RESEARCH   **Open Access**

Estimating physical activity from self-reported behaviours in large-scale population studies using network harmonisation: findings from UK Biobank and associations with disease outcomes

Matthew Pearce[1], Tessa Strain[1], Youngwon Kim[2], Stephen J. Sharp[1], Kate Westgate[1], Katrien Wijndaele[1], Tomas Gonzales[1], Nicholas J. Wareham[1] and Søren Brage[1*]

Abstract

Background: UK Biobank is a large prospective cohort study containing accelerometer-based physical activity data with strong validity collected from 100,000 participants approximately 5 years after baseline. In contrast, the main cohort has multiple self-reported physical behaviours from > 500,000 participants with longer follow-up time, offering several epidemiological advantages. However, questionnaire methods typically suffer from greater measurement error, and at present there is no tested method for combining these diverse self-reported data to more comprehensively assess the overall dose of physical activity. This study aimed to use the accelerometry sub-cohort to calibrate the self-reported behavioural variables to produce a harmonised estimate of physical activity energy expenditure, and subsequently examine its reliability, validity, and associations with disease outcomes.

Methods: We calibrated 14 self-reported behavioural variables from the UK Biobank main cohort using the wrist accelerometry sub-cohort ($n = 93,425$), and used published equations to estimate physical activity energy expenditure ($PAEE_{SR}$). For comparison, we estimated physical activity based on the scoring criteria of the International Physical Activity Questionnaire, and by summing variables for occupational and leisure-time physical activity with no calibration. Test-retest reliability was assessed using data from the UK Biobank repeat assessment ($n = 18,905$) collected a mean of 4.3 years after baseline. Validity was assessed in an independent validation study ($n = 98$) with estimates based on doubly labelled water ($PAEE_{DLW}$). In the main UK Biobank cohort ($n = 374,352$), Cox regression was used to estimate associations between $PAEE_{SR}$ and fatal and non-fatal outcomes including all-cause, cardiovascular diseases, respiratory diseases, and cancers.

Results: $PAEE_{SR}$ explained 27% variance in gold-standard $PAEE_{DLW}$ estimates, with no mean bias. However, error was strongly correlated with $PAEE_{DLW}$ ($r = -.98$; $p < 0.001$), and $PAEE_{SR}$ had narrower range than the criterion. Test-retest reliability ($\Lambda = .67$) and relative validity (Spearman = .52) of $PAEE_{SR}$ outperformed two common approaches for processing self-report data with no calibration. Predictive validity was demonstrated by associations with morbidity and mortality, e.g. 14% (95%CI: 11–17%) lower mortality for individuals meeting lower physical activity guidelines.

(Continued on next page)

* Correspondence: soren.brage@mrc-epid.cam.ac.uk
[1]MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge Biomedical Campus, Box 285, Cambridge CB2 0QQ, UK
Full list of author information is available at the end of the article

(Continued from previous page)

Conclusions: The PAEE$_{SR}$ variable has good reliability and validity for ranking individuals, with no mean bias but correlated error at individual-level. PAEE$_{SR}$ outperformed uncalibrated estimates and showed stronger inverse associations with disease outcomes.

Keywords: Accelerometer, Physical activity energy expenditure, Questionnaire, Calibration, Doubly labelled water

Background

Higher levels of physical activity have been shown to be associated with a lower risk of morbidity and mortality [1], but accurately assessing the dose of physical activity in large population studies remains challenging. Most large cohort studies with long follow-up have utilised self-report questionnaires to assess physical activity. These methods typically have lower cost and higher feasibility than more objective methods but are prone to measurement error [2], and may not capture physical activity across all activity domains meaning the full dose is not characterised [3]. UK Biobank has shown that it is feasible to collect accelerometer-based physical activity data with strong validity [4] on a large scale ($n > 100,000$) [5]. Despite this, the main UK Biobank cohort is five times larger and has longer follow-up time to morbidity and mortality outcomes, which offers several epidemiological advantages compared to the more recent accelerometer sub-cohort. However, there is currently no tested method for estimating total volume of physical activity from the self-report information in UK Biobank collected at baseline.

The baseline questionnaire includes items adapted from the International Physical Activity Questionnaire (IPAQ) [6] and the Recent Physical Activity Questionnaire (RPAQ) [7, 8]. Responses could theoretically be processed separately using methods developed specifically for those two questionnaires, but using the totality of the available data should provide a more comprehensive estimate of the total dose, as they capture information about complimentary types, intensities and domains of activity. Previous work has shown how these self-reported behaviours relate to a summary of movement volume from 24-h wrist acceleration [9], and how wrist acceleration relates to physical activity energy expenditure (PAEE) as measured by the gold-standard method of doubly labelled water [4]. Despite the paucity of validation studies describing the direct relationship between these self-report data and those from the gold-standard method, it is possible to use network harmonisation [10] to combine the above strands of evidence to estimate PAEE; this would capitalise on the very large sample size of strand one and the more robust relationship between two objective measures in strand two, but the reliability and validity of this approach have not yet been tested in this context.

This study aimed to: 1) use the UK Biobank accelerometry sub-cohort to harmonise the self-reported behavioural variables and produce a summary estimate of PAEE; 2) examine test-retest reliability of this estimate using the UK Biobank repeat assessment sub-cohort; 3) assess validity of the PAEE estimate using values from a gold-standard doubly labelled water (DLW) based assessment in an independent validation study; 4) investigate associations of the PAEE estimate with morbidity and mortality in the main UK Biobank cohort.

Methods

The following sections set out the collection and processing of relevant data in UK Biobank, the methods of the DLW validation study, and the statistical analyses.

UK Biobank

Participants and study design

UK Biobank is an ongoing prospective cohort study of 502,625 adults aged 40–69 years residing within 25 miles of one of 22 assessment centres in England, Scotland, and Wales. Additional file 1: Figure S1 describes the exclusion criteria and sample sizes used in different components of the present study. Participants were identified from National Health Service general practitioner registries and invited to a baseline assessment between 2006 and 2010 [11]. A subsample of 20,346 participants attended a repeat assessment visit (2012–2013), and between 2013 and 2015 another partially overlapping subsample of 106,053 participated in a follow-up study during which they wore a wrist-mounted accelerometer for 7 days [5]. The UK Biobank study was approved by the North West Multicentre Research Ethics Committee and all participants provided written informed consent. Data for the current analysis were downloaded on 4th April 2019, containing information from 502,536 participants with baseline measures following withdrawals.

Self-reported behaviours

Physical activity, television viewing, computer use, and sleep were self-reported using a touch-screen questionnaire and responses were used to generate behavioural variables as previously described [9]. There are a total of 14 behavioural variables which are detailed in

Supplementary Table S1; data for these were collected at baseline (2006–2010) and in a subsample during the repeat-assessment visit (2012–2013). IPAQ-based questions were used to derive minutes per day of moderate-to-vigorous physical activity (MVPA), as well as the IPAQ score in metabolic equivalent of task (MET) minutes/day for comparison [6] (Supplementary Table S2). Similarly, RPAQ-based questions were used to derive (minutes per day unless stated otherwise): walking for pleasure, strenuous sports, other exercises, light do-it-yourself (DIY), heavy DIY, heavy physical work, walking/standing work, sedentary work, getting about method (categorical: car or public transport, mixed use, walking or cycling), commuting method (categorical: car or public transport, mixed use, walking or cycling), television viewing (hours per day), computer use (hours per day). The questions are similar but not identical to those used in the original RPAQ [7]. Therefore, an alternative summary was computed for this instrument following the same scoring principles; this score in MET-minutes/day comprised the sum of leisure-time and occupational physical activity and is denoted LTPA+OPA in the present analysis (Supplementary Table S2). Sleep and nap time was categorised as: ≤ 5 h per day, 6 h per day, 7 h per day, 8 h per day, ≥ 9 h per day. As part of pilot testing, some participants completed a different baseline questionnaire to the rest of the main cohort; the data were incompatible and we therefore excluded these participants ($n = 3797$). We also removed participants for whom the sum of daily MVPA, television viewing, computer use and sleep was greater than 24 h ($n = 4514$). These variables were chosen as they should be mutually exclusive and thus used to detect generic misunderstanding of the behavioural questions.

Accelerometer sub-cohort
The collection and processing of the accelerometer data have been described in greater detail previously [5]. Between 2013 and 2015 invitations to participate in the accelerometer sub-cohort were sent to 236,519 participants who had provided a valid email address at recruitment. Consenting participants ($n = 106,053$) were sent an accelerometer (Axivity AX3, Newcastle upon Tyne, UK) initialised to capture three-dimensional acceleration at 100 Hz continuously for 7 days which they were asked to begin wearing immediately on their dominant wrist. Participants were asked to return the accelerometer via pre-paid envelope after the monitoring period. Euclidean norm minus one (ENMO) was calculated as the Euclidean norm (vector magnitude) of calibrated acceleration [12] in three axes minus one gravitational unit (1000 m-g) and negative values were truncated to zero [13]. Periods of ≥ 60 min during which the standard deviations (SD) of all three axes were <

13.0 m-g were identified as non-wear. Mean wrist ENMO in m-g was summarised across valid wear-time (data across the full 24 h spectrum and at least 72 h of wear in total) for each individual whilst minimising diurnal bias caused by non-wear [14].

Calibration models
In order to utilise the totality of the self-report information in UK Biobank, linear regression models were fitted to estimate the association between the 14 behavioural variables and movement volume (ENMO) using data from the accelerometry sub-cohort. Continuous self-report variables were natural log ($\log_e(x + 1)$) transformed (+ 1 due to zero values). Coefficients were mutually adjusted (i.e. entered in the same regression model) and derived separately for men and women. We also accounted for change in both age and season between baseline and the accelerometry assessment by adding delta terms to the regression models. Participants with < 72 h of wear time ($n = 6310$) or mean wrist ENMO ≥ 500 m-g ($n = 4$) were excluded. The standard error (SE) of each predicted PAEE was calculated using the variance-covariance matrix from the model and the values of each variable.

Prediction of PAEE from self-report (PAEE$_{SR}$)
The sex-specific regression models developed in the accelerometry sub-cohort were used to predict mean wrist ENMO from self-report data in the main UK Biobank cohort. These predicted wrist ENMO values were then converted to PAEE$_{SR}$ in kJ/day/kg using data from a similarly aged UK cohort [15] and a previously reported scaling equation for dominant wrist acceleration [4]. To assess reliability, this process was repeated for participants with complete self-report data collected during the repeat assessment visit ($n = 18,905$).

To propagate the uncertainty of the initial prediction of wrist ENMO and subsequent conversion to PAEE$_{SR}$, predicted wrist ENMO values were resampled 100 times at random from normal distributions centered at each individual's estimated wrist ENMO and its SE. In the same way, we sampled 100 beta and alpha coefficients used to convert wrist ENMO to PAEE$_{SR}$. Wrist ENMO was then converted to PAEE$_{SR}$ using the 100 sets of sampled values and coefficients. The mean and SD of the 100 predictions for each individual were used as the point estimate of PAEE$_{SR}$ and its SE, respectively.

Outcome assessment for survival analyses
Vital status and primary or secondary diagnoses of hospital episodes of participants were established by linkage to national death registry data obtained from the Health and Social Care Information Centre for England and Wales and the Information Services Department for

Scotland [11]. Censoring dates were 31st January 2018 in England and Wales, and 30th November 2016 in Scotland. International Classification of Diseases 10th edition codes were used to define disease outcomes as shown in Supplementary Table S3. Non-fatal outcomes were hospital episodes of heart failure, stroke, ischaemic heart disease, atrial fibrillation, all cardiovascular disease, chronic obstructive pulmonary disease, all respiratory disease, cancers including breast, prostate, endometrial, lung, colon, oesophageal, liver, gastric cardia, myeloid leukaemia, myeloma, rectum, bladder, malignant melanoma, and all cancer. Selection of site-specific cancer outcomes was based upon a previous review [16] and at least 25 events in the follow-up period. Fatal outcomes were all-cause mortality, cardiovascular disease mortality, respiratory disease mortality, and cancer mortality.

Covariate assessment for survival analyses

Demographic, lifestyle, and clinical variables were assessed at baseline by the aforementioned touch-screen questionnaire, verbal interview, or physical measurement. The following variables were considered as potential confounders of the relationship between $PAEE_{SR}$ and all-cause mortality: age, sex, ethnicity (white/non-white), Townsend deprivation index, highest educational level (degree or above/any other qualification/no qualification), employment status (unemployed/in paid or self-employment), alcohol consumption (never/previous/current), smoking (never/previous/current), salt added to food (never/sometimes), oily fish intake (never/sometimes), fruit and vegetable intake (score from 0 to 4), processed and red meat intake (average weekly frequency in days per week), body mass index (BMI) in three categories (< 25, $25-30$, $\geq 30 \, kg \cdot m^{-2}$), parental cancer history including history of bowel, lung, maternal breast cancer, or paternal prostate cancer (yes/no), parental history of heart disease, stroke, hypertension or diabetes (yes/no), use of blood pressure medication (yes/no), use of cholesterol lowering medication (yes/no), doctor-diagnosed diabetes or treatment with insulin (yes/no), doctor-diagnosed coronary heart disease, stroke or cancer (yes/no).

DLW validation study

The validity of $PAEE_{SR}$ values was assessed using DLW-based PAEE values ($PAEE_{DLW}$) in an independent validation study, details of which have previously been reported [4]. Participants were 100 adults aged 40–70 years recruited from the Fenland Study [17, 18] and invited to two assessment visits separated by 9–14 days for gold-standard assessment of total energy expenditure [19–30]. Resting energy expenditure and diet-induced thermogenesis values were subtracted from total energy expenditure and divided by body mass yielding an estimate of total daily $PAEE_{DLW}$ in kJ/day/kg. Participants also answered the UK Biobank questions needed to generate $PAEE_{SR}$ using the calibration model described above, although data were incomplete for some ($n = 2$). Ethical approval for this study was obtained from Cambridge University Human Biology Research Ethics Committee (Ref: HBREC/2015.16). All participants provided written informed consent.

Statistical analyses

Test-retest reliability of behavioural variables, $PAEE_{SR}$, IPAQ, and LTPA+OPA

Test-retest reliability (repeatability) of the 14 behavioural variables as well as the $PAEE_{SR}$, IPAQ, and LTPA+OPA summary scores was examined by regression of the repeat assessment measures (2012–2013) on baseline measures (2006–2010) yielding lambda coefficients [31] and their standard errors, while (weighted) Cohen's kappa coefficients [32] were calculated for ordinal variables.

Validity of $PAEE_{SR}$, IPAQ, and LTPA+OPA

Absolute validity (agreement) of the $PAEE_{SR}$ values was assessed by calculating the mean bias and 95% limits of agreement [33] compared with $PAEE_{DLW}$. We used $PAEE_{DLW}$ as the criterion in the main analysis rather than the average between $PAEE_{SR}$ and $PAEE_{DLW}$, which has been recommended [34]. However, error in $PAEE_{DLW}$ is very low compared to self-report, meaning $PAEE_{DLW}$ is likely to be closer to the latent 'true' level of the exposure. The plot of $PAEE_{SR}$ vs the average of $PAEE_{SR}$ and $PAEE_{DLW}$ was conducted as a sensitivity analysis. Precision was assessed by calculating root mean square error (RMSE), i.e. the square-root of the mean squared differences. Individual differences between $PAEE_{SR}$ and $PAEE_{DLW}$ were examined visually across the measurement range of the criterion. The association between each of $PAEE_{SR}$, IPAQ, and LTPA + OPA with $PAEE_{DLW}$ was modelled using linear regression. The relative validity (similar ranking of individuals) of the three summary scores was examined with Spearman's rank-order correlation using $PAEE_{DLW}$.

Survival analyses

In the main UK Biobank cohort, Cox regression with age as the underlying timescale was used to estimate associations between $PAEE_{SR}$ and each of the fatal and non-fatal outcomes, adjusted for all covariates listed above, and in a separate model omitting BMI. Hazard ratios were presented per 5 kJ/day/kg of PAEE as this is approximately equivalent to the lower World Health Organization guideline of 150 min of moderate intensity activity per week [35]. Models were weighted using the inverse of the individual-level SE; weights were normalised such that the sum of weights equalled the analytical sample size. Individuals with missing exposure data ($n = 20,133$) or

covariate data ($n = 19,778$) were excluded for the survival analyses, as were individuals with pre-baseline hospital episodes of ischaemic heart disease, stroke, respiratory disease or cancer as defined above ($n = 55,574$), and those with only self-reported doctor-diagnosed ischaemic heart disease, stroke, or cancer ($n = 23,402$). Finally, we excluded participants experiencing events in the first 2 years of follow-up ($n = 986$ for mortality; range 22 to 24,084 for non-fatal outcomes), meaning the final analysis sample for mortality analyses included 374,352 participants, with fewer for analyses of non-fatal outcomes. Breast and prostate cancer analyses were conducted in women only and men only, respectively.

For fatal outcomes, we compared the associations of each of the three summary scores ($PAEE_{SR}$, IPAQ, and LTPA+OPA) using the modelling approach described above, and presented hazard ratios per 1 SD increment of each exposure. We also repeated this adding sleep as a covariate in the Cox regression model when using IPAQ and LTPA+OPA. In sensitivity analyses, hazard ratios were also estimated by quartile of $PAEE_{SR}$ using all covariates, and in a separate model omitting BMI. We also replicated the main analysis described above in only those participants reporting pre-baseline disease and who did not die within 2 years of follow-up ($n = 77,843$). In addition, the associations between $PAEE_{SR}$ and each of the disease outcomes were assessed using cubic spline regression models (3 knots) using all the covariates. For this analysis, we used a reference $PAEE_{SR}$ level of a hypothetical man or woman reporting: no leisure-time physical activity, 8 hours per day of sedentary occupation, 2 hours per day of television viewing, 2 hours per day of computer use, motorised transport for commuting and getting about, and sleeping for $\geq 9\,h$ per day. All analyses were conducted using STATA/SE 14.2 (StataCorp, TX, USA).

Results

Baseline characteristics of participants from the studies included in analyses are shown in Table 1. Participants in the DLW validation study were, on average, 2 years younger and more active than those in UK Biobank. Following exclusions, 52,507 women and 41,918 men were included in the two separate regression analyses to predict wrist movement from self-report data. The resulting models explained 14 and 17% of variance in mean wrist ENMO (m-g) in women and men respectively. The sex-specific coefficients for the 14 behavioural variables are shown in Additional file 1: Table S4.

Test-retest reliability of behavioural variables, $PAEE_{SR}$, IPAQ, and LTPA+OPA scores

The mean (SD) time between baseline (2006–2010) and repeat assessment (2012–2013) was 4.3 (0.9) years.

Table 2 summarises self-reported behaviours at both time points: the largest change in reported behaviours between baseline and repeat assessment was for occupational variables, all of which decreased in duration. Test-retest reliability was higher for $PAEE_{SR}$ than for the IPAQ or LTPA+OPA scores of MET-minutes per day.

Validity of $PAEE_{SR}$, IPAQ, and LTPA+OPA scores

Self-report data were complete for 98 out of 100 participants in the DLW validation study. Figure 1 shows $PAEE_{SR}$ minus $PAEE_{DLW}$ plotted against $PAEE_{DLW}$. $PAEE_{DLW}$ mean (SD) was 50.0 (16.1) kJ/day/kg compared with 48.9 (3.7) kJ/day/kg for $PAEE_{SR}$. The mean bias was -1.1 (95%CI: -4.0 to 1.8 kJ/day/kg), or -2% of the criterion mean, and the limits of agreement were -30.2 to 28.1 kJ/day/kg ($\pm 58\%$). The RMSE was 14.5 kJ/day/kg, or 29% of the criterion mean. Error of $PAEE_{SR}$ was strongly correlated with $PAEE_{DLW}$ ($r = -.98$; $p < 0.001$); $PAEE_{SR}$ was an overestimate for less active individuals and an underestimate for the more active. Plotting error of $PAEE_{SR}$ vs the average of $PAEE_{SR}$ and $PAEE_{DLW}$ showed a similar proportional bias ($r = -.93$; $p < 0.001$, Supplemental Fig. S2). The range of $PAEE_{SR}$ (40.5 to 56.2 kJ/day/kg) was 81% narrower than $PAEE_{DLW}$ (9 to 91 kJ/day/kg). Spearman correlation between $PAEE_{SR}$ and $PAEE_{DLW}$ was $r_s = .52$ ($p < 0.001$), while for IPAQ and LTPA+OPA, Spearman correlations with $PAEE_{DLW}$ were $r_s = .23$ ($p = 0.022$) and $r_s = .41$ ($p < 0.001$), respectively. $PAEE_{SR}$ explained 27% of variance in $PAEE_{DLW}$ with a large negative intercept (Fig. 1). By comparison, IPAQ and LTPA+OPA scores explained 5 and 8%, respectively.

Survival analyses

During a median (interquartile range) 8.9 (8.3–9.5) years of follow-up (3,311,773 person-years), 9372 participants died. Each 5 kJ/day/kg of $PAEE_{SR}$ (equivalent to meeting the lower activity recommendations) was associated with an approximate 14% lower hazard of all-cause mortality (Fig. 2). Incidence of non-fatal respiratory disease (but severe enough to require hospital admission) was more strongly associated with $PAEE_{SR}$ than non-fatal cardiovascular disease or cancer incidence. Amongst site-specific cancers, $PAEE_{SR}$ was only associated with non-fatal breast and kidney cancers; numbers of people with most site-specific cancers were small. Similar associations were observed when omitting BMI as a covariate (Additional file 1: Figure S4), but associations were generally stronger in those with pre-baseline disease than the main cohort (Additional file 1: Figure S5; characteristics presented in Table S6). Comparing mortality associations of the three summary scores, hazard ratios for mortality per 1 SD increment were consistently strongest for $PAEE_{SR}$ (Fig. 3). The IPAQ and LTPA+OPA scores

Table 1 Characteristics of participants in UK Biobank and the DLW validation study

	UK Biobank		Independent DLW validation study
	Main cohort	Accelerometer sub-cohort	
Analysis sample (n)	374,352	93,425	98
Age at baseline (years)	56 (8)	56 (8)	54 (7)
Age at postal follow-up (years)	–	62 (8)	–
Proportion of women	56%	56%	50%
Weight (kg)	78 (16)	77 (15)	77 (14)
Body mass index (kg/m^2)	27 (5)	27 (5)	27 (3)
Minutes per day of:			
Heavy physical work	21 (51)	16 (43)	38 (65)
Walking/standing work	56 (91)	48 (84)	91 (107)
Sedentary work	110 (143)	122 (146)	144 (153)
MVPA	90 (110)	85 (99)	124 (125)
Walking for pleasure	14 (22)	15 (22)	16 (27)
Strenuous sports	2 (10)	3 (10)	4 (12)
Other exercises	9 (17)	10 (17)	12 (16)
Light DIY	10 (24)	11 (24)	11 (28)
Heavy DIY	6 (19)	6 (18)	7 (24)
Hours per day of:			
Television viewing	3 (2)	2 (1)	2 (1)
Computer use	1 (1)	1 (1)	2 (2)
Getting about method:			
Car or public transportation	48%	44%	41%
Mixed use	43%	47%	32%
Walking or cycling	9%	9%	27%
Commuting method:			
Car or public transportation	87%	85%	59%
Mixed use	8%	10%	19%
Walking or cycling	5%	5%	22%
Hours per day of sleep:			
≤ 5.0	5%	4%	0%
6.0	19%	18%	4%
7.0	40%	43%	32%
8.0	29%	29%	49%
≥ 9.0	7%	6%	15%
PAEE$_{SR}$ (kJ/day/kg)	47 (4)	48 (4)	49 (4)
IPAQ scoring (MET-minutes/day)	373 (458)	357 (412)	509 (468)
LTPA+OPA scoring (MET-minutes/day)	380 (425)	349 (383)	572 (560)

DIY do-it-yourself, *DLW* doubly labelled water, *IPAQ* International Physical Activity Questionnaire, *LTPA+OPA* leisure-time and occupational physical activity, *MET* metabolic equivalent of task, *MVPA* Moderate-to-vigorous physical activity, *PAEE$_{SR}$* physical activity energy expenditure predicted from self-report
Values are means (standard deviations) unless otherwise stated

showed no association with cancer mortality in contrast to PAEE$_{SR}$. Additionally adjusting for sleep in the Cox model did not meaningfully alter associations for IPAQ and LTPA+OPA scores (data not shown).

There were dose-response associations across quartiles of PAEE$_{SR}$, with lower hazard in higher quartiles, and attenuation of the effect with additional adjustment for BMI (Supplementary Table S5). There was a non-linear

Table 2 Reliability of self-reported behaviours using baseline and repeat assessment in UK Biobank (*n* = 18,905)

	Baseline	Repeat	Lambda/kappa (SE)
Years between baseline and repeat:			
< 3		11%	
≥ 3 to < 4		24%	
≥ 4 to < 5		43%	
≥ 5 to < 6		21%	
≥ 6		1%	
Minutes per day of:			
Heavy physical work	16 (43)	11 (36)	Λ = 0.527 (0.005)
Walking/standing work	47 (82)	31 (70)	Λ = 0.482 (0.005)
Sedentary work	112 (143)	76 (128)	Λ = 0.609 (0.005)
MVPA	83 (99)	81 (94)	Λ = 0.479 (0.006)
Walking for pleasure	15 (22)	16 (24)	Λ = 0.520 (0.007)
Strenuous sports	3 (10)	2 (10)	Λ = 0.453 (0.006)
Other exercises	10 (18)	10 (17)	Λ = 0.456 (0.006)
Light DIY	11 (26)	10 (23)	Λ = 0.256 (0.006)
Heavy DIY	7 (19)	6 (17)	Λ = 0.273 (0.006)
Hours per day of:			
Television viewing	3 (1)	3 (2)	Λ = 0.821 (0.005)
Computer use	1 (1)	1 (1)	Λ = 0.547 (0.007)
Getting about method:			K = 0.324 (0.006)
Car or public transportation	49%	48%	
Mixed use	44%	45%	
Walking or cycling	7%	7%	
Commuting method:			K = 0.487 (0.006)
Car or public transportation	90%	93%	
Mixed use	7%	5%	
Walking or cycling	3%	2%	
Hours per day of sleep:			K = 0.500 (0.005)
≤ 5.0	3%	4%	
6.0	18%	17%	
7.0	42%	40%	
8.0	30%	31%	
≥ 9.0	7%	8%	
PAEE$_{SR}$ (kJ/day/kg)	46 (4)	47 (4)	Λ = 0.671 (0.005)
IPAQ scoring (MET-minutes/day)	345 (411)	342 (395)	Λ = 0.489 (0.006)
LTPA+OPA scoring (MET-minutes/day)	349 (382)	281 (337)	Λ = 0.552 (0.005)

DIY do-it-yourself, *IPAQ* International Physical Activity Questionnaire, *LTPA+OPA* leisure-time and occupational physical activity, *MET* metabolic equivalent of task, *MVPA* moderate-to-vigorous physical activity, *PAEE$_{SR}$* physical activity energy expenditure predicted from self-report, *SE* standard error, Λ Lambda coefficient, K weighted kappa coefficient
Values are mean (standard deviation) unless otherwise stated

inverse association of PAEE$_{SR}$ with all-cause mortality (Supplementary Fig. S3), with steeper gradient of the relationship moving from the least active individual to ~ 15 kJ/day/kg PAEE$_{SR}$, and shallower gradient above that level with greater uncertainty.

Discussion

This study reports the reliability and validity of PAEE predicted from a range of self-reported behaviours using a network harmonisation approach which included calibration to 7-day wrist accelerometry in approximately
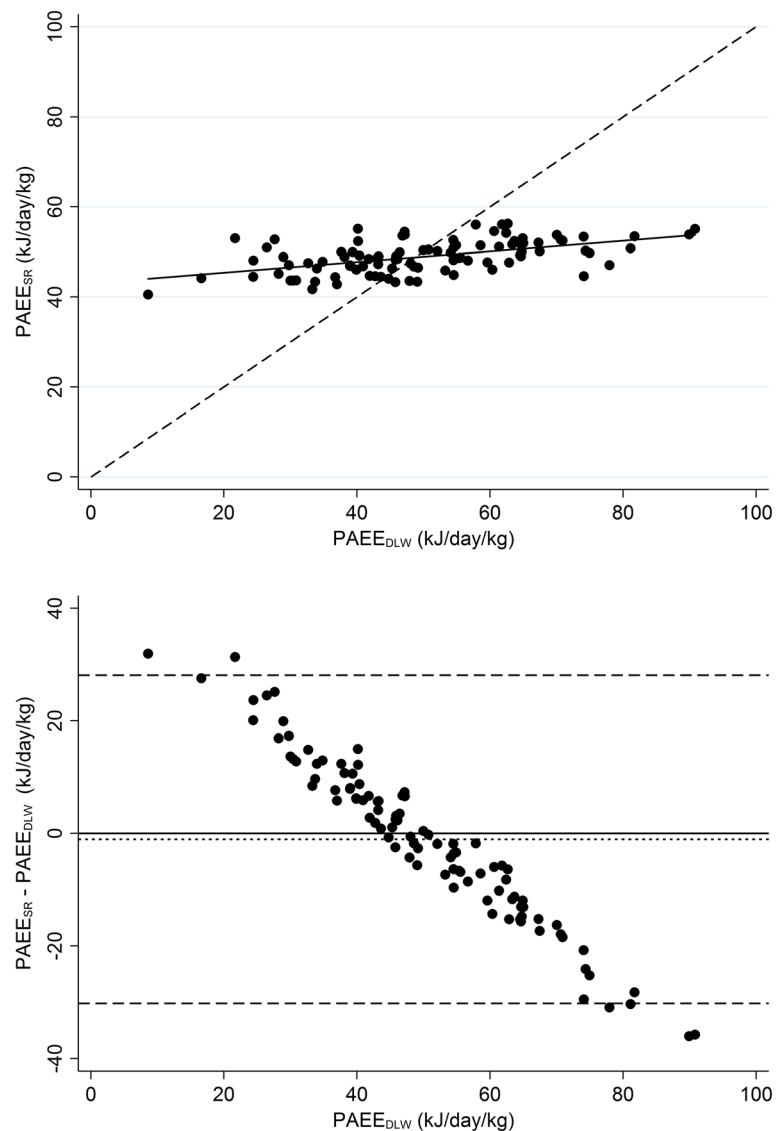
Fig. 1 Validity of physical activity energy expenditure predicted from self-report (PAEE_SR) vs. doubly labelled water based PAEE (PAEE_DLW). Upper panel shows scatter plot with line of unity (dashed) and regression line (solid); lower panel shows differences between physical activity energy expenditure predicted from self-report (PAEE_SR) and PAEE_DLW, plotted against PAEE_DLW. Reference lines indicate mean difference (dotted) and 95% limits of agreement (dashed). $n = 98$

100,000 free-living individuals. Our findings suggest that this method of combining behavioural data in UK Biobank produces PAEE values suitable for ranking individuals (based on Spearman's rank-order correlation) and demonstrates predictive validity when examining associations with morbidity and mortality, for example showing 14% lower mortality for individuals accumulating PAEE equivalent to meeting the lower World Health Organization physical activity guidelines [35]. However there are challenges with interpretation on an absolute scale due to marked under- and over-estimation at the exposure extremes.

Test-retest reliability of PAEE_SR outperformed MET-minute scores from IPAQ and LTPA+OPA and many previous self-reported estimates [2] despite an average of 4 years between baseline and repeat assessment, during which it might be expected for physical activity to decline in this population. We were not able to examine whether there were 'true' within-individual changes in PAEE between time-points using a criterion, but accounting for such changes would likely serve to improve reliability coefficients observed here. It is encouraging to note that although the behaviours demonstrated relatively poor test-retest
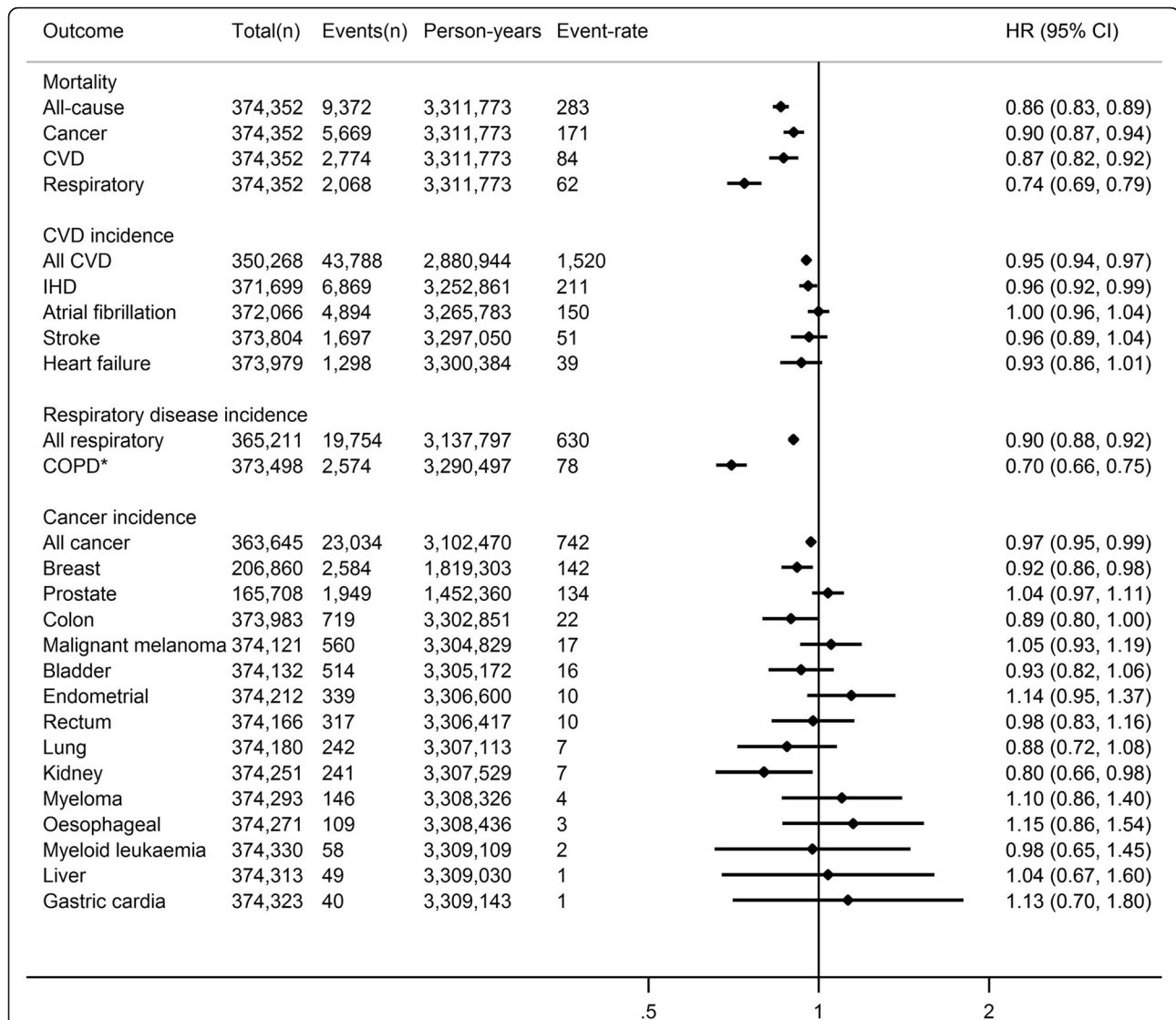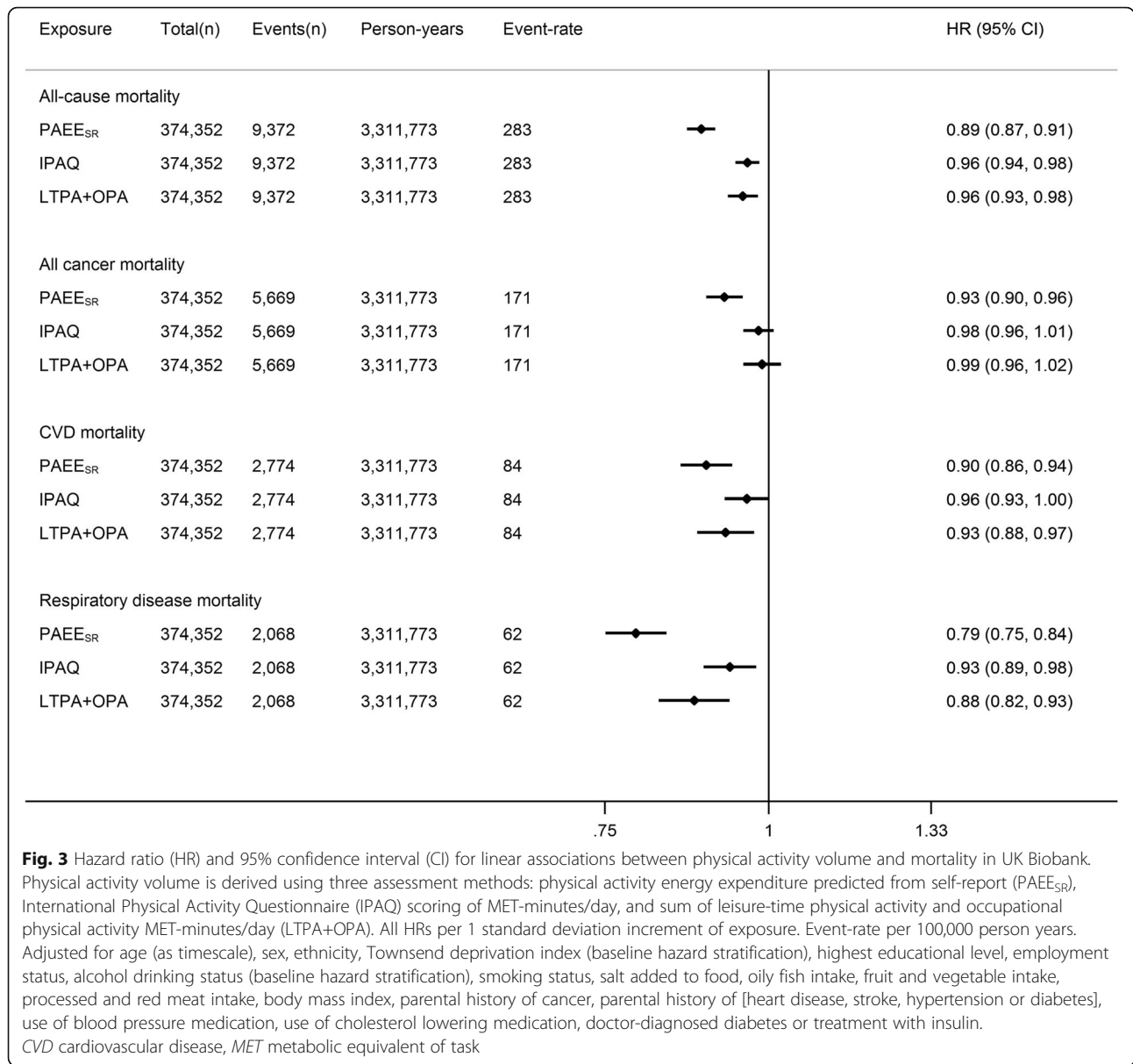
Outcome	Total(n)	Events(n)	Person-years	Event-rate		HR (95% CI)
Mortality						
All-cause	374,352	9,372	3,311,773	283		0.86 (0.83, 0.89)
Cancer	374,352	5,669	3,311,773	171		0.90 (0.87, 0.94)
CVD	374,352	2,774	3,311,773	84		0.87 (0.82, 0.92)
Respiratory	374,352	2,068	3,311,773	62		0.74 (0.69, 0.79)
CVD incidence						
All CVD	350,268	43,788	2,880,944	1,520		0.95 (0.94, 0.97)
IHD	371,699	6,869	3,252,861	211		0.96 (0.92, 0.99)
Atrial fibrillation	372,066	4,894	3,265,783	150		1.00 (0.96, 1.04)
Stroke	373,804	1,697	3,297,050	51		0.96 (0.89, 1.04)
Heart failure	373,979	1,298	3,300,384	39		0.93 (0.86, 1.01)
Respiratory disease incidence						
All respiratory	365,211	19,754	3,137,797	630		0.90 (0.88, 0.92)
COPD*	373,498	2,574	3,290,497	78		0.70 (0.66, 0.75)
Cancer incidence						
All cancer	363,645	23,034	3,102,470	742		0.97 (0.95, 0.99)
Breast	206,860	2,584	1,819,303	142		0.92 (0.86, 0.98)
Prostate	165,708	1,949	1,452,360	134		1.04 (0.97, 1.11)
Colon	373,983	719	3,302,851	22		0.89 (0.80, 1.00)
Malignant melanoma	374,121	560	3,304,829	17		1.05 (0.93, 1.19)
Bladder	374,132	514	3,305,172	16		0.93 (0.82, 1.06)
Endometrial	374,212	339	3,306,600	10		1.14 (0.95, 1.37)
Rectum	374,166	317	3,306,417	10		0.98 (0.83, 1.16)
Lung	374,180	242	3,307,113	7		0.88 (0.72, 1.08)
Kidney	374,251	241	3,307,529	7		0.80 (0.66, 0.98)
Myeloma	374,293	146	3,308,326	4		1.10 (0.86, 1.40)
Oesophageal	374,271	109	3,308,436	3		1.15 (0.86, 1.54)
Myeloid leukaemia	374,330	58	3,309,109	2		0.98 (0.65, 1.45)
Liver	374,313	49	3,309,030	1		1.04 (0.67, 1.60)
Gastric cardia	374,323	40	3,309,143	1		1.13 (0.70, 1.80)

.5   1   2

Fig. 2 Hazard ratio (HR) and 95% confidence interval (CI) for linear associations of physical activity energy expenditure predicted from self-report (PAEE$_{SR}$, per 5 kJ/day/kg increments) with fatal and non-fatal outcomes in UK Biobank. Event-rate per 100,000 person years. Adjusted for age (as timescale), sex, ethnicity, Townsend deprivation index (baseline hazard stratification), highest educational level, employment status, alcohol drinking status (baseline hazard stratification), smoking status, salt added to food, oily fish intake, fruit and vegetable intake, processed and red meat intake, body mass index, parental history of cancer, parental history of [heart disease, stroke, hypertension or diabetes], use of blood pressure medication, use of cholesterol lowering medication, doctor-diagnosed diabetes or treatment with insulin. *COPD* chronic obstructive pulmonary disease; *CVD* cardiovascular disease; *IHD* ischaemic heart disease. *COPD incidence likely only represents the most severe cases as only approximately 25% of COPD cases are picked up in Hospital Episode Statistics data, compared to national surveys [36]

reliability in isolation, combining them provides an estimate of PAEE$_{SR}$ which seems to better reflect a habitual level of activity.

In the separate DLW validation study, PAEE$_{SR}$ showed a non-significant 2% underestimation and explained 27% variance in PAEE$_{DLW}$. This compares favourably to the relative validity of scores from IPAQ and LTPA+OPA reported here, as well as self-reported activity volume in previous work [2], with stronger criterion validity than estimates from IPAQ [6, 37, 38] and RPAQ [7, 8], on which

the questions are based. This may be explained by inclusion of a more comprehensive and complimentary list of physical activity behaviours, as well as sleep and sedentary behaviours which also provide information about the total volume of movement each day. Our validation study findings indicate that PAEE$_{SR}$ explains much higher levels of variance in the 'true' volume of physical activity assessed by PAEE$_{DLW}$, and this is reflected in stronger associations with mortality in UK Biobank compared with IPAQ and LTPA+OPA, which were more attenuated.

Exposure	Total(n)	Events(n)	Person-years	Event-rate		HR (95% CI)
All-cause mortality						
$PAEE_{SR}$	374,352	9,372	3,311,773	283		0.89 (0.87, 0.91)
IPAQ	374,352	9,372	3,311,773	283		0.96 (0.94, 0.98)
LTPA+OPA	374,352	9,372	3,311,773	283		0.96 (0.93, 0.98)
All cancer mortality						
$PAEE_{SR}$	374,352	5,669	3,311,773	171		0.93 (0.90, 0.96)
IPAQ	374,352	5,669	3,311,773	171		0.98 (0.96, 1.01)
LTPA+OPA	374,352	5,669	3,311,773	171		0.99 (0.96, 1.02)
CVD mortality						
$PAEE_{SR}$	374,352	2,774	3,311,773	84		0.90 (0.86, 0.94)
IPAQ	374,352	2,774	3,311,773	84		0.96 (0.93, 1.00)
LTPA+OPA	374,352	2,774	3,311,773	84		0.93 (0.88, 0.97)
Respiratory disease mortality						
$PAEE_{SR}$	374,352	2,068	3,311,773	62		0.79 (0.75, 0.84)
IPAQ	374,352	2,068	3,311,773	62		0.93 (0.89, 0.98)
LTPA+OPA	374,352	2,068	3,311,773	62		0.88 (0.82, 0.93)

.75   1   1.33

Fig. 3 Hazard ratio (HR) and 95% confidence interval (CI) for linear associations between physical activity volume and mortality in UK Biobank. Physical activity volume is derived using three assessment methods: physical activity energy expenditure predicted from self-report ($PAEE_{SR}$), International Physical Activity Questionnaire (IPAQ) scoring of MET-minutes/day, and sum of leisure-time physical activity and occupational physical activity MET-minutes/day (LTPA+OPA). All HRs per 1 standard deviation increment of exposure. Event-rate per 100,000 person years. Adjusted for age (as timescale), sex, ethnicity, Townsend deprivation index (baseline hazard stratification), highest educational level, employment status, alcohol drinking status (baseline hazard stratification), smoking status, salt added to food, oily fish intake, fruit and vegetable intake, processed and red meat intake, body mass index, parental history of cancer, parental history of [heart disease, stroke, hypertension or diabetes], use of blood pressure medication, use of cholesterol lowering medication, doctor-diagnosed diabetes or treatment with insulin. *CVD* cardiovascular disease, *MET* metabolic equivalent of task

Estimation errors were strongly negatively correlated with the criterion $PAEE_{DLW}$, i.e. displaying regression to the mean which is a consequence of using a relatively weak self-report instrument and prediction equations explaining relatively low levels of variance in wrist ENMO. The explanatory power of our models could have been strengthened using additional predictors (e.g. age, adiposity, etc.), but these are not directly representative of activity, and inclusion of more complicated predictors could hinder the transferability of newly derived models even if the relevant behavioural variables are available. Therefore, in order to make results more useful in answering epidemiological questions about the role of physical activity, we employed a model using behavioural data. Weak prediction models with a large constant narrowed the

observed range of predicted values substantially resulting in overestimation at the lower end and underestimation for more active individuals, widening the 95% limits of agreement. The component of $PAEE_{SR}$ from the constant is mathematically insensitive to differences in behaviour between individuals and does not influence correlations with criterion $PAEE_{DLW}$ or health associations; it does, however, impact interpretation of the exposure on an absolute scale, which presents a challenge for translation of observed associations with mortality to public health recommendations [39]. To facilitate such interpretation, we marginalised $PAEE_{SR}$ by subtracting the level of exposure of the least active individual from all participants in the analytical sample. The resulting dose-response curve for all-cause mortality is consistent with messages

emphasising greater benefits of increasing PAEE at the lower end of the exposure range [40]. Future work should explore methods to remedy these prediction errors and make use of alterative statistical approaches which combine data to give an integrated score [41]; the present study aimed to predict physical activity volume rather than characterise the overall pattern of health-related behaviours.

Limitations of this study include a healthy volunteer selection bias in UK Biobank such that it is not representative of the general population [42]; the accelerometer sub-cohort may also suffer from selection bias, although no major differences in self-reported behaviours or $PAEE_{SR}$ were observed here. There was an average 5.7 year gap between baseline self-reported behaviours and the accelerometer data used for calibration. We cannot rule out that physical activity may have changed in this time, although $PAEE_{SR}$ in the repeat assessment sub-cohort was relatively stable over a similar period and we accounted for change in age and season between these time points when deriving the prediction equations. The generalisability of prediction equations to those who did not survive until the accelerometry sub-cohort commenced must also be considered. This would be a concern if individuals who died during this period exhibited different relationships between self-reported behaviours and wrist ENMO, rather than just different behaviours. Given the size of the calibration samples, we argue that the heterogeneity of relationships included when deriving the models is sufficient. Furthermore, the accelerometry sub-study occurred over a number of years, meaning that some individuals who died relatively early in the follow-up period would have been included. Further work is necessary to explore the effects of using calibration equations with relatively weak self-report instruments, as these will be important for future harmonisation efforts (e.g. for synthesis of data from studies using different self-report methods). In particular, it is necessary to understand how calibrated and non-calibrated self-reported data should be used to estimate associations with disease outcomes across the full dose range, given the challenges of interpretation we have reported. Strengths of the work include use of $PAEE_{DLW}$ for examining validity, and propagation of the uncertainty (prediction errors) accrued at each step of our method for estimating PAEE to the analyses of associations with disease outcomes. Wrist accelerometry has strong validity compared to $PAEE_{DLW}$ [4], but is not available in the whole UK Biobank cohort and there is much less follow-up time in the sub-cohort where the measure is available. We used a robust criterion to calibrate and harmonise 14 self-report variables, with the added advantage that the necessary self-report data exist for approximately 475,000 participants, permitting use as an exposure, outcome, or covariate in future analyses.

Conclusions

In conclusion, we have successfully utilised a network harmonisation approach to exploit the diverse behavioural data in UK Biobank and derive an overall summary estimate of PAEE. The $PAEE_{SR}$ variable has good reliability and validity for ranking individuals compared with other self-report methods. It is the only estimate of PAEE available in the main UK Biobank cohort which has been tested against the gold-standard DLW-based criterion, showing no mean bias but a systematic bias at individual level stemming from inherent weaknesses of the self-report data. It does however have predictive validity in that it is prospectively associated with morbidity and mortality, and in a way that can be interpreted in a public health framework.

Supplementary information

Additional file 1: Table S1. Questions used to generate domain-specific and composite behavioural variables. **Table S2.** Calculation of comparison summary scores using METs. **Table S3.** International Classification of Diseases 10th edition (ICD-10) codes for outcome definition. **Table S4.** Mutually adjusted sex-specific coefficients (standard errors) for prediction of average daily wrist acceleration (m-g) from 14 self-reported behaviours. **Table S5.** Hazard ratio and 95% confidence interval for fatal and non-fatal outcomes by quartile of $PAEE_{SR}$ in UK Biobank. **Table S6** Baseline characteristics of participants with prevalent chronic disease in UK Biobank. **Figure S1.** Exclusions and sample sizes for analyses. **Figure S2.** Differences between physical activity energy expenditure predicted from self-report ($PAEE_{SR}$) and doubly labelled water based PAEE ($PAEE_{DLW}$), plotted against their mean. **Figure S3.** Hazard ratio and 95% confidence intervals for association between physical activity energy expenditure predicted from self-report ($PAEE_{SR}$) and disease outcomes in UK Biobank. **Figure S4.** Hazard ratio (HR) and 95% confidence interval (CI) for linear associations of physical activity energy expenditure predicted from self-report ($PAEE_{SR}$, per 5 kJ/day/kg increments) with fatal and non-fatal outcomes in UK Biobank. **Figure S5.** Hazard ratio (HR) and 95% confidence interval (CI) for linear associations of physical activity energy expenditure predicted from self-report ($PAEE_{SR}$, per 5 kJ/day/kg increments) with fatal and non-fatal outcomes in UK Biobank.

Abbreviations

BMI: Body mass index; CI: Confidence interval; COPD: Chronic obstructive pulmonary disease; CVD: Cardiovascular disease; DIY: Do-it-yourself; DLW: Doubly labelled water; ENMO: Euclidean norm minus one; HR: Hazard ratio; ICD-10: International Classification of Diseases 10th edition; IHD: Ischaemic heart disease; IPAQ: International Physical Activity Questionnaire; LTPA+OPA: Leisure-time and occupational physical activity; MET: Metabolic equivalent of task; MVPA: Moderate-to-vigorous physical activity; PAEE: Physical activity energy expenditure; $PAEE_{DLW}$: Physical activity energy expenditure from doubly labelled water; $PAEE_{SR}$: Physical activity energy expenditure predicted from self-report; RMSE: Root mean square error; RPAQ: Recent Physical Activity Questionnaire; SD: Standard deviation; SE: Standard error; TV: Television

Acknowledgements

Authors' contributions

Funding

Availability of data and materials

The UK Biobank data (Application Number 20684) that support the findings of this study are available to all bona fide researchers for health related research that is in the public interest, https://www.ukbiobank.ac.uk/register-apply/. The Biobank Validation Study data that support the findings of this study are available on request at https://epi-meta.mrc-epid.cam.ac.uk/.

Ethics approval and consent to participate

The UK Biobank study was approved by the North West Multicentre Research Ethics Committee and all participants provided written informed consent. Ethical approval for the Biobank Validation Study was obtained from Cambridge University Human Biology Research Ethics Committee (Ref: HBREC/2015.16). All participants provided written informed consent.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

[1]MRC Epidemiology Unit, University of Cambridge School of Clinical Medicine, Institute of Metabolic Science, Cambridge Biomedical Campus, Box 285, Cambridge CB2 0QQ, UK. [2]School of Public Health, The University of Hong Kong Li Ka Shing Faculty of Medicine, Room 301D 3/F, Jockey Club Building for Interdisciplinary Research, 5 Sassoon Road, Pokfulam, Hong Kong.

References

1. Lee I-M, Shiroma EJ, Lobelo F, Puska P, Blair SN, Katzmarzyk PT. Effect of physical inactivity on major non-communicable diseases worldwide: an analysis of burden of disease and life expectancy. Lancet. 2012;380(9838): 219–29.
2. Helmerhorst HJF, Brage S, Warren J, Besson H, Ekelund U. A systematic review of reliability and objective criterion-related validity of physical activity questionnaires. Int J Behav Nutr Phys Act. 2012;9:103.
3. Matthews CE, Moore SC, George SM, Sampson J, Bowles HR. Improving self-reports of active and sedentary behaviors in large epidemiologic studies. Exerc Sport Sci Rev. 2012;40(3):118–26.
4. White T, Westgate K, Hollidge S, Venables M, Olivier P, Wareham N, et al. Estimating energy expenditure from wrist and thigh accelerometry in free-living adults: a doubly labelled water study. Int J Obes. 2019. https://doi.org/10.1038/s41366-019-0352-x.
5. Doherty A, Jackson D, Hammerla N, Plötz T, Olivier P, Granat MH, et al. Large scale population assessment of physical activity using wrist worn accelerometers: the UK Biobank Study. PLoS One. Public Library of Science. 2017;12(2):e0169649. https://doi.org/10.1371/journal.pone.0169649.
6. Craig CL, Marshall AL, Sjostrom M, Bauman AE, Booth ML, Ainsworth BE, et al. International physical activity questionnaire: 12-country reliability and validity. Med Sci Sports Exerc. 2003;35(8):1381–95.
7. 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. https://doi.org/10.3945/ajcn.2009.28432.
8. Golubic R, May AM, Benjaminsen Borch K, Overvad K, Charles M-A, Diaz MJT, et al. Validity of electronically administered recent physical activity questionnaire (RPAQ) in ten European countries. PLoS One. Public library of science. 2014;9(3):e92829. https://doi.org/10.1371/journal.pone.0092829.
9. Kim Y, Wijndaele K, Sharp SJ, Strain T, Pearce M, White T, et al. Specific physical activities, sedentary behaviours and sleep as long-term predictors of accelerometer-measured physical activity in 91,648 adults: a prospective cohort study. Int J Behav Nutr Phys Act. 2019;16(1):41. https://doi.org/10.1186/s12966-019-0802-9.
10. Pearce M, Bishop TRP, Sharp S, Westgate K, Venables M, Wareham NJ, et al. Network harmonization of physical activity variables through indirect validation. J Meas Phys Behav. 2020;3:1. https://doi.org/10.1123/jmpb.2019-0001.
11. Littlejohns TJ, Sudlow C, Allen NE, Collins R. UK Biobank: opportunities for cardiovascular research. Eur Heart J. 2017;44:1–10.
12. van Hees VT, Fang Z, Langford J, Assah F, Mohammad A, da Silva ICM, et al. Autocalibration of accelerometer data for free-living physical activity assessment using local gravity and temperature: an evaluation on four continents. J Appl Physiol. American Physiological Society. 2014;117(7):738–44. https://doi.org/10.1152/japplphysiol.00421.2014.
13. van Hees VT, Gorzelniak L, Dean Leon EC, Eder M, Pias M, Taherian S, et al. Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. PLoS One. 2013/04/30. 2013;8(4):e61691.
14. Brage S, Westgate K, Wijndaele K, Godinho J, WN GS. Evaluation of a method for minimising diurnal information bias in objective sensor data. Amherst: ICAMPAM; 2013.
15. White T, Westgate K, Wareham NJ, Brage S. Estimation of physical activity energy expenditure during free-living from wrist accelerometry in UK adults. PLoS One. 2016;11(12):e0167472 San Francisco, CA USA: Public Library of Science. http://www.ncbi.nlm.nih.gov/pmc/articles/PMC5147924/.
16. Moore SC, Lee I-M, Weiderpass E, Campbell PT, Sampson JN, Kitahara CM, et al. Association of leisure-time physical activity with risk of 26 types of cancer in 1.44 million adults. JAMA Intern Med. 2016;176(6):816–25. https://doi.org/10.1001/jamainternmed.2016.1548.
17. O'Connor L, Brage S, Griffin SJ, Wareham NJ, Forouhi NG. The cross-sectional association between snacking behaviour and measures of adiposity: the Fenland Study, UK. Br J Nutr. 2015/09/08. 2015;114(8):1286–93.
18. Lindsay T, Westgate K, Wijndaele K, Hollidge S, Kerrison N, Forouhi N, et al. Descriptive epidemiology of physical activity energy expenditure in UK adults. The Fenland Study. medRxiv. 2019;1:19003442 http://medrxiv.org/content/early/2019/08/27/19003442.abstract.
19. Craig H. Isotopic standards for carbon and oxygen and correction factors for mass-spectrometric analysis of carbon dioxide. Geochim Cosmochim Acta. 1957;12(1):133–49 http://www.sciencedirect.com/science/article/pii/0016703757900248.
20. Schoeller DA. Recent advances from application of doubly labeled water to measurement of human energy expenditure. J Nutr. 1999;129(10):1765–8. https://doi.org/10.1093/jn/129.10.1765.
21. Elia M, Livesey G. Theory and validity of indirect calorimetry during net lipid synthesis. Am J Clin Nutr. 1988;47:591–607.
22. Haugen HA, Melanson EL, Tran ZV, Kearney JT, Hill JO. Variability of measured resting metabolic rate. Am J Clin Nutr. 2003/12/12. 2003;78(6): 1141–5.
23. Henry CJ. Basal metabolic rate studies in humans: measurement and development of new equations. Public Health Nutr. 2005/11/10. 2005;8(7a): 1133–52.
24. Nielsen S, Hensrud DD, Romanski S, Levine JA, Burguera B, Jensen MD. Body composition and resting energy expenditure in humans: role of fat, fat-free mass and extracellular fluid. Int J Obes Relat Metab Disord. 2000/10/18. 2000;24(9):1153–7.
25. Watson LP, Raymond-Barker P, Moran C, Schoenmakers N, Mitchell C, Bluck L, et al. An approach to quantifying abnormalities in energy expenditure

and lean mass in metabolic disease. Eur J Clin Nutr. 2013/11/28. 2014;68(2): 234–40.

26. Goldberg GR, Prentice AM, Davies HL, Murgatroyd PR. Overnight and basal metabolic rates in men and women. Eur J Clin Nutr. 1988/02/01. 1988;42(2): 137–44.

27. Bingham SA, Gill C, Welch A, Cassidy A, Runswick SA, Oakes S, et al. Validation of dietary assessment methods in the UK arm of EPIC using weighed records, and 24-hour urinary nitrogen and potassium and serum vitamin C and carotenoids as biomarkers. Int J Epidemiol. 1997;26(suppl_1): S137. https://doi.org/10.1093/ije/26.suppl_1.S137.

28. Mulligan AA, Luben RN, Bhaniani A, Parry-Smith DJ, Connor L, Khawaja AP, et al. A new tool for converting food frequency questionnaire data into nutrient and food group values: FETA research methods and availability. BMJ Open. 2014;4(3):e004503 http://bmjopen.bmj.com/content/4/3/e004503.abstract.

29. Jequier E. Pathways to obesity. Int J Obes Relat Metab Disord. 2002/08/14. 2002;26(Suppl 2):S12.

30. Brage S, Westgate K, Franks PW, Stegle O, Wright A, Ekelund U, et al. Estimation of free-living energy expenditure by heart rate and movement sensing: a doubly-labelled water study. PLoS One. 2015/09/09. 2015;10(9): e0137206.

31. Keogh RH, White IR. A toolkit for measurement error correction, with a focus on nutritional epidemiology. Stat Med. 2014/02/06. 2014;33(12):2137– 55.

32. Cohen J. Weighted kappa: nominal scale agreement with provision for scaled disagreement or partial credit. Psychol Bull. 1968;70:213–20.

33. Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. Lancet. 1986;1(8476):307–10.

34. Bland JM, Altman DG. Comparing methods of measurement: why plotting difference against standard method is misleading. Lancet (London, England). 1995;346(8982):1085–7.

35. World Health Organization. Global Recommendations on Physical Activity for Health. Geneva; 2010. Available from: https://www.ncbi.nlm.nih.gov/books/NBK305057/.

36. Rothnie KJ, Su B, Newson R, Quint JK, Soljak M. COPD prevalence model for small populations: Technical Document produced for Public Health England. 2019. Available from: https://fingertips.phe.org.uk/documents/COPD-prevalence-model-Technical-Document-v1.2-0fc (1).docx.

37. 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.

38. Hansen AW, Dahl-Petersen I, Helge JW, Brage S, Gronbaek M, Flensborg-Madsen T. Validation of an Internet-based long version of the international Physical Activity questionnaire in Danish adults using combined accelerometry and heart rate monitoring. J Phys Act Health. 2014;11(3):654– 64.

39. Matthews CE, Kozey Keadle S, Moore SC, Schoeller DS, Carroll RJ, Troiano RP, et al. Measurement of active and sedentary behavior in context of large epidemiologic studies. Med Sci Sports Exerc. 2018;50(2):266–76.

40. Piercy KL, Troiano RP, Ballard RM, Carlson SA, Fulton JE, Galuska DA, et al. The Physical Activity Guidelines for Americans. JAMA. 2018;320(19):2020–8. https://doi.org/10.1001/jama.2018.14854.

41. Keadle SK, Kravitz ES, Matthews CE, Tseng M, Carroll RJ. Development and testing of an integrated score for physical behaviors. Med Sci Sports Exerc. 2019;51(8):1759–66.

42. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. Am J Epidemiol. 2017;186(9):1026–34.

Publisher's Note