3	TEE		**0.77**		0.29	0.13
	AEE				0.08	0.27
DynaPort	TEE				**0.32**	0.11

Correlations between different outcomes for the six monitors explored. Table 3A variables ***measured*** by the physical activity monitors. No patient wore simultaneously the Lifecorder Plus and the Actigraph GT3X monitor. Table 3B, variables ***estimated*** by the physical activity monitors.
No patient wore simultaneously the RT3 and the DynaPort MoveMonitor. AC=activity counts, VMU=vector magnitude units, Mint=movement intensity, WT=walking time, AT=activity time, TPA=time on physical activity, AS=activity score, TEE=total energy expenditure, AEE=activity energy expenditure, MET's=metabolic equivalents. Black numbers represent statistical significant correlations. Non statistically significant correlations are represented by grey numbers.

Figure S1. ABILITY OF PHYSICAL ACTIVITY MONITORS TO ESTIMATE TEE

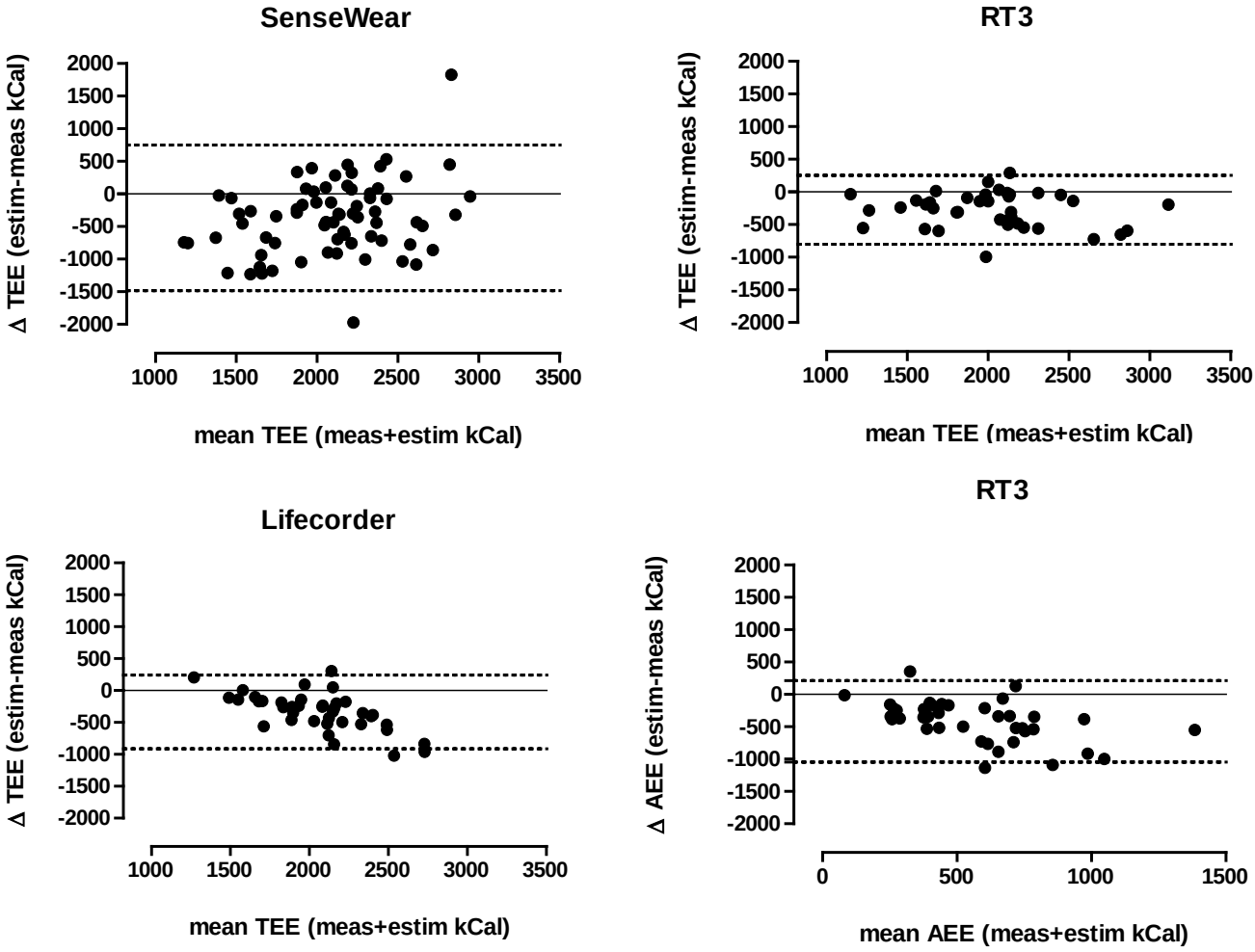

Figure S1. Bland and Altman plot using SenseWear Armband, RT3 and Lifecorder output (kilo calories [kCal]) to estimate total energy expenditure (TEE) (kCal) and activity energy expenditure (AEE) for the RT3 monitor. The dashed lines represent the 95% confidence intervals of the observations. Each dot represents each individual subject using the monitor. Data correspond to the average kCal of all days of assessment for each monitor and DLW measured daily energy expenditure (TEE and AEE).

Figure S2. REPEATABILITY OF PHYSICAL ACTIVITY ASSESSMENT IN TWO CONSECUTIVE WEEKS

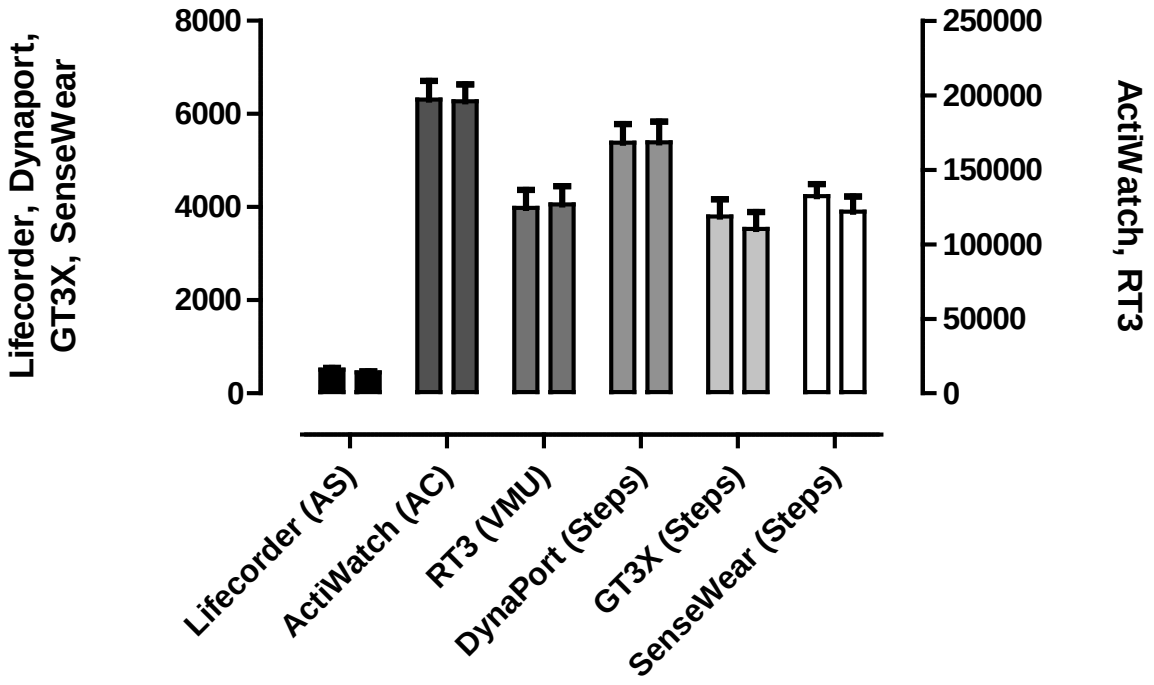

Figure S2. Repeatability of physical activity outcomes in two consecutive weeks for the six explored monitors: Lifecorder plus, ActiWatch Spectrum, RT3, DynaPort MoveMonitor, Actigraph GT3X, SenseWear Armband. Each colour represents each of the monitors. The first column represents the first week while the second column represents the second week. No statistical difference between the first and the second week was found for any of the variables for each monitor. AS=activity score, AC=activity counts, VMU=vector magnitude units.

Figure S2. IDENTITY PLOTS OF PHYSICAL ACTIVITY ASSESSMENT IN TWO CONSECUTIVE WEEKS

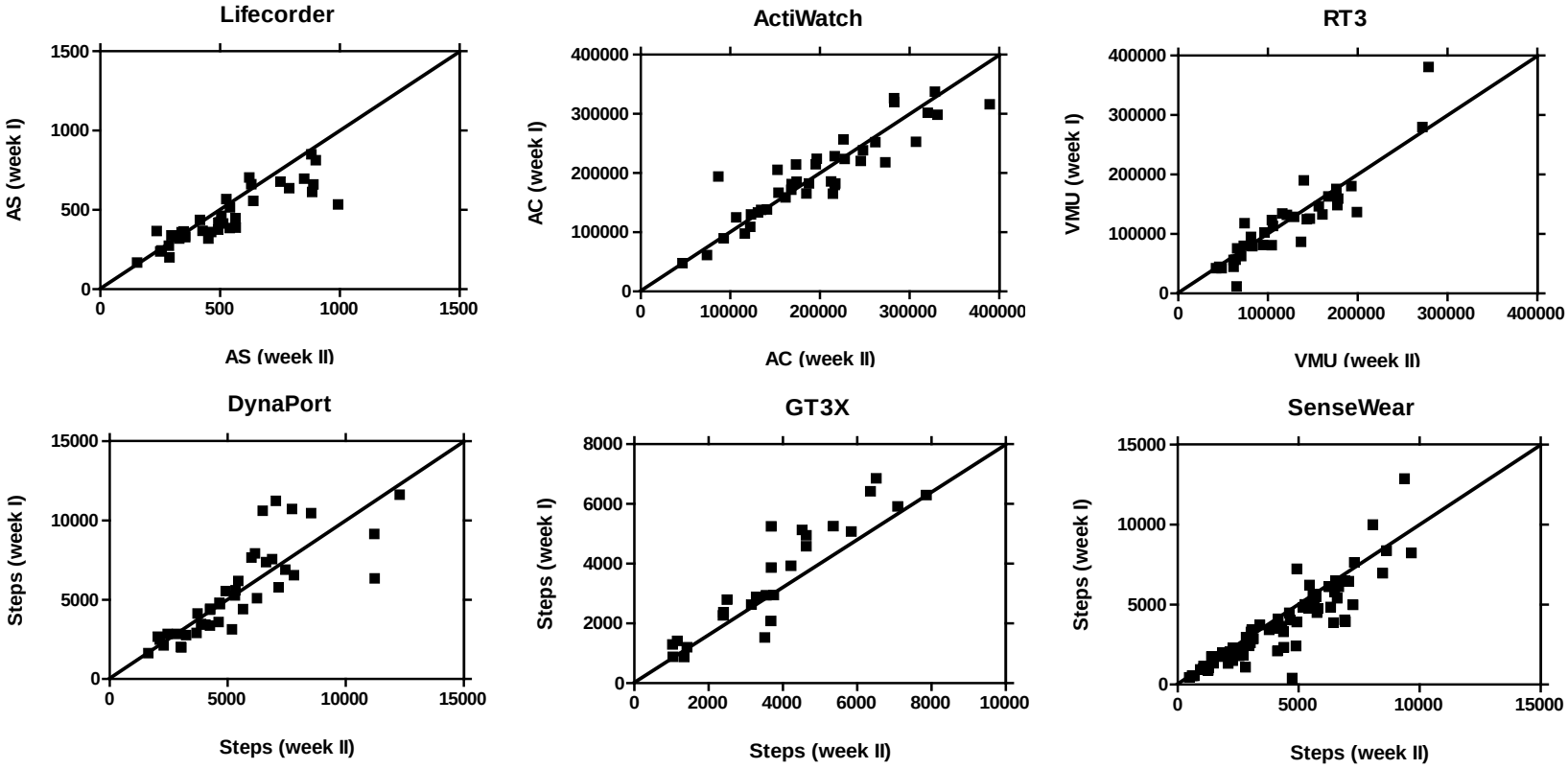

Figure S3. Identity plots of physical activity outputs between two consecutive weeks of assessment for the six explored monitors: Lifecorder plus, ActiWatch Spectrum, RT3, DynaPort MoveMonitor, Actigraph GT3X, SenseWear Armband. AS=activity score, AC=activity counts, VMU=vector magnitude units

REFERENCES

1. Rabe KF, Hurd S, Anzueto A, Barnes PJ, Buist SA, Calverley P, et al. Global strategy for the diagnosis, management, and prevention of chronic obstructive pulmonary disease: GOLD executive summary. *Am J Respir Crit Care Med*. 2007;176(6):532-55. Epub 2007/05/18.

2. Miller MR, Hankinson J, Brusasco V, Burgos F, Casaburi R, Coates A, et al. Standardisation of spirometry. *Eur Respir J*. 2005;26(2):319-38. Epub 2005/08/02.

3. ATS statement: guidelines for the six-minute walk test. *American Journal of Respiratory and Critical Care Medicine*. 2002;166(1):111-7.

4. Jones PW, Quirk FH, Baveystock CM, Littlejohns P. A Self-complete Measure of Health Status for Chronic Airflow Limitation. *American Review of Respiratory Disease*. 1992;145:1321-7.

5. Jones PW, Harding G, Berry P, Wiklund I, Chen WH, Kline Leidy N. Development and first validation of the COPD Assessment Test. *Eur Respir J*. 2009;34(3):648-54. Epub 2009/09/02.

6. British Standard Institution. Precision of test methods 1: Guide for the determination and reproducibility for a standard test method (BS 597, Part 1). *London: (BSI)*. 1975.

EXAMPLE OF THE QUESTIONAIRE DESGNATED TO EVALUATE USABILITY OF THE ACTIVITY MONITORS

PROactive questionnaire on user friendliness of activity monitors

In the PROactive project we aim at using an activity monitor in patients with chronic obstructive pulmonary disease. Activity monitors measure the amount and intensity of activities you carry out in your daily life. Since patients will be asked to wear these monitors in future clinical studies, we highly value your opinion on the monitors we tested. Given that you did wear several of these monitors for a couple of days, we believe you are best placed to evaluate the positive and negative aspects of each monitor. Your experiences and opinions will help us to choose the most convenient and most user friendly monitor.

Please respond to the questions below. There are no correct or incorrect answers. We are interested in your honest opinion and value any comments. Thank you very much in advance for your collaboration!

THESE QUESTIONS RELATE TO THE "**ACTI***WATCH* SPECTRUM" MONITOR:

Section A

I experienced technical problems with the Monitor

- O All the time
- O Frequently
- O Sometimes
- O Seldom
- O Never

The Monitor interfered with my normal activities

- O All the time
- O Frequently
- O Sometimes
- O Occasionally
- O Never

I felt comfortable wearing the monitor

- O All the time
- O Frequently
- O Sometimes
- O Occasionally
- O Never

I felt embarrassed wearing the Monitor

- O All the time
- O Frequently
- O Sometimes
- O Occasionally
- O Never

The instruction on how to use the Monitor were clear

- O Strongly disagree
- O Disagree
- O Neutral
- O Agree
- O Strongly agree

Using the monitor on a daily basis was easy

- O Strongly disagree
- O Disagree
- O Neutral
- O Agree
- O Strongly agree

How much trouble did you have getting started with the monitor?

- O No trouble to start up
- O Sometimes trouble to start up
- O Regularly caused trouble to start up
- O Always trouble to start up
- O I had to call the centre to get help in starting-up

The Monitor was easy to put on/take off

- O Yes this was very easy
- O This worked just fine
- O I found it somewhat difficult
- O I found it difficult
- O I was unable to manage this on my own

The Monitor was bulky/heavy.

- O Yes , very much so
- O Yes much
- O Not particularly
- O Not at all
- O No opinion

The Monitor A bothered me in the bed.

- O Yes , very much so
- O Yes much
- O Not particularly
- O Not at all
- O No opinion, I did not wear the monitor at night

I felt my privacy was invaded by the Monitor

- O Strongly disagree
- O Disagree
- O Neutral
- O Agree
- O Strongly agree

If my doctor would like to use the Monitor to assess my physical activity I would be willing to wear the monitor for

- O Less than 1 day
- O 2-4 days
- O 1 week
- O Longer than 1 week.
- O I would not mind wearing the monitor continuously (longer then 1 month)

We want to ask you to provide a final score for the Monitor. All things considered can you give the monitor a score from 0% to 100%, where 0 means the worst possible monitor, and 100% means the ideal monitor in your opinion.
My score for MONITOR : ……/100

In the section below we want to give you the opportunity to give other comments on the Monitor.

I experienced the following problems with the Monitor:
……
……
……
……
……
……
……
……
……
……
……
……
I liked these features of the monitor in particular:
……
……
……
……
……
……
……
……
……
……
……

……

……

……

……