Are Self-report Measures Able to Define Individuals as Physically Active or Inactive?

Jostein Steene-Johannessen[1], **Sigmund A. Anderssen**[1], **Hidde P. van der Ploeg**[2], **Ingrid J.M. Hendriksen**[3,4], **Alan E. Donnelly**[5], **Søren Brage**[6], and **Ulf Ekelund**[1]

[1]Department of Sports Medicine, Norwegian School of Sport Sciences, Oslo, Norway [2]VU University Medical Center, Department of Public and Occupational Health, EMGO Institute for Health and Care Research, Amsterdam, The Netherlands [3]TNO Expertise Centre Lifestyle, Leiden, The Netherlands [4]Body@Work TNO-VUmc, Research Center Physical Activity, Work and Health, Amsterdam, The Netherlands [5]Department of Physical Education and Sport Sciences, University of Limerick, Limerick, Ireland [6]Medical Research Council Epidemiology Unit, University of Cambridge, Cambridge, United Kingdom

Abstract

Purpose—Assess the agreement between commonly used self-report methods compared with objectively measured physical activity (PA) in defining the prevalence of individuals compliant with PA recommendations.

Methods—Time spent in moderate and vigorous PA (MVPA) was measured at two time points in 1713 healthy individuals from 9 European countries using individually-calibrated combined heart-rate and movement sensing. Participants also completed the Recent Physical Activity Questionnaire (RPAQ), the short form of the International Physical Activity Questionnaire (IPAQ) and the short European Prospective Investigation into Cancer and Nutrition Physical Activity Questionnaire (EPIC-PAQ). Individuals were categorized as active (e.g. reporting 150 min of MVPA per week) or inactive, based on the information derived from the different measures. Sensitivity and specificity analyses and Kappa statistics were performed to evaluate the three PAQs ability to correctly categorize individuals as active or inactive.

Results—Prevalence estimates of being sufficiently active varied significantly (p for all <0.001) between self-report measures (IPAQ 84.2 % [95% CI 82.5 to 85.9], RPAQ 87.6% [95% CI 85.9 to 89.1], EPIC-PAQ 39.9% [95% CI 37.5 to 42.1] and objective measure 48.5% [95% CI 41.6 to 50.9]. All self-report methods showed low or moderate sensitivity (IPAQ 20.0 %, RPAQ 18.7% and EPIC-PAQ 69.8%) to correctly classify inactive people and the agreement between objective and self-reported PA was low (κ =0.07 [95% CI 0.02 to 0.12], 0.12 [95% CI 0.06 to 0.18] and 0.19 [95% CI 0.13 to 0.24] for IPAQ, RPAQ, and EPIC-PAQ, respectively).

Corresponding author: Jostein Steene-Johannessen, Department of Sports Medicine, Norwegian School of Sport Sciences, P.O Box 4014 Ullevål Stadion, 0806 Oslo, jostein.steene-johannessen@nih.no, Phone/Fax: +47 23262295/ +47 22 23 42 20.

The authors have no conflict of interest to report.

Conclusion—The modest agreement between self-reported and objectively measured PA suggests that population levels of PA derived from self-report should be interpreted cautiously. Implementation of objective measures in large-scale cohort studies and surveillance systems is recommended.

Keywords

Physical activity; Sensitivity; Self-report; Questionnaire; Accelerometry

Introduction

Physical activity (PA) is one of the leading risk factors for non-communicable diseases and it has been suggested that physical inactivity is one of the greatest public health problem in the 21st century (37). Thus, increasing PA has been proposed as an important public health strategy. Many health authorities worldwide recommend that adults should engage in moderate-to vigorous PA (MVPA) for at least 150 minutes per week (9, 26, 38). Although total PA energy expenditure (PAEE) and time spent in different intensities are important dimensions when measuring PA in relation to health outcomes, accurate measures of the proportion of the population meeting above-mentioned recommendations is fundamental for public health policy and informing intervention strategies.

For practical reasons, PA questionnaires (PAQs) are the most commonly used assessment method in large population based cohort studies and surveillance systems. Based on data from 122 countries, 31.1 % of adults worldwide were estimated to be inactive (e.g. not meeting PA recommendations), with substantial between-country variation (19), and data from Europe (5) reveal large variation in population levels of sitting even across industrialised countries, with levels ranging from 191 to 407 minutes/day. However, there is a paucity of data documenting the agreement between different PAQs for correctly classifying individuals as physically active (e.g. meeting the adult PA recommendations of > 150 min of MVPA per week) compared with an objective criterion method.

Frequently used PAQs include the International Physical Activity Questionnaire (IPAQ) (11) which has been used in large scale population surveys such as the Eurobarometer(27) and the World Health Organization (WHO) world health survey (17), the Recent Physical Activity Questionnaire (RPAQ) which is used in the Fenland study and National Diet European Prospective Investigation into Cancer and Nutrition Survey in the UK (6, 13) and the short EPIC Physical Activity Questionnaire (EPIC-PAQ) (36) used in one of the largest pan-European cohorts including approximately 520.000 individuals (29). These three PAQs have been extensively validated and the overall results reveal validity for ranking individuals and group level assessment of PAEE, MVPA and sedentary time (11, 16, 23). However, reported associations with criterion measures rarely exceed correlations of 0.3 (28), and, although significant, should be interpret as low to moderate. In addition, there are several well-known limitations with self-report, especially with regards to cultural differences, recall bias and misinterpretation of questions (10). Over-reporting of PA appears to be a measurement issue with respect to IPAQ (15, 25, 31) and a recent validation of RPAQ revealed higher estimates of PAEE and MVPA compared with an objective criterion measure

(16). Taken together, this might result in an overestimation of the proportion of respondents being categorised as sufficiently physical active when PA is assessed with self-report.

Accurate data on population levels of PA are required for policymakers and researchers, in order to be able to answer fundamental public health questions by means of exploring trends in PA behaviour both within and across countries, to evaluate effect of different initiatives and to be able to reach target populations. Thus, based on the large number of data available in European adults using different PAQs the main aim of the current study was to assess the agreement between different commonly used self-report methods and compared with objectively measured PA (criterion method) in defining the prevalence of individuals compliant with PA recommendations.

Methods

Details of the study population and study design have been published elsewhere (16, 23). In short, a convenient sample of healthy individuals from 10 European countries was recruited based on a centre-specific age and gender distribution similar to the original EPIC-Europe cohort (29). Approximately 200 men and women were recruited from each country [Denmark (n= 177); France (n=171) Germany (n=208), Greece (n=182), Italy (n=195), Netherlands (n=206), Norway (n=176), Spain (n=202), Sweden (n=194) and UK (n=196)]. In France and Norway, only women were included. In addition, only those individuals comprising complete data on objective measured PA and all three PAQs are included in the analyses. In Umeå (Sweden) the original EPIC-PAQ was not used and consequently all participants from Sweden were excluded from the present analyses. The final study population therefore included 1713 participants. The study consisted of two visits held four to five months apart (mean time between visits = 4.5 months; SD = 1.0). Height, weight and free-living PA were measured at both visits according to standard procedures (23) with the additional administration of IPAQ, RPAQ and EPIC-PAQ at the second visit. For standardisation and quality control across centres, MRC Epidemiology Unit staff organised a workshop prior to testing and visited study centres during the testing phase. Each centre obtained ethical approval from a local ethics board prior to participant recruitment, and informed consent was obtained from all participants.

Objective PA measurement

PA was objectively measured using a combined heart rate (HR) and movement sensor (Actiheart, CamNtech Ltd, Cambridge, UK) attached to the chest via standard ECG electrodes. All participants performed an eight-minute submaximal ramped step test (200-mm step; Reebok, Lancaster, UK) to determine the individual relationship between HR and workload (8). Following the step test, the Actiheart sensor was reinitialized to collect data in one minute epochs and the participants were instructed to wear the monitor constantly (24 hours/day) for at least four consecutive days with a mean wear time for both measuring periods of 4.7 (1.0) days. We excluded all participants with less than 3 days of wear data. Furthermore, MVPA was averaged based on the mean of each of the two four-day measurements.

PA intensity ($J \cdot min^{-1} \cdot kg^{-1}$) for each time point was estimated from the combination of movement registration and individually calibrated HR (8) using a branched equation framework (7). To handle potential measurement noise in the heart rate trace HR data from free-living was pre-processed (32) and identification of non-wear periods from the combination of non-physiological heart rate and prolonged periods (> 60 min) of inactivity were performed. This method yields quantification of uncertainty (error bars on the estimate of latent heart rate at any given time) which is heavily influenced by the randomness of the measurement (which is non-physiological over short time-scales) and we simply use the size of this uncertainty in combination with prolonged periods of no-movement to decide if it is likely that the person was wearing the monitor or not The threshold for MVPA was set at >3.5 MET, with 1 MET based on the standard definition (32) of 1 MET = 3.5 $mlO_2 \cdot min^{-1} \cdot kg^{-1}$. This threshold was chosen to match METS values derived for reported time spent walking from IPAQ (3.3 METS) and RPAQ (3.5 METS). For the purpose of defining the individuals as inactive or active, the threshold for meeting PA recommendations was defined as achieving at least 150 $min \cdot week^{-1}$ of moderate-to-vigorous activity (> 3.5 MET). We included all time above this level, without any stipulation of bouts.

PA questionnaires

The three different PAQs were electronically administered as previously described (6, 11, 16, 23). Prior to the objective monitoring of PA, each participant completed the short EPIC-PAQ, the short version of the IPAQ and the RPAQ. All PAQs were translated from English to each of the specific languages and then back-translated before administration (23).

In short, EPIC-PAQ is comprised of four questions related to PA during the last year. The four questions cover; 1) category of occupational activity, 2) participation in several activities (walking, cycling, do-it-yourself, gardening, sports and household chores) during both summer and winter, 3) participation in vigorous non-occupational activity and 4) number of floors stair flights climbed per day. Based on the information on occupational category and time spent in sports and cycling we derived the Cambridge Index (36). To define individuals as active or inactive we collapsed these categories into either *not meeting PA recommendations* (inactive and moderately inactive) or *meeting PA recommendations* (moderately active and active). Participants categorized as inactive reported an inactive occupation and no leisure time PA and those categorized as moderately inactive reported both a sedentary occupation and less than 3.5 $hours \cdot week^{-1}$ of moderate to vigorous intensity leisure time PA or a standing occupation in combination with no leisure time PA.

RPAQ consists of 9 different questions referring to the last 4 weeks. The RPAQ is comprised of closed questions and with ordered categories of frequencies paired with duration. RPAQ covers 4 domains of PA; domestic life, work, recreation and transport. The domestic section includes questions about TV-viewing, computer use and stair climbing. The Modified Tecumseh Occupational Activity Questionnaire (1) was adopted for deriving the occupational categories of PA (mostly sitting, standing, manual or heavy manual). Questions about recreational PA were adopted from The Minnesota Leisure Time Activity Questionnaire (30), and transport related PA was categorised as walking, cycling, and use of car/public transport. To estimate time spent in different intensities all activities were

categorised as follows; sedentary (<1.5 MET); light (1.5 to <3 MET) and MVPA (>3 MET). Based on time (min·week^{-1}) spent >3 MET those individuals achieving at least 150 min·week^{-1} of MVPA were categorized as meeting PA recommendations. The majority of recreation activities accessible to report were referring to MET-scores above 3.5 METS.

The short, last 7-days IPAQ asks the respondents to report time (i.e. number of sessions and average time per session) spent walking (3.3 METS), in moderate intensity PA (4.0 METS), in vigorous-intensity PA (8 METS) and sitting (<1.5 METS) (only weekdays). Questions regarding intensity were supplemented by examples of commonly performed activities. Based on the information within each intensity category we estimated the total amount of time spent in PA per week. MVPA (min·week^{-1}) was estimated by summing the reported time spent walking (3.3 METS) and in MVPA (>4 METS) intensity and then categorized those individuals achieving at least 150 min·week^{-1} of MVPA as meeting the PA recommendations.

Statistics

Values in tables are presented as mean ± SD, unless otherwise stated. Differences in participants' age and BMI were assessed by independent-samples t-test. Based on data from each PAQ differences were assessed between proportions of individuals meeting PA recommendations using chi-square analyses. Sensitivity and specificity analyses with 95% confidence intervals were performed to evaluate the PAQ's ability to correctly categorize individuals as active or inactive using combined heart rate and motion sensing as criterion method. Specificity is the PAQ's ability to correctly identify an individual as physically active whereas the sensitivity refers to the ability to correctly identify individuals as physically inactive. Kappa statistics was used to evaluate the level of agreement between PAQs and the criterion method in defining prevalence of compliance with recommendations. Random effect meta-analyses were used to calculate the combined agreement across countries. Heterogeneity across countries in the agreement of each PAQ was evaluated by Forest plots and assessed using I-squared (I^2) statistics. Kappa correlations coefficients of 0.81-1.00 are generally interpreted as very good, 0.61-0.80 as good, 0.41-0.60 as moderate, 0.21-0.40 as fair, and <0.20 as poor (2). Finally, Pearson's correlation coefficient (r) was used to evaluate the relationship (on country level) of the proportion meeting the PA recommendations from each of the three self-report instruments with the objective criterion measure.

All analyses were performed in IBM SPSS statistics version 21 except for the random effect meta-analyses, which was performed using STATA version 13.0 (StataCorp, College Station, TX). Threshold for significance was set at p< 0.05.

Results

The baseline characteristics stratified by study location are shown in table 1. Across all locations the average age was 54.7 (SD 9.5) years and the average BMI was 25.7 (SD 4.0) kg·m^{-2}. The majority of the study population were women (72 %) and they were younger (p= 0.009) and leaner (p< 0.001) compared with men. Mean age varied across countries ranging from 48.4 to 61.5 years. BMI also varied across countries, with Greek women and

men having the highest BMI (27.0 and 27.7 kg·m^{-2}, respectively) whereas the women and men from the Netherlands had the lowest BMI (22.6 and 23.5 kg·m^{-2}, respectively).

Table 2 displays the prevalence of of participants meeting PA recommendations according to different PAQs and the objective measure by country, sex and overall. The results revealed substantial discrepancies in prevalence estimates from the three PAQs. Overall estimations based on IPAQ and RPAQ suggested that more than 4 out of 5 participants were categorized as sufficiently active, whereas the proportion of participants categorized as sufficiently active was 39.9% when estimated from the short EPIC-PAQ and 48.5% based on the objective measure. There were no sex differences in the proportions meeting PA recommendations based on IPAQ (p=0.991) and objectively measured PA (p=0.098), whereas a lower proportion of women were meeting PA recommendations based on RPAQ (p=0.015) and EPIC-PAQ (p=0.003). A similar pattern was found within all countries revealing that the prevalence of being categorized as active was highest when based on results from RPAQ and IPAQ, whereas prevalence estimates were consistently lower when based on EPIC-PAQ or the objective measure. However, comparing prevalence estimates within each PA measure between countries the results revealed significant differences (p for all < 0.001), with prevalence of meeting PA recommendations based on EPIC-PAQ ranging from 17.6% in Greece to 54% in Norway. For both RPAQ and IPAQ respectively, participants from Greece had the lowest prevalence (74.2 % and 67.0%) whereas the Netherlands showed the highest prevalence (99% and 94.7%). Finally, based the objective measure, 24.5% of UK participants and 65.5% of the participants from the Netherlands where categorized as meeting PA recommendations. Figure 1 a, b and c reveals no significant correlations for the self-report instruments with the objective measure in ranking country levels of proportion meeting PA recommendations (IPAQ r = 0.28 [p= 0.47], RPAQ r = 0.35 [p= 0.36] and EPIC-PAQ r = 0.41 [p= 0.27]).

Using objectively measured PA as the criterion method we found good specificity but poor sensitivity for IPAQ and RPAQ when evaluating the ability of capturing participants meeting PA recommendations (specificity) or not meeting PA recommendations (sensitivity) (table 3). In detail, the overall specificity for IPAQ to capture sufficiently active individuals was 88.7% whereas only 20.0% were correctly captured as insufficiently active. A similar pattern was found for RPAQ where specificity and sensitivity were 94.2% and 18.8%, respectively. For EPIC-PAQ results revealed slightly lower specificity (50.2%) but better sensitivity (69.8%). Furthermore, table 3 provides a more detailed description of the sensitivity and specificity by sex and country, revealing large differences between countries.

The level of agreement between all three PAQs and the objective measure in defining the proportion of compliance with recommendations were poor. The pooled estimates of agreement (Figure 2) revealed κ - values of 0.07 (95% CI 0.02 to 0.12), and 0.12 (95% CI 0.06 to 0.18) for IPAQ and RPAQ, respectively, whereas the overall agreement was slightly stronger for EPIC-PAQ (κ =0.19 [95% CI 0.13 to 0.24]). We observed significant heterogeneity across countries in agreement for IPAQ (I^2 = 94.9%, P < 0.001), RPAQ (I^2 = 97.5%, P < 0.001) and EPIC-PAQ (I^2 = 91.9%, P < 0.001).

Discussion

We present data on the ability of three commonly used PAQs to assess the prevalence of sufficiently active adults compared with objectively measured PA from nine European countries. Our results demonstrate substantial discrepancies in prevalence estimates of being sufficiently physically active derived from the three PAQs. Moreover, all three self-report methods showed low to moderate sensitivity to correctly classify inactive people and the agreement with the objective PA measure was low. Both IPAQ and RPAQ tended to substantially overestimate the number of people meeting the PA recommendations, whereas EPIC-PAQ underestimated the number of people not meeting the PA recommendations, which was illustrated with lower specificity than that found for IPAQ and RPAQ.

Results from the present study confirm recent findings that the proportion of individuals being categorized as sufficiently active varies substantially between countries (19). Although the same PAQs and objective measure were used in each country there are well known limitations and potential measurement errors in self-reported PA. However, the objective measure used in the present study confirmed differences between countries suggesting that previously reported differences between countries are not entirely explained by differential bias in the PAQs. Thus, it is likely that the present, and to some extent previously reported findings, reflect geographical and cultural differences in overall PA level across European countries.

Nonetheless, our results indicate that the three self-report methods do not match well in ranking population levels of prevalence of individuals meeting PA recommendations. The low to moderate correlation observed on country level (figure 1 a, b and c) suggests that cross country comparison is difficult even if the same self-report instrument is used. For example, the IPAQ, which is used for surveillance purposes in the Eurobarometer, performs poorly when ranking countries according to their proportion of individuals meeting PA recommendations. We can only speculate on why countries differ in how methods disagree, but cultural differences in how people understand and interpret with certain PA questions is likely the most obvious reason. However, use of only nine data points need to be considered when interpreting the present results. Nevertheless, our results indicate that observations from cross-country comparisons using IPAQ and RPAQ should be interpreted with some caution, and ideally objective assessment methods should be used.

Another important question is whether PAQs are able to classify or identify sufficiently active or insufficiently active individuals according to recommendations compared to a criterion method. This information is important when deciding on the best methods available to answer fundamental public health questions such as exploring PA behaviour in a population or evaluating effect of different public health initiatives. The present findings reveal substantial differences in the proportion of individuals classified as meeting the current recommendations for PA between different self-report measures and also when compared to the objective measure (Table 3). These results corroborate to some extent with previous observations from the NHANES where Troiano et al (33) found that less than 5% met the PA recommendations based on accelerometer derived results, whereas 51% met PA recommendations based on self-report questionnaire data. Others have also found limited

ability of IPAQ to classify inactive people when compared to objectively measured PA. Dyrstad et al (14) showed that in a sample of 1751 Norwegian men and women, 67% of participants were categorized as sufficiently active by IPAQ, whereas the corresponding number for accelerometry was 22%. Similarly, Ekelund et al (15) showed that the sensitivity of IPAQ to capture insufficiently active individuals was only 45% in a sample of 187 Swedish adults. Moreover, data from many population-based studies using the IPAQ suggest that approximately three quarters of individuals meet or exceed 150 min-week of MVPA (11, 14, 17). To summarize, the available data suggest that the proportion of individuals categorised as sufficiently active based on IPAQ appears substantially overestimated when compared with objective PA maesures.

The present results also reveal low κ-values suggesting limited agreement between the PAQs and the objective measure in defining individuals as physically active according to proposed PA recommendations. For both IPAQ and RPAQ the sensitivity to identify individuals not meeting PA recommendations was poor, whereas EPIC-PAQ showed somewhat better sensitivity. Thus, suggesting that a simple derived PA index may be superior to the interpretative framework overlaid on IPAQ and RPAQ when used to identify those who are physically inactive.

There are several factors potentially contributing to the observed large discrepancies and poor levels of agreements. First, self-reports are unreliable especially for housework and occupational activity; this may be particularly problematic especially in low- and middle-income countries, where transport, occupational, and housework activities often are mixed with daily life (20). Moreover, social desirability recall bias and cultural differences in perceptions of the meaning of PA could introduce systematic errors that might lead to overestimations of the respondents' PA level assessed by self-report (35). Rzewnicki et al (31) have suggested that a possible problem with IPAQ is that the respondents need to report an average time per day for each activity performed, which increases the likelihood that the respondent refers to the most active day. In addition, respondents have to calculate an average amount of time per day across many activities which might also increase the possibility of over-reporting. Participants in this study reported a daily average of respectively 51, 28 and 17 minutes of walking, MPA and VPA based on IPAQ. Corresponding numbers derived from RPAQ were 98 minutes in MPA and 14 minutes in VPA, whereas values derived from the combined heart rate and movement sensing showed a daily average of 27 minutes in MPA and 1minute in VPA. These substantial differences in estimated time spent in different intensities from different PA measures underscore that over-reporting is a major challenge affecting population prevalence estimates. Although standardized questionnaires (i.e. IPAQ and the Global Physical Activity Questionnaire) have been successfully implemented globally (19), our results suggest that estimates of population levels of PA and differences in these estimates between countries are likely overestimations of actual levels of PA and should be interpreted cautiously.

Second, the specific criteria used to categorise individuals as meeting or not meeting PA recommendations from self-report were somewhat arbitrary. Thus, it cannot completely be ruled out that this might have biased the results. For example, there is no standard method available for deriving prevalence estimates with respect to PA recommendations based on

the RPAQ. Nevertheless, the criteria applied (at least 150 min·week^{-1} of activity >3 MET) are in agreement with the proposed guidelines (26, 38). Moreover, when defining sufficiently active individuals according to the IPAQ the original scoring protocol (www.ipaq.ki.se) was slightly modified. All individuals with a total self-reported PA level >150 MVPA min·week^{-1} were considered physically active in agreement with the data derived from the RPAQ. Thus, our criteria for categorising individuals as sufficiently active, is less strict compared to the original IPAQ scoring protocol. On the other hand, summarising the total amount of activity per week regardless how the accumulated time is distributed across days is in accordance with the latest recommendations in many countries (26, 38). For EPIC-PAQ we adopted exactly the same criteria for defining the prevalence of sufficiently active individuals (i.e. >150 MVPA min·week^{-1}). On the other hand, the Cambridge index is based on categorising individuals into four groups of PA based on occupation and recreational PA. Participants categorised as 'inactive' or 'moderately inactive' were classified as not being sufficiently active according to the 150 MVPA min·week^{-1} threshold. The 'moderately inactive' category is defined as reporting a sedentary occupation in combination with < 3.5 hours of recreational activity which is higher, but as close as possible to the 150 MVPA min·week^{-1} threshold. Further, the Cambridge index appears accurate for ranking individuals according to their PA levels (23) and predict increased risk for mortality (24) suggesting both criterion and face validity. Thus, using this simple derived PA index to assign participants into either active or inactive seems reasonable. MVPA was defined as equivalent to 3.5 MET from our combined heart rate and movement sensing method to closely match the MET-values used to define MVPA from IPAQ and RPAQ. For example, walking is defined to have an intensity of 3.3 METS and 3.5 METs in IPAQ and RPAQ, respectively. Further, self-reported time in MVPA is defined as 3.5 METs and 4 METS in RPAQ and IPAQ (www.ipaq.ki.se), respectively. In sensitivity analyses using 3 METS as defining MVPA from our objective criterion method we observed slight improved, although still poor agreement for IPAQ (κ =0.138 [95% CI 0.085 to 0.191]) and RPAQ (κ =0.192 [95% CI 0.137 to 0.247]), whereas agreement for EPIC-PAQ, was slightly attenuated (κ =0.114 [95% CI 0.083 to 0.145]). Thus, the definition of MVPA from our objective measure did not affect the overall result that agreement for self-report measures are poor in general, even though the relative performance of the three instrument depend on the definition of the criterion MVPA measure.

Finally, reference timeframe differs between all three PAQs. The EPIQ-PAQ refers to PA during the last year, the RPAQ to the last four weeks, whereas, the IPAQ refers to the last seven days. However, the poor agreement with our criterion method which was equal in magnitude and evident for all three PAQs thus unlikely affected by the differences in recall periods. This is because time spent in MVPA from combined heart rate and movement sensing was estimated by the average of two time points 4-5 months apart.

Objective measures have the potential to overcome limitations associated with self-report and accelerometers have been suggested as the minimum standard in epidemiological research (10). There is also recent work showing, at least among adults, that objective measured PA are more strongly correlated with several cardio-metabolic risk markers (i.e. lipids, triglyserides, insulin and glucose) compared to self-report (3). This underscores the impression that devices might measures physiological meaningful activity. In line with this,

population-based surveys and observational cohort studies using objective assessment methods (i.e. accelerometry) have recently been successfully conducted in several countries (4, 12, 18, 21) suggesting that objective measurement of PA is feasible in large scale cohort studies and PA surveillance systems in developed and developing countries.

Thus, it may be timely to increase the efforts to implement objective measures of PA in large-scale surveillance systems. Although great progress has been made in this field, there are still comparability issues due to the variety of monitors and differences in study protocols, data cleaning and data reduction procedures used. One possible solution to overcome comparability issues across brands of accelerometers might be the use of raw acceleration data rather than relying on proprietary activity counts. For example, raw accelerometer output from two different accelerometer brands appears comparable when attached to the same body location (22) suggesting that the output from different brands are comparable when expressed in SI units (i.e. mg).

Nevertheless, it should be acknowledged that measures derived from self-report and objective methods are not equivalent. As recently pointed out in a review by Troiano et al (34) summary measures are often expressed using the same metrics (i.e., PAEE, time spent in different intensities or MET-min·week^{-1}). However, combined HR and movement sensing quantify the acceleration of the trunk in combination with individually calibrated heart rate to estimate PAEE over a short period, whereas self-report instruments attempt to quantify PA based on reported time periods engaged in specific behaviours. Thus, these two methods are in fact measuring different aspects of the concept of physical activity and thereby leading to challenges for direct comparison. Troiano et al (34) also argue that the epidemiological studies that are the basis for the PA recommendations rely on self-report and thereby questioning estimations of proportion sufficiently physically active based on objective measures. On the other hand, few questionnaires are designed to estimate prevalence of PA according to the recommendations. Moreover, the fact that we still are comparing results based on different measures highlights the importance of documenting measurement errors between self-report and objective measures.

Although measured at two visits five months apart, a limitation of our objective method is that it provides only a snapshot (at least one day) of PA and daily variability might have been better captured with a full week of monitoring. Moreover, an accelerometer located on the trunk is likely to underestimate certain activities such as cycling, swimming or upper body movement and single HR monitoring is a less valid measure of energy expenditure during sedentary and light activity (23). However, the combination of HR and movement sensing has the potential to circumvent some of the limitations of the two respective methods. We also acknowledge that our individual calibration could be limited by only using a step test calibration and a treadmill would have been better. In addition, our estimation of time spent in MVPA included an accumulation of all minutes spent above the 3.5 METs threshold and did not consider for example continuous 10-minute bouts which is part of the current PA guidelines for public health (9, 26, 38). Thus, our estimate of time spent in MVPA likely overestimated the proportion of participants meeting PA recommendation from the objective criterion measure; however such an overestimation would only imply an even larger discrepancy between objectively measured and self-reported PA.

Among the strengths of the present study is the large and diverse sample of men and women from 9 different European countries in which data collection procedures and methods were standardised across study locations and PA assessed by combined heart rate and movement sensing at two different time points five months apart.

Conclusion

Our results reveal substantial differences in prevalence estimates between self-reported measures when assessing compliance with PA recommendations compared with an objective criterion measure. The three self-reports do not perform well in ranking country levels of the proportion of individuals meeting PA recommendations. Further, all three self-report methods; the IPAQ, RPAQ and EPIC-PAQ demonstrated low to moderate sensitivity to correctly classify inactive people and the agreement between PA measures was low, suggesting weak relationships between PAQs and the criterion method. Nevertheless, self-report is vital for measuring attitudes, perception of environment, activity types and context. Thus, implementation of a combination of objective and subjective assessment methods in large scale cohort studies and surveillance systems should be a priority in future PA research.

Acknowledgements

References

1. Ainsworth BE, Jacobs DR Jr, Leon AS, Richardson MT, Montoye HJ. Assessment of the accuracy of physical activity questionnaire occupational data. J Occup Med. 1993; 35(10):1017–27. [PubMed: 8271073]

2. Altman D. Some common problems in medical researchPractical statistics for medical research. Altman DG, editorLondon: Chapman and Hall; 1991. 403–409.

3. Atienza AA, Moser RP, Perna F, et al. Self-reported and objectively measured activity related to biomarkers using NHANES. Med Sci Sports Exerc. 2011; 43(5):815–21. [PubMed: 20962693]

4. Baptista F, Santos DA, Silva AM, et al. Prevalence of the Portuguese population attaining sufficient physical activity. Med Sci Sports Exerc. 2012; 44(3):466–73. [PubMed: 21844823]

5. Bennie JA, Chau JY, van der Ploeg HP, Stamatakis E, Do A, Bauman A. The prevalence and correlates of sitting in European adults - a comparison of 32 Eurobarometer-participating countries. Int J Behav Nutr Phys Act. 2013; 10:107. [PubMed: 24020702]

6. Besson H, Brage S, Jakes RW, Ekelund U, Wareham NJ. Estimating physical activity energy expenditure, sedentary time, and physical activity intensity by self-report in adults. Am J Clin Nutr. 2010; 91(1):106–14. [PubMed: 19889820]

7. Brage S, Brage N, Franks PW, et al. Branched equation modeling of simultaneous accelerometry and heart rate monitoring improves estimate of directly measured physical activity energy expenditure. J Appl Physiol. 2004; 96(1):343–51. [PubMed: 12972441]

8. Brage S, Ekelund U, Brage N, et al. Hierarchy of individual calibration levels for heart rate and accelerometry to measure physical activity. J Appl Physiol. 2007; 103(2):682–92. [PubMed: 17463305]

9. Bull FC, the Expert Working Groups. Physical Activity Guidelines in the U.K.: Review and Recommendations. School of Sport, Exercise and Health Sciences, Loughborough University; 2010 May.

10. Corder K, van Sluijs EM. Invited commentary: comparing physical activity across countries - current strengths and weaknesses. Am J Epidemiol. 2010; 171(10):1065–8. [PubMed: 20406761]

11. Craig CL, Marshall AL, Sjostrom M, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc. 2003; 35(8):1381–95. [PubMed: 12900694]

12. da Silva IC, van Hees VT, Ramires VV, et al. Physical activity levels in three Brazilian birth cohorts as assessed with raw triaxial wrist accelerometry. Int J Epidemiol. 2014; 43(6):1959–1968. [PubMed: 25361583]

13. Day N, Oakes S, Luben R, et al. EPIC-Norfolk: study design and characteristics of the cohort. European Prospective Investigation of Cancer. Br J Cancer. 1999; 80(Suppl 1):95–103. [PubMed: 10466767]

14. Dyrstad SM, Hansen BH, Holme IM, Anderssen SA. Comparison of self-reported versus accelerometer-measured physical activity. Med Sci Sports Exerc. 2014; 46(1):99–106. [PubMed: 23793232]

15. Ekelund U, Sepp H, Brage S, et al. Criterion-related validity of the last 7-day, short form of the International Physical Activity Questionnaire in Swedish adults. Public Health Nutr. 2006; 9(2):258–65. [PubMed: 16571181]

16. Golubic R, May AM, Benjaminsen Borch K, et al. Validity of electronically administered Recent Physical Activity Questionnaire (RPAQ) in ten European countries. PLoS One. 2014; 9(3):e92829. [PubMed: 24667343]

17. Guthold R, Ono T, Strong KL, Chatterji S, Morabia A. Worldwide variability in physical inactivity a 51-country survey. Am J Prev Med. 2008; 34(6):486–94. [PubMed: 18471584]

18. Hagstromer M, Troiano RP, Sjostrom M, Berrigan D. Levels and patterns of objectively assessed physical activity--a comparison between Sweden and the United States. Am J Epidemiol. 2010; 171(10):1055–64. [PubMed: 20406758]

19. Hallal PC, Andersen LB, Bull FC, et al. Global physical activity levels: surveillance progress, pitfalls, and prospects. Lancet. 2012; 380(9838):247–57. [PubMed: 22818937]

20. Hallal PC, Gomez LF, Parra DC, et al. Lessons learned after 10 years of IPAQ use in Brazil and Colombia. J Phys Act Health. 2010; 7(Suppl 2):S259–64. [PubMed: 20702914]

21. Hansen BH, Kolle E, Dyrstad SM, Holme I, Anderssen SA. Accelerometer-determined physical activity in adults and older people. Med Sci Sports Exerc. 2012; 44(2):266–72. [PubMed: 21796052]

22. Hildebrand M, VANH VT, Hansen BH, Ekelund U. Age group comparability of raw accelerometer output from wrist- and hip-worn monitors. Med Sci Sports Exerc. 2014; 46(9):1816–24. [PubMed: 24887173]

23. InterAct C. Validity of a short questionnaire to assess physical activity in 10 European countries. Eur J Epidemiol. 2012; 27(1):15–25. [PubMed: 22089423]

24. Khaw KT, Jakes R, Bingham S, et al. Work and leisure time physical activity assessed using a simple, pragmatic, validated questionnaire and incident cardiovascular disease and all-cause mortality in men and women: The European Prospective Investigation into Cancer in Norfolk prospective population study. Int J Epidemiol. 2006; 35(4):1034–43. [PubMed: 16709620]

25. Lee PH, Macfarlane DJ, Lam TH, Stewart SM. Validity of the International Physical Activity Questionnaire Short Form (IPAQ-SF): a systematic review. Int J Behav Nutr Phys Act. 2011; 8:115. [PubMed: 22018588]

26. Nordic Council of Ministers. Nordic nutrition recommendations 2012. Copenhagen: Nordic Council of Ministers; 2014. 627

27. Papacostas A. Eurobarometer 64.3: Foreign Languages, Biotechnology, Organized Crime, and Health Items, November-December 2005. Inter-university Consortium for Political and Social Research (ICPSR). 2010; [distributor]. doi: 10.3886/ICPSR04590.v3

28. Prince SA, Adamo KB, Hamel ME, Hardt J, Connor Gorber S, Tremblay M. A comparison of direct versus self-report measures for assessing physical activity in adults: a systematic review. Int J Behav Nutr Phys Act. 2008; 5:56. [PubMed: 18990237]

29. Riboli E, Hunt KJ, Slimani N, et al. European prospective investigation into cancer and nutrition (EPIC): study populations and data collection. Public Health Nutr. 2002; 5(6B):1113–24. [PubMed: 12639222]

30. Richardson MT, Leon AS, Jacobs DR Jr, Ainsworth BE, Serfass R. Comprehensive evaluation of the Minnesota Leisure Time Physical Activity Questionnaire. J Clin Epidemiol. 1994; 47(3):271–81. [PubMed: 8138837]

31. Rzewnicki R, Vanden Auweele Y, De Bourdeaudhuij I. Addressing overreporting on the International Physical Activity Questionnaire (IPAQ) telephone survey with a population sample. Public Health Nutr. 2003; 6(3):299–305. [PubMed: 12740079]

32. Stegle O, Fallert SV, MacKay DJ, Brage S. Gaussian process robust regression for noisy heart rate data. IEEE Trans Biomed Eng. 2008; 55(9):2143–51. [PubMed: 18713683]

33. Troiano RP, Berrigan D, Dodd KW, Masse LC, Tilert T, McDowell M. Physical activity in the United States measured by accelerometer. Med Sci Sports Exerc. 2008; 40(1):181–8. [PubMed: 18091006]

34. Troiano RP, McClain JJ, Brychta RJ, Chen KY. Evolution of accelerometer methods for physical activity research. Br J Sports Med. 2014; 48(13):1019–23. [PubMed: 24782483]

35. Tucker JM, Welk GJ, Beyler NK. Physical activity in U.S.: adults compliance with the Physical Activity Guidelines for Americans. Am J Prev Med. 2011; 40(4):454–61. [PubMed: 21406280]

36. Wareham NJ, Jakes RW, Rennie KL, et al. Validity and repeatability of a simple index derived from the short physical activity questionnaire used in the European Prospective Investigation into Cancer and Nutrition (EPIC) study. Public Health Nutr. 2003; 6(4):407–13. [PubMed: 12795830]

37. World Health Organization. Global health risks: mortality and burden of disease attributable to selected major risks. Geneva: World Health Organization; 2009. 62

38. World Health Organization. Global recommendations on physical activity for health. Geneva: World Health Organization; 2010. 58
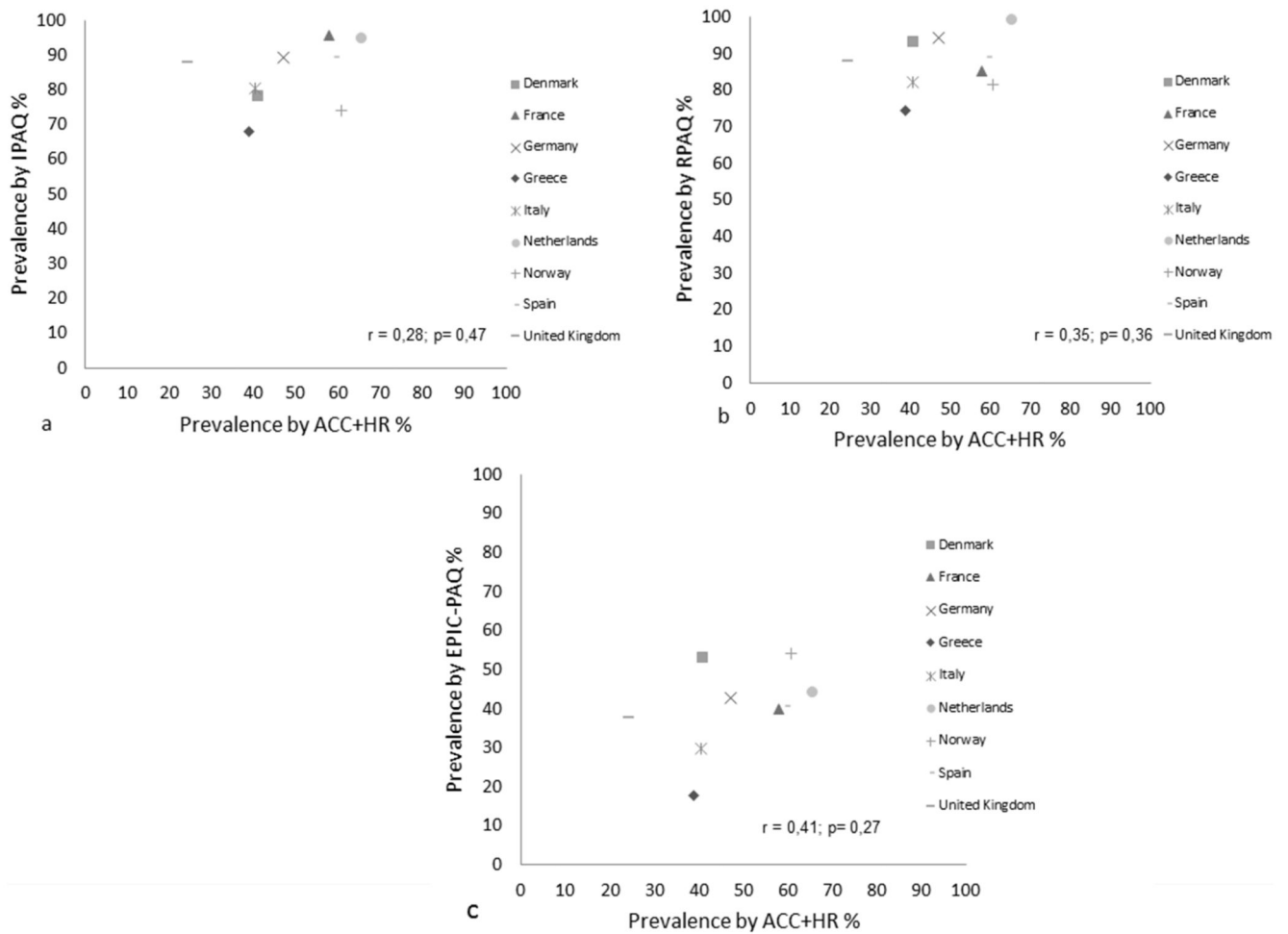
Figure 1.
Scatterplot of the proportion defined as sufficiently active by IPAQ vs ACC+HR (a), RPAQ vs ACC+HR (b) and EPIC-PAQ vs ACC+HR (c) by country. Overall correlation (r) between prevalence estimates revealed no relationship (r = 0,28; p= 0.47, r = 0,35; p= 0.36 and r = 0.42; p= 0.27) for IPAQ vs ACC+HR, RPAQ vs ACC+HR and EPIC-PAQ vs ACC+HR, respectively.

Figure 2.
Meta-analysis of the agreement (Kappa coefficients) between IPAQ (a), RPAQ (b) and EPIC-PAQ (c) and compared with objectively measured physical activity in defining the prevalence of compliance with recommendations, by country and overall.

Table 1

Participants characteristics (mean, SD [standard deviation])

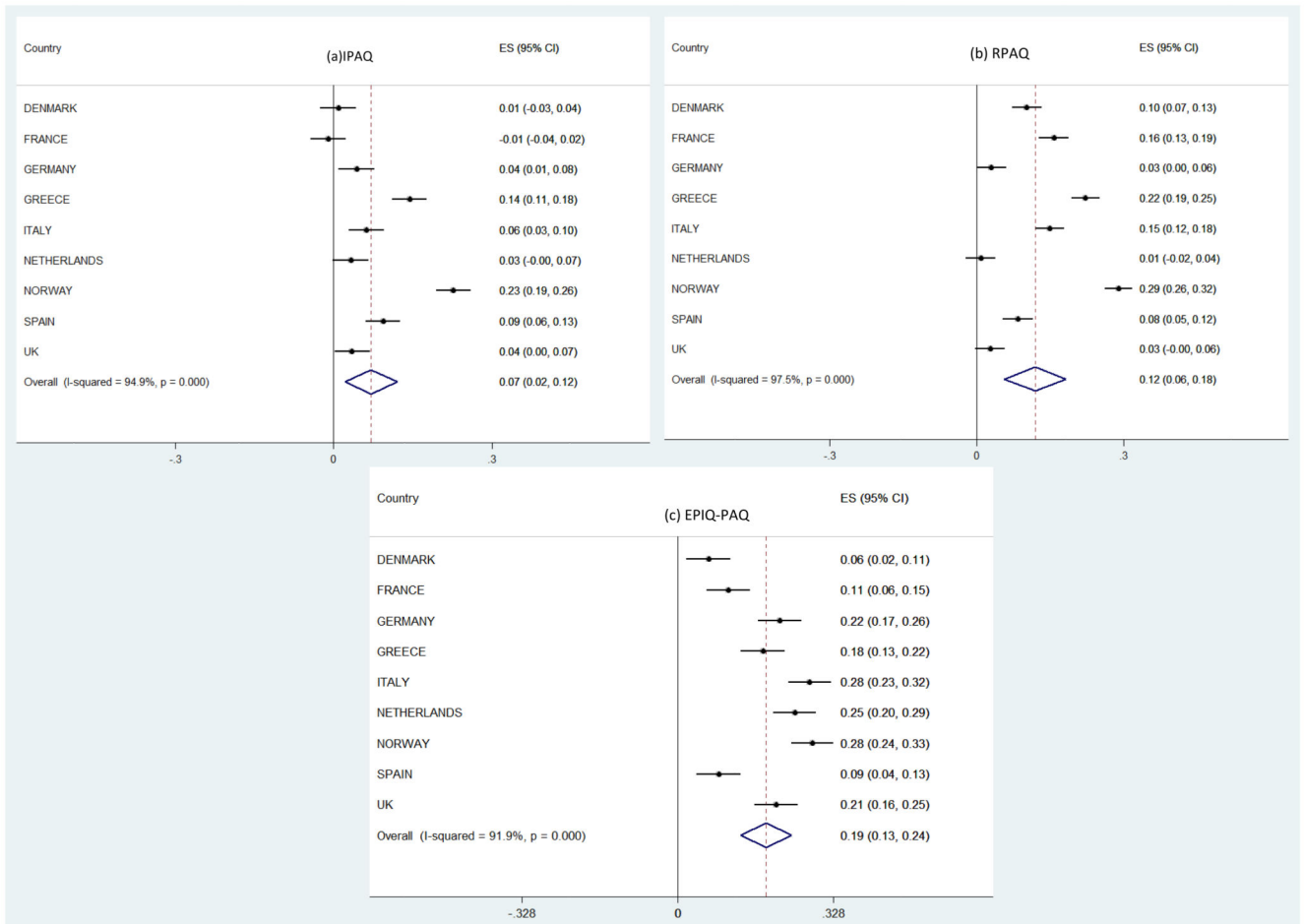| Country	N	Age (years)	SD	Height (m)	SD	Weight (kg)	SD	BMI (kg·m⁻²)	SD
Women									
Denmark	111	57.8	4.2	1.6	0.05	69.9	12.7	26.0	4.4
France	171	54.9	7.5	1.6	0.06	61.4	9.3	23.3	3.3
Germany	125	55.5	4.5	1.6	0.06	69.0	10.9	25.7	4.0
Greece	117	51.6	16.2	1.7	0.08	69.5	13.4	27.0	5.5
Italy	142	53.2	6.5	1.6	0.06	63.9	10.6	25.0	3.9
Netherlands	177	59.2	10.0	1.7	0.07	62.9	7.7	22.6	2.4
Norway	176	48.4	4.9	1.6	0.06	71.1	10.8	26.1	3.5
Spain	112	49.2	8.4	1.6	0.06	65.2	9.6	25.5	3.6
United Kingdom	103	60.1	7.5	1.6	0.06	69.6	11.2	26.6	4.1
Total, women	1234	54.3	9.2	1.6	0.07	66.6	11.2	25.1	4.1
Men									
Denmark	66	58.6	3.6	1.8	0.06	87.2	10.4	27.6	3.2
France									
Germany	83	58.2	3.1	1.8	0.05	86.2	12.5	27.5	3.5
Greece	65	51.0	18.9	1.7	0.08	84.4	12.2	27.7	3.5
Italy	53	53.5	6.3	1.7	0.06	79.1	13.7	26.2	4.0
Netherlands	29	50.3	11.2	1.8	0.06	77.9	9.6	23.5	2.2
Norway									
Spain	90	51.9	7.2	1.7	0.07	80.7	11.0	27.2	3.3
United Kingdom	93	61.5	7.9	1.8	0.06	85.2	12.4	27.5	3.3
Total, men	479	55.7	10.1	1.8	0.07	83.6	12.2	27.1	3.5
Total both sexes	1713	54.7	9.5	1.7	0.09	71.4	13.8	25.7	4.0

Table 2

Proportion (95 % CI) of participants meeting the physical activity recommendations according to the different PAQs and objective monitoring (Acc+HR) by country, sex and overall (N=1713).

	Men	95 % CI	Women	95 % CI	All	95 % CI	Mean sex difference
IPAQ							
Denmark	80.3	70.7 to 89.9	76.6	68.7 to 84.0	78.0	71.9 to 84.1	4.3
France[†]			95.3	92.1 to 98.5			
Germany	89.2	82.5 to 95.9	89.6	84.2 to 95.0	89.4	85.2 to 93.6	0.4
Greece	67.7	56.3 to 79.1	66.7	58.2 to 75.2	67.0	60.2 to 73.8	1.0
Italy	83.0	72.9 to 93.1	79.6	73.0 to 86.2	80.5	74.9 to 86.1	3.4
Netherlands	96.6	90.3 to 103.2	94.4	91.0 to 97.8	94.7	91.6 to 97.8	2.2
Norway[†]			73.9	67.4 to 80.4			
Spain	91.1	85.2 to 96.9	88.4	82.5 to 94.3	89.6	85.4 to 93.8	1.7
UK	84.9	77.6 to 92.2	88.3	82.1 to 94.5	86.7	81.9 to 91.5	3.6
Overall	84.3	81.0 to 87.6	84.1	82.1 to 86.1	84.2	82.5 to 85.9	0.2
RPAQ							
Denmark	93.9	88.3 to 99.7	91.9	86.8 to 97.0	92.7	89.0 to 96.6	2.0
France[†]			85.4	80.1 to 90.6			
Germany	97.6	94.3 to 100.9	92.0	87.2 to 96.8	94.2	91.0 to 97.4	5.6
Greece	73.8	63.1 to 84.5	74.4	66.5 to 82.3	74.2	67.8 to 80.6	0.6
Italy	88.7	80.2 to 97.2	79.6	73.0 to 86.2	82.1	84.5 to 93.1	9.1
Netherlands	96.6	90.3 to 103.2	99.4	98.3 to 100.5	99.0	97.6 to 100.4	2.8
Norway[†]			81.3	75.5 to 87.1			
Spain	93.3	88.1 to 98.5	85.7	79.2 to 92.2	89.1	84.8 to 93.4	7.6
UK	92.5	87.1 to 97.6	83.5	76.3 to 90.7	87.8	83.2 to 92.4	9.0
Overall	91.0	88.1 to 93.3	86.2	84.3 to 88.1	87.6	85.9 to 89.1	4.8[*]
EPIC-PAQ							

	Men	95 % CI	Women	95 % CI	All	95 % CI	Mean sex difference
IPAQ							
Denmark	56.1	44.8 to 68.6	50.5	41.2 to 59.6	52.5	45.5 to 60.1	6.1
France†			40.9	33.5 to 48.3	40.8	33.5 to 48.1	
Germany	41.0	30.4 to 51.6	44.0	35.3 to 52.7	42.8	36.1 to 49.5	3.0
Greece	20.0	11.2 to 30.6	16.2	9.2 to 22.2	17.6	12.2 to 23.0	3.8
Italy	35.8	22.9 to 48.7	27.5	20.2 to 34.8	29.7	23.3 to 36.1	8.3
Netherlands	79.3	64.6 to 94.0	38.4	31.2 to 45.6	44.2	37.2 to 50.8	40.9 *
Norway†			54.0	46.6 to 61.4	54.0	46.6 to 61.4	
Spain	51.1	40.8 to 61.4	32.1	23.4 to 40.7	40.6	33.8 to 47.4	19.0 *
UK	50.5	40.3 to 60.7	26.2	17.7 to 34.7	37.8	31.0 to 44.6	24.3 *
Overall	45.7	41.2 to 50.1	37.7	35.0 to 40.4	39.9	37.5 to 42.1	8.0 *
Acc+HR							
Denmark	47.0	35.8 to 59.8	36.9	28.3 to 46.11	40.7	33.9 to 48.3	10.1
France†			57.9	50.7 to 65.3	58.0	50.7 to 65.3	
Germany	49.2	38.6 to 60.2	45.6	36.9 to 54.3	47.1	40.3 to 53.9	3.6
Greece	47.8	35.8 to 59.8	34.2	25.5 to 42.3	39.6	32.5 to 46.7	13.6 *
Italy	54.7	41.3 to 68.1	35.2	27.3 to 43.1	40.5	33.6 to 47.4	19.5 *
Netherlands	89.7	90.3 to 103.1	61.6	54.4 to 68.8	65.5	59.2 to 72.2	28.1 *
Norway†			60.8	53.6 to 68.0	60.8	53.6 to 68.0	
Spain	63.3	53.3 to 73.3	56.3	47.1 to 65.5	59.4	52.6 to 66.2	7.0
UK	34.4	24.7 to 44.1	15.5	8.5 to 22.5	24.5	18.0 to 30.0	18.9 *
Overall	51.8	47.3 to 56.3	47.2	44.4 to 50.0	48.5	46.1 to 50.9	4.6 *

*
P< 0.05 for sex within PAQ's and objective measure

†
Only females included

Table 3

Specificity and sensitivity (95% CI) for the agreement between the different PAQs and objective method.

Country	IPAQ		RPAQ		EPIC	
	Specificity % (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)	Sensitivity % (95% CI)	Specificity % (95% CI)	Sensitivity % (95% CI)
Denmark	77.8 (71.7 to 84.0)	21.9 (15.8 to 30.0)	100.0 (95.1 to 100.0)	12.4 (7.5 to 17.3)	55.6 (46.8 to 61.4)	49.5 (42.1 to 56.9)
France	94.9 (91.6 to 98.7)	4.2 (1.2 to 7.2)	91.9 (87.8 to 96.0)	23.6 (17.2 to 30.0)	45.5 (36.2 to 55.2)	65.3 (58.2 to 72.4)
Germany	91.8 (84.7 to 95.8)	12.7 (7.7 to 20.2)	96.0 (90.0 to 98.4)	7.3 (3.8 to 10.9)	54.1 (44.3 to 63.6)	67.3 (58.0 to 75.3)
Greece	77.8 (71.8 to 83.8)	40.0 (32.9 to 47.1)	90.3 (86.0 to 94.6)	36.4 (29.4 to 43.4)	27.8 (21.3 to 34.3)	89.1 (84.6 to 93.6)
Italy	84.8 (75.3 to 91.1)	22.4 (15.8 to 30.8)	92.4 (84.4 to 96.5)	25.0 (18.1 to 33.6)	45.6 (35.1 to 56.5)	81.0 (72.9 to 87.3)
Netherlands	95.6 (90.7 to 98.0)	7.0 (3.1 to 15.5)	99.3 (96.5 to 99.9)	1.4 (0.2 to 7.6)	54.1 (47.3 to 60.9)	74.6 (68.7 to 88.2)
Norway	82.2 (74.2 to 88.3)	39.1 (28.5 to 50.1)	91.6 (84.8 to 95.5)	34.8 (24.6 to 47.5)	65.4 (56.0 to 73.8)	63.8 (52.8 to 74.1)
Spain	93.3 (87.7 to 96.1)	15.9 (9.4 to 25.0)	92.5 (85.6 to 95.5)	15.7 (10.7 to 20.7)	44.2 (37.5 to 51.0)	64.6 (58.0 to 71.2)
United Kingdom	91.7 (80.5 to 96.7)	14.9 (10.0 to 21.5)	91.7 (80.5 to 96.7)	13.5 (8.9 to 19.5)	56.3 (42.3.3 to 69.3)	68.2 (60.4 to 75.2)
Total women	89.1 (86.3 to 91.4)	20.2 (17.4 to 23.5)	93.3 (91.0 to 95.0)	20.1 (17.9 to 22.3)	47.2 (44.4 to 50.0)	68.7 (67.8 to 73.7)
Total men	87.9 (85.0 to 91.0)	19.5 (16.0 to 23.0)	96.4 (94.7 to 98.1)	14.7 (11.5 to 17.9)	56.5 (52.0 to 61.0)	65.8 (59.5 to 71.6)
Total both sexes	88.7 (86.3 to 91.0)	20.0 (17.5 to 22.7)	94.2 (93.1 to 95.3)	18.7 (16.9 to 20.5)	50.2 (45.4 to 55.0)	69.8 (67.6 to 72.0)