Network Harmonization of Physical Activity Variables Through Indirect Validation

**Matthew Pearce, Tom R.P. Bishop, Stephen Sharp, Kate Westgate,
Michelle Venables, Nicholas J. Wareham, and Søren Brage**

University of Cambridge

Harmonization of data for pooled analysis relies on the principle of inferential equivalence between variables from different sources. Ideally, this is achieved using models of the direct relationship with gold standard criterion measures, but the necessary validation study data are often unavailable. This study examines an alternative method of network harmonization using indirect models. Starting methods were self-report or accelerometry, from which we derived indirect models of relationships with doubly labelled water (DLW)-based physical activity energy expenditure (PAEE) using sets of two bridge equations via one of three intermediate measures. Coefficients and performance of indirect models were compared to corresponding direct models (linear regression of DLW-based PAEE on starting methods). Indirect model beta coefficients were attenuated compared to direct model betas (10%–63%), narrowing the range of PAEE values; attenuation was greater when bridge equations were weak. Directly and indirectly harmonized models had similar error variance but most indirectly derived values were biased at group-level. Correlations with DLW-based PAEE were identical after harmonization using continuous linear but not categorical models. Wrist acceleration harmonized to DLW-based PAEE via combined accelerometry and heart rate sensing had the lowest error variance (24.5%) and non-significant mean bias 0.9 (95%CI: −1.6; 3.4) kJ·day^{-1}·kg^{-1}. Associations between PAEE and BMI were similar for directly and indirectly harmonized values, but most fell outside the confidence interval of the criterion PAEE-to-BMI association. Indirect models can be used for harmonization. Performance depends on the measurement properties of original data, variance explained by available bridge equations, and similarity of population characteristics.

Keywords: calibration, DLW, epidemiology, inferential equivalence

Harmonization of exposure and outcome variables is an essential step when integrating different sources of data for the same analysis, such as in meta-analysis of published results, pooled or federated meta-analysis of individual-level data, and global surveillance of risk factors for disease. Analyses of this nature which use information from multiple sources are often constrained by the quality and compatibility of the original data (Fortier et al., 2010). Harmonization aims to bring together various types and levels of data which represent the same underlying construct (e.g., physical activity, energy intake, body fat percentage, etc.) in order to achieve compatibility when methods vary between studies or study phases (Granda & Blasczyk, 2011). The process does not strictly require that precisely the same original collection and processing methods are employed in each study (Fortier et al., 2010), but the harmonized data should be "inferentially equivalent" (i.e., their format, function, and meaning are the same) (Atkin et al., 2017). This inferential equivalence will depend upon the scientific context and the type of analysis being undertaken.

A common harmonization approach is conversion to the level of the least detailed information, for example transformation of continuous data to a binary categorization of low vs. high physical activity level (Kilpelainen et al., 2011). However, this approach loses the resolution of the more detailed data, and may therefore limit the power and scope of subsequent analyses. It is also unclear how well variables harmonized in this way relate to the latent true value of the exposure. An alternative approach to harmonization is to restrict analyses to only those studies which have assessed and expressed the exposure and outcome in the desired way. This maintains the detail of the contributing data, but—as highlighted by Aune, Norat, Leitzmann, Tonstad, and Vatten (2015)—greatly reduces the proportion of the available data that can be included in evidence synthesis. At best, this leads to loss of power. At worst, this leads to bias if the studies that are included with optimal data have specific characteristics.

Another approach to harmonization is to use validation studies which report the statistical (e.g., regression) models of relationships between values from the less precise methods and the latent true level of exposure, as assessed by a construct-specific gold-standard criterion method. A direct model permits transformation of original data to the desired harmonized format. The problem is that this mapping approach is often not possible because the ideal validation study employing gold-standard criterion methods has either not been conducted, does not report the necessary model, or is not applicable to the population or setting in question. This limitation may be more common in particular populations or settings, such as those in which the feasibility or cost of gold (or even silver) standard methods is prohibitive. Consequently, some populations

and settings may be studied with unsatisfactorily harmonized data or excluded from analyses altogether.

When the ideal direct model is unavailable, a potential solution may be to use a combination of models in the same network, whereby estimates from the less precise but more feasible field method are mapped to a criterion using "bridge equations" which form an indirect route (Figure 1). This concept is analogous to network meta-analysis in which multiple comparisons can be inferred despite not being directly tested (Lu & Ades, 2004). "Network harmonization" adopts similar principles in that it utilizes existing, ideally published, validity data to derive a new indirect model without the requirement to conduct *de novo* fieldwork. By utilizing a combination of published bridge equations and existing datasets, this study examines such an approach by comparing the inferential equivalence of data harmonized to a gold standard format via both direct and indirect models.

Methods

Study Outline

We present a hypothetical analysis task requiring harmonization of, at-first-glance, incompatible starting point values to the format of values arising from a target criterion. One could use a direct model between the two sets of data to complete this task, but as mentioned above often an alternative approach is needed. Here, we use intermediary values from a third method, for which separate links (bridge equations) to the starting values and target criterion are available. The regression models between starting point values and the intermediate (Bridge Equation 1) and between the intermediate and the target criterion (Bridge Equation 2) are used to derive the indirect model as outlined in Figure 1. In order to assess the validity of this network harmonization approach, the indirect model is compared to the direct model.

Description of Data

Here, we use the example of harmonizing four distinct sets of starting point values to the target of total daily physical activity energy expenditure (PAEE) expressed in $kJ \cdot day^{-1} \cdot kg^{-1}$ as measured using the gold standard criterion. Although the gold standard assessment method for PAEE has been used in the present study,
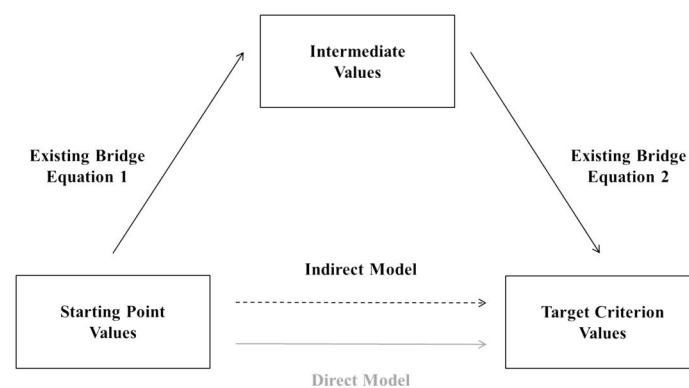

Figure 1 — Indirect modelling of the relationship between starting point values and target criterion values via intermediate values (broken black arrow). Intermediates are characterized by already established (published) relationships with both the target criterion and the starting point as indicated by the solid black arrows. The new indirect model is evaluated against the direct model (solid grey arrow).

the following methods are applicable to other target variables (e.g., moderate-intensity activity) with their own gold standard, or to target variables assessed by methods which are not of gold standard. For simplicity, we will use linear models to describe the links between data.

Starting Point Values. The four sets of starting point values were: (1) duration (minutes per day) of moderate-to-vigorous physical activity (MVPA) derived from the Recent Physical Activity Questionnaire (RPAQ) (Besson, Brage, Jakes, Ekelund, & Wareham, 2010); (2) total daily PAEE expressed in $kJ \cdot day^{-1} \cdot kg^{-1}$ derived from RPAQ; (3) the four-level categorical Cambridge Index (Golubic et al., 2014; Peters et al., 2012; Wareham et al., 2003); and (4) mean wrist acceleration expressed in milli-g (ACC_{WRIST}) (White et al., 2019).

Target Criterion Values. The gold-standard target criterion for assessing PAEE ($kJ \cdot day^{-1} \cdot kg^{-1}$) was the difference between total and resting energy expenditure as measured by the DLW method and two lab-based assessments of resting metabolic rate, coupled with allowance for the diet-induced thermogenic effect.

Intermediate Values. To derive the indirect models between the starting point data and the target criterion PAEE, one of three intermediates was used: (1) mean daily trunk acceleration in $m \cdot s^{-2}$ (ACC_{TRUNK}) (Brage, Brage, Franks, Ekelund, & Wareham, 2005); (2) total daily PAEE in $kJ \cdot day^{-1} \cdot kg^{-1}$ derived from the individually calibrated flex heart rate method (HR) (Brage et al., 2007; Spurr et al., 1988); (3) total daily PAEE in $kJ \cdot day^{-1} \cdot kg^{-1}$ from combined ACC_{TRUNK} and HR (ACCHR) (Brage et al., 2004, 2007). This allows examination of the measurement properties of the intermediate values on the final model.

Bridge Equations

To examine different aspects of the performance of network harmonization using indirect models, one of seven variations on Bridge Equation 1 was combined with one of three variations on Bridge Equation 2. In addition, we examine the performance of meta-analyzing multiple indirect routes.

If available, we used published bridge equations. If relevant equations were unavailable but correlation coefficients and basic (mean and *SD*) summary statistics were, we derived the equations through back-transformation of standardized coefficients using the corr2data STATA command. If this was not possible, we used individual-level data from existing datasets to derive equations (pretending these were published validation studies), and subsequently used them alongside existing bridge equations sourced from published work.

Bridge Equation 1. Five variations on Bridge Equation 1 were derived from the Fenland Study, a population-based cohort study of 12,435 adults born between 1950 and 1975 and registered with general practices in Cambridgeshire, United Kingdom. We randomly split this dataset into five subsamples to represent five independent validation studies. Participants attended our research facility and completed RPAQ (Besson et al., 2010) and underwent treadmill testing for individual calibration (Brage et al., 2007) whilst fitted with a chest-worn combined heart rate and movement sensor (Actiheart, CamNtech Ltd, Papworth, UK) (Brage et al., 2005). At the end of the clinical assessment, they were instructed to wear this device continuously for six days and nights and carry on with their normal behaviors. Data from RPAQ were used to derive duration (minutes per day) of MVPA by summing duration reported participating in activities with intensity > 3.0 METs (Ainsworth et al., 2011),

estimates of PAEE calculated as frequency × duration × intensity (Besson et al., 2010), and the four-level Cambridge Index (Golubic et al., 2014; Wareham et al., 2003). The ACC_{TRUNK} signal from the combined heart and movement sensor was used in the format of mean daily trunk acceleration in m·s^{-2}, while the HR signal together with treadmill test data was used to derive an individually calibrated estimate of PAEE as previously described (Brage et al., 2007, 2015). The two signals were also combined (ACCHR) to predict PAEE using branched equation modelling (Brage et al., 2004). The Fenland Study was approved by the Health Research Authority National Research Ethics Service Committee East of England-Cambridge Central, and participants provided written informed consent. Five variations on Bridge Equation 1 using self-report as the starting point were derived from the linear regression of: 1) ACC_{TRUNK} on RPAQ MVPA; (2) HR PAEE on RPAQ MVPA; (3) ACCHR PAEE on RPAQ MVPA; (4) ACCHR PAEE on RPAQ PAEE; (5a) ACCHR PAEE on RPAQ Cambridge Index. To examine whether indirect harmonization is robust to variations in measurement protocol and population (i.e., a potentially non-ideal bridge equation), we used an additional linear regression of (5b) ACCHR PAEE on RPAQ Cambridge Index. This was derived from the similar short European Prospective Investigation into Cancer and Nutrition Study Physical Activity Questionnaire (short EPIC-PAQ) administered in the EPIC cohort across 10 European countries (Peters et al., 2012; Wareham et al., 2003) denoted by $ACCHR_{EUROPE}$.

To contrast the harmonization process of self-report data as starting points with that of objective data as the starting point, two additional AC bridge equations derived from the linear regression of: (6) ACC_{TRUNK} on ACC_{WRIST}; and (7) ACCHR PAEE on ACC_{WRIST}, were obtained from published work (White, Westgate, Wareham, & Brage, 2016).

Bridge Equation 2.   Three bridge equations were obtained from published data (Brage et al., 2015) using the linear regression of DLW-based total daily PAEE expressed in kJ·day^{-1}·kg^{-1} on: (1) ACC_{TRUNK}; (2) HR PAEE; and (3) ACCHR PAEE, in doing so linking to the prediction output from a Bridge Equation 1 described above.

Direct Model

The inferential equivalence of harmonized PAEE values was assessed using gold standard DLW-based PAEE values from the UK Biobank Validation Study (BBVS) reported in detail elsewhere (White et al., 2019). Briefly, rate of carbon dioxide production (rCO_2) was measured using the 10-day DLW method of Schoeller et al. (1986) and converted to total energy expenditure (TEE) using the energy equivalents of CO_2 of Elia and Livesey (1988) in 100 participants. Resting metabolic rate (RMR) was measured on two separate days during clinic visits with a fifteen-minute rest test by respired gas analysis (OxyconPro, Jaeger, Germany), and scaled by a factor of 0.94 to account for RMR measurements being conducted in the afternoon rather than the morning (Haugen, Melanson, Tran, Kearney, & Hill, 2003). The closest measurement value (visit 1, visit 2, or their mean) by proximity to the within-person median of predictions of RMR using three equations (Henry, 2005; Nielsen et al., 2000; Watson et al., 2014) was used in analysis. Total daily REE was calculated, with an additional adjustment of sleeping metabolic rate being 5% lower than awake resting metabolic rate (Goldberg, Prentice, Davies, & Murgatroyd, 1988). Diet-induced thermogenesis was estimated using macronutrient intake assessed by food frequency questionnaire as previously described (Brage et al., 2015; Jequier, 2002). The REE and diet-induced

thermogenesis were subtracted from TEE and divided by body mass yielding an estimate of total daily PAEE in kJ·day^{-1}·kg^{-1}.

The four variations on the starting point values described above were replicated in BBVS so that four corresponding direct models predicting the target criterion DLW PAEE could be derived. Participants completed RPAQ and the raw data were used to derive duration of MVPA, PAEE, and the four-level Cambridge Index. In addition, participants were fitted with a tri-axial accelerometer (AX3, Axivity, Newcastle, UK) on the wrist for 9 days and nights whilst continuing with their usual activities. The ACC_{WRIST} signal was used to approximate acceleration as a result of human movement and expressed in milli-g (van Hees et al., 2013). Ethical approval for the study was obtained from Cambridge University Human Biology Research Ethics Committee (Ref: HBREC/2015.16). All participants provided written informed consent.

Deriving the Indirect Model

Beta and alpha coefficients for the indirect models were derived by substituting Bridge Equation 1 into Bridge Equation 2 to give Formula 3:

1. Intermediate Values = $Beta_1$ × Starting Point Values + $Alpha_1$
2. Target Criterion Values = $Beta_2$ × Intermediate Values + $Alpha_2$
3. Target Criterion Values = $Beta_2$ × ($Beta_1$ × Starting Point Values + $Alpha_1$) + $Alpha_2$.

Formula 3 simplifies to give the following formulae for deriving the new coefficients:

$$Alpha_{Indirect} = Alpha_1 * Beta_2 + Alpha_2$$

$$Beta_{Indirect} = Beta_1 * Beta_2.$$

These formulae provide parameter estimates for the indirect model coefficients but not their standard errors. To propagate the uncertainty in the parameter estimates from Bridge Equation 1 and Bridge Equation 2 to the new indirect model, 10,000 values of each parameter were sampled from a normal distribution with mean equal to the observed parameter estimate and standard deviation equal to the standard error of that parameter estimate; the formulae above were then applied to the sampled values. The means and standard deviations of the resulting distributions for $Beta_{Indirect}$ and $Alpha_{Indirect}$ were used as the coefficient point estimates and standard errors, respectively.

For the indirect model using the categorical Cambridge Index derived from RPAQ, the categorical data were replaced by one constant and three dummy variables to represent four levels. The above steps were then applied in the same way, but repeated for each of the four values of $Beta_{Indirect}$.

We meta-analyzed the newly derived beta and alpha coefficients from each of three indirect models predicting PAEE from duration of MVPA, thus generating a fifth indirect combined prediction equation using all available information; this represents the scenario where harmonization is performed using multiple validation studies of the same instrument.

Analysis

The inferential equivalence of each permutation of the indirect model was assessed alongside an equivalent direct model derived from the linear regression of target criterion PAEE data on corresponding starting point data available for 100 participants in the BBVS. PAEE was predicted from four types of starting point data using direct and indirect models and compared with values from

the observed criterion PAEE (i.e., the "true" PAEE exposure using the DLW method) by calculating the mean bias and 95% limits of agreement, root mean square error (RMSE), and Spearman correlation to assess the similarity with which individuals were ranked. Note in this evaluation scenario, the mean bias of directly mapped relationships is always zero. We derived the theoretical combined explained variance as the product of the r^2 values from the two linear bridge equations for each indirect model.

Finally, to demonstrate utility, we examined the associations between all PAEE estimates and body mass index (BMI) using multivariable linear regression adjusted for age and sex in a subset of 1695 participants in the Fenland Study. All data processing and analyses were performed in STATA/SE 14.2 (StataCorp, TX, USA).

Results

The characteristics of participants from each of the sources of data are described in Table 1. The participants in the Brage et al. (2015) study were younger and more active with lower BMI than participants in BBVS and the Fenland Study, including the subset reported in White et al. (2016). Participants in EPIC were less active than those in the Fenland Study. The criterion value of PAEE from the DLW method in the BBVS had a mean (SD) of 49.7 (16.2) $kJ \cdot kg^{-1} \cdot day^{-1}$ and a range of 8.6 to 90.8 $kJ \cdot kg^{-1} \cdot day^{-1}$.

The combinations of Bridge Equation AC and Bridge Equation CB used to generate the indirect model are shown in Table 2 alongside their r^2 values. The newly derived indirect models are plotted alongside the direct models in Figure 2.

Table 3 reports the coefficients and performance of models predicting target criterion DLW-based PAEE ($kJ \cdot kg^{-1} \cdot day^{-1}$) from a continuous estimate of MVPA duration (minutes$\cdot day^{-1}$) derived from RPAQ. The beta coefficients for indirect models were attenuated compared with the direct model beta (also see Figure 2). The correlation between the predicted and DLW-based PAEE values

was preserved irrespective of harmonization method; however, at group-level the values from ACCHR and HR indirect models were significantly biased. Each model was characterized by a narrowing of the range of PAEE values, and this was particularly pronounced when using ACC_{TRUNK} values as the intermediate method, the indirect model with the smallest combined explained variance, most attenuated beta coefficient, and widest limits of agreement. The HR indirect model resulted in predictions with the largest RMSE while the ACC_{TRUNK} and ACCHR indirect model predictions were similarly precise. Coefficients generated by meta-analyzing the three indirect models had smaller standard errors and performance was similar to that of harmonization via ACCHR.

Table 4 reports the coefficients and performance of models predicting DLW-based PAEE ($kJ \cdot kg^{-1} \cdot day^{-1}$) from a continuous estimate of total daily PAEE ($kJ \cdot kg^{-1} \cdot day^{-1}$) derived from RPAQ. Compared with the raw values of RPAQ-derived PAEE, the indirect and direct models reduced group-level mean bias and approximately halved RMSE. As for MVPA estimates, the correlation between RPAQ PAEE values and criterion PAEE was maintained regardless of whether values were raw, directly or indirectly harmonized. The ranges of the indirectly and directly harmonized values were much reduced compared with those from the raw RPAQ and those of the criterion PAEE.

Table 5 reports the coefficients and performance of models predicting DLW-based PAEE ($kJ \cdot kg^{-1} \cdot day^{-1}$) using the categorical Cambridge Index derived from RPAQ. Correlations with criterion DLW-based PAEE were weaker, and differed for direct and indirect harmonization. For the direct model, the value of PAEE assigned to being "inactive" was greater than that assigned to being "moderately inactive" and "moderately active"; the indirect model values of PAEE for each category were ordered more intuitively (also see Figure 2). RMSEs for the indirectly and directly harmonized values were similar, and also similar to RMSEs using continuous RPAQ data described above, however the group-level values were again biased when using the Fenland Study data for

Table 1   Participant Characteristics

	Fenland Study	White et al. (2016)	Peters et al. (2012)	Brage et al. (2015)	BBVS
N	10 602	1050	1941	46	100
Age (years)	48 (7)	50 (7)	53 (8)	34 (9)	54 (7)
Female (%)	52	48	70	50	50
BMI ($kg \cdot m^{-2}$)	27 (5)	26 (4)	26 (4)	25 (3)	27 (3)
ACC_{TRUNK} ($m \cdot s^{-2}$)	.126 (.055)	.125 (.055)	–	.237 (.090)	–
HR PAEE ($kJ \cdot day^{-1} \cdot kg^{-1}$)	71 (43)	–	–	67 (42)	–
ACCHR PAEE ($kJ \cdot day^{-1} \cdot kg^{-1}$)	55 (22)	–	44 (16)	69 (25)	–
RPAQ MVPA (minutes$\cdot day^{-1}$)	112 (152)	–	–	–	128 (146)
RPAQ PAEE ($kJ \cdot day^{-1} \cdot kg^{-1}$)	38 (32)	–	–	–	40 (29)
CAM Inactive (%)	13	–	15	–	7
CAM Moderately inactive (%)	54	–	33	–	37
CAM Moderately active (%)	28	–	28	–	30
CAM Active (%)	17	–	24	–	26
DLW-based PAEE ($kJ \cdot day^{-1} \cdot kg^{-1}$)	–	–	–	66 (24)	50 (16)
ACC_{WRIST} (milli-g)	–	48 (11)	–	–	43 (10)*

Note. Values are mean (SD) unless specified otherwise. *Abbreviations:* ACC_{TRUNK} = mean daily trunk acceleration; ACC_{WRIST} = mean daily high pass filter vector magnitude wrist acceleration signal; ACCHR = combined heart rate and movement sensing; BBVS = Biobank Validation Study; BMI = body mass index; CAM = Cambridge Index; DLW = doubly labelled water; HR = heart rate; MVPA = moderate-to-vigorous physical activity; PAEE = physical activity energy expenditure; RPAQ = Recent Physical Activity Questionnaire.
*$n = 97$.

Table 2   Bridge Equations Used to Derive Indirect Models

Bridge Equation	Starting Point Value	Intermediate Value	Target Criterion Value	N	β (SE)	α (SE)	r^2
Harmonization of RPAQ MVPA via ACC$_{TRUNK}$							
1	RPAQ MVPA (minutes·day^{-1})	ACC$_{TRUNK}$ (m·s^{-2})	–	2121	$5.84 \cdot 10^{-5}$ $(7.9 \cdot 10^{-6})$	.1199 (.0015)	.02
2	–	ACC$_{TRUNK}$ (m·s^{-2})	DLW PAEE (kJ·day^{-1}·kg^{-1})	46	165 (32)	26.7 (8.2)	.37
Harmonization of RPAQ MVPA via PAEE from HR							
1	RPAQ MVPA (minutes·day^{-1})	HR PAEE (kJ·day^{-1}·kg^{-1})	–	2121	.0840 (.0061)	60.9 (1.2)	.08
2	–	HR PAEE (kJ·day^{-1}·kg^{-1})	DLW PAEE (kJ·day^{-1}·kg^{-1})	46	.34 (.07)	42.7 (5.8)	.34
Harmonization of RPAQ MVPA via PAEE from ACCHR							
1	RPAQ MVPA (minutes·day^{-1})	ACCHR PAEE (kJ·day^{-1}·kg^{-1})	–	2120	.0390 (.0030)	50.69 (.57)	.07
2	–	ACCHR PAEE (kJ·day^{-1}·kg^{-1})	DLW PAEE (kJ·day^{-1}·kg^{-1})	46	.66 (.11)	20.0 (8.1)	.45
Harmonization of RPAQ PAEE via PAEE from ACCHR							
1	RPAQ PAEE (kJ·day^{-1}·kg^{-1})	ACCHR PAEE (kJ·day^{-1}·kg^{-1})	–	2120	.239 (.014)	45.63 (.69)	.12
2		ACCHR PAEE (kJ·day^{-1}·kg^{-1})	DLW PAEE (kJ·day^{-1}·kg^{-1})	46	.66 (.11)	20.0 (8.1)	.45
Harmonization of Cambridge Index via PAEE from ACCHR							
1	RPAQ Cambridge Index	ACCHR PAEE (kJ·day^{-1}·kg^{-1})	–	2120	*Inactive = 0; Moderately inactive = 5.6 (1.5); Moderately active = 13.0 (1.6); Active = 24.2 (1.7)	44.2 (1.3)	.12
2		ACCHR PAEE (kJ·day^{-1}·kg^{-1})	DLW PAEE (kJ·day^{-1}·kg^{-1})	46	.66 (.11)	20.0 (8.1)	.45
Harmonization of Cambridge Index via PAEE from ACCHR$_{EUROPE}$							
1	RPAQ Cambridge Index	ACCHR$_{EUROPE}$ PAEE (kJ·day^{-1}·kg^{-1})	–	1941	*Inactive = 0; Moderately inactive = 4.6 (1.1); Moderately active = 9.1 (1.1); Active = 14.8 (1.1)	36.14 (.88)	.10
2		ACCHR$_{EUROPE}$ PAEE (kJ·day^{-1}·kg^{-1})	DLW PAEE (kJ·day^{-1}·kg^{-1})	46	.66 (.11)	20.0 (8.1)	.45
Harmonization of ACC$_{WRIST}$ via ACC$_{TRUNK}$							
1	ACC$_{WRIST}$ (milli-g)	ACC$_{TRUNK}$ (m·s^{-2})	–	1050	$4.78 \cdot 10^{-3}$ $(9.0 \cdot 10^{-5})$	-.097 (.0036)	.53
2		ACC$_{TRUNK}$ (m·s^{-2})	DLW PAEE (kJ·day^{-1}·kg^{-1})	46	165 (32)	26.7 (8.2)	.37
Harmonization of ACC$_{WRIST}$ via PAEE from ACCHR							
1	ACC$_{WRIST}$ (milli-g)	ACCHR PAEE (kJ·day^{-1}·kg^{-1})	–	1050	1.232 (.012)	−6.90 (.45)	.67
2		ACCHR PAEE (kJ·day^{-1}·kg^{-1})	DLW PAEE (kJ·day^{-1}·kg^{-1})	46	.66 (.11)	20.0 (8.1)	.45

Abbreviations: ACC$_{TRUNK}$ = mean daily trunk acceleration; ACC$_{WRIST}$ = mean daily high pass filter vector magnitude wrist acceleration signal; ACCHR = combined heart rate and movement sensing; ACCHR$_{EUROPE}$ = combined heart rate and movement sensing from European population; DLW = doubly labelled water method; HR = heart rate; MVPA = moderate-to-vigorous physical activity; PAEE = physical activity energy expenditure; RPAQ = Recent Physical Activity Questionnaire; SE = standard error.
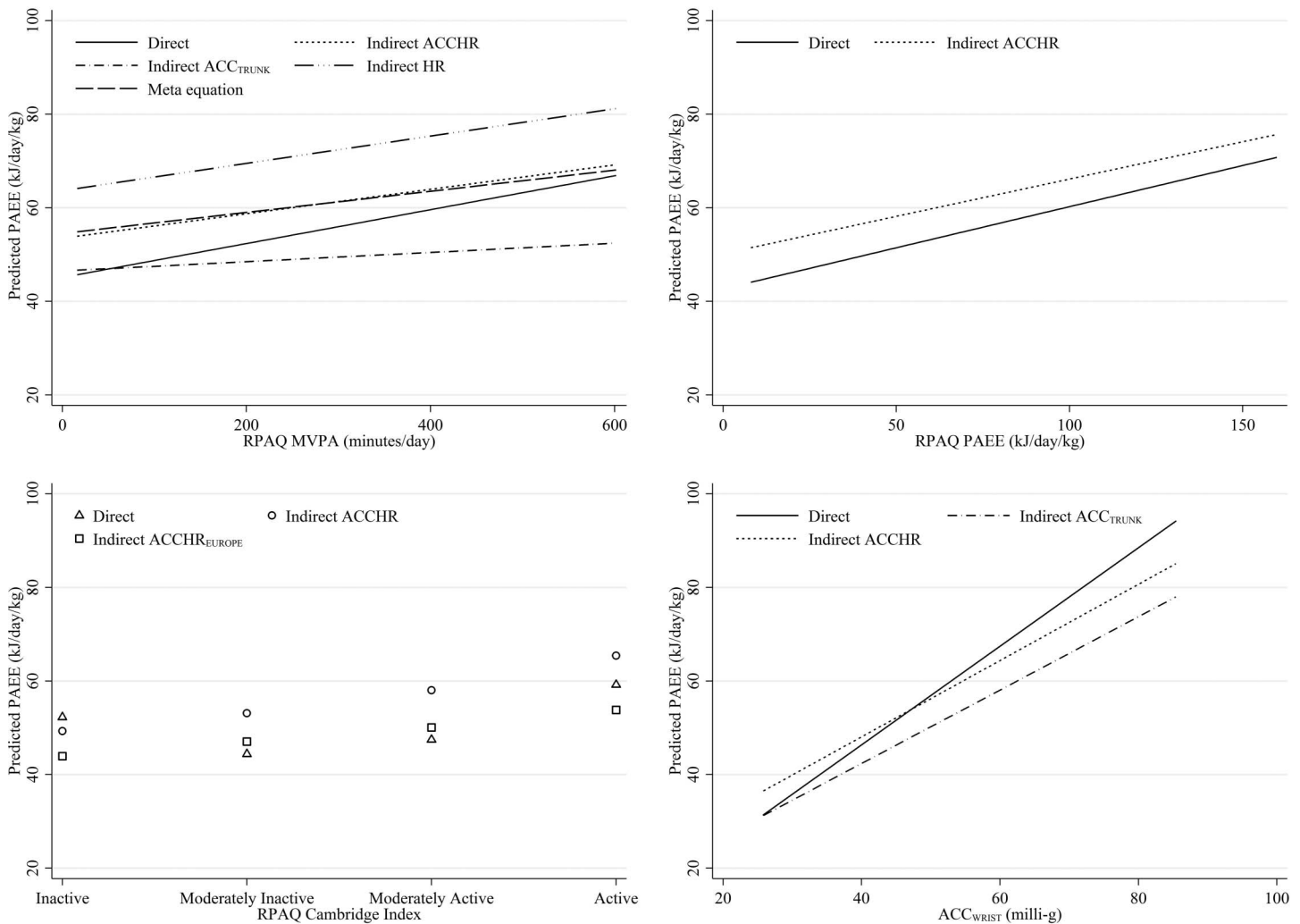*Constant and three dummy variables used to represent four-level Cambridge Index.

Figure 2 — Comparison of direct and indirect harmonization models by starting data type. *Abbreviations*: ACC_{TRUNK} = mean daily trunk acceleration; ACCHR = combined heart rate and movement sensing; $ACCHR_{EUROPE}$ = combined heart rate and movement sensing from European population; ACC_{WRIST} = mean daily high pass filter vector magnitude wrist acceleration signal; HR = heart rate; MVPA = moderate-to-vigorous physical activity; PAEE = physical activity energy expenditure; RPAQ = Recent Physical Activity Questionnaire.

Bridge Equation 1. The indirect model derived using the Bridge Equation 1 from a less active European population showed unbiased group-level estimates of PAEE.

Table 6 reports the coefficients and performance of models predicting DLW-based PAEE ($kJ·kg^{-1}·day^{-1}$) from ACC_{WRIST} (milli-g). Compared with the RPAQ-derived models described above, models based on accelerometer data as starting point and mapping via other objective methods resulted in stronger correlations of estimated PAEE with DLW-based PAEE. The RMSEs observed for both the indirect and direct models were smaller than for any of the RPAQ models, and with higher combined explained variance; the range of PAEE values was also better preserved as shown in Figure 2.

To demonstrate utility of the different harmonization approaches, we examined the associations of PAEE with BMI in the Fenland study ($n = 1695$ subsample); Figure 3 shows the beta (95% confidence intervals) of these linear associations by harmonization method and starting data format, alongside that from the silver-standard ACCHR PAEE. PAEE estimates harmonized from ACC_{WRIST} showed statistically significant inverse associations

with BMI. There was a striking difference between models for the Cambridge Index, with indirectly harmonized PAEE showing statistically significant inverse associations with BMI and the directly harmonized PAEE showing no relationship. Other associations of indirectly harmonized PAEE values with BMI were similar but more uncertain when compared to the direct equivalent. All associations using PAEE harmonized from continuous RPAQ data were weak with 95% confidence intervals crossing zero. MVPA data indirectly harmonized to PAEE via ACC_{TRUNK} values resulted in the widest confidence intervals.

Discussion

This study is the first to examine and evaluate an indirect validation technique for harmonization, an essential step in any study which combines data from different sources in the same analysis. Our findings indicate that indirect models can be employed to harmonize data to a compatible format in the absence of the ideal validation study, but that the harmonized values may still be biased at group-level, have a narrower value range compared to the

Table 3   Model Coefficients and Performance for Predicting DLW-Based PAEE from RPAQ MVPA

	Direct BBVS	Indirect via ACC$_{TRUNK}$	Indirect via HR	Indirect via ACCHR	Meta-equation
Coefficients					
β (SE)	.036 (.011)	.0099 (.0032)	.0291 (.0081)	.0261 (.0062)	.0225 (.0012)
α (SE)	45.1 (2.1)	46.5 (12.2)	63.7 (10.6)	53.5 (13.9)	54.5 (.23)
Combined explained variance	–	.007	.027	.032	–
Performance vs DLW PAEE					
Mean bias (kJ·day^{-1}·kg^{-1})[95% CI]	–	−1.9 [−5.1; 1.2]	17.7 [14.6; 20.7]	7.1 [4.1; 10.1]	7.7 [4.6; 10.7]
Mean bias (%)	–	−3.9	35.6	14.3	15.4
Limits of agreement* (kJ·day^{-1}·kg^{-1})	–	−34; 30	−13; 48	−24; 37	23; 39
RMSE (kJ·day^{-1}·kg^{-1})	15.3	15.8	23.4	16.9	17.2
RMSE (%)	30.7	32.0	47.0	34.0	34.6
Spearman correlation* (rho)	.37	.37	. 37	.37	.37
Range (kJ·day^{-1}·kg^{-1})	46; 67	47; 52	64; 81	54; 69	55; 68

Note: Combined explained variance is the product of the explained variance of the two bridge equations. It is not the explained variance (r^2) of the indirect model.
Abbreviations: ACC$_{TRUNK}$ = mean daily trunk acceleration; ACCHR = combined heart rate and movement sensing; BBVS = Biobank Validation Study; DLW = doubly labelled water; HR = heart rate; MVPA = moderate-to-vigorous physical activity; PAEE = physical activity energy expenditure; RMSE = root mean square error; RPAQ = Recent Physical Activity Questionnaire; SE = standard error.

Table 4   Model Coefficients and Performance for Predicting DLW-Based PAEE from RPAQ PAEE

	Direct BBVS	Indirect via ACCHR	Non-harmonized RPAQ PAEE
Coefficients			
β (SE)	.176 (.054)	.159 (.035)	–
α (SE)	42.7 (2.7)	50.2 (13.5)	–
Combined explained variance	–	.054	–
Performance vs DLW PAEE			
Mean bias (kJ·day^{-1}·kg^{-1}) [95%CI]	–	6.9 [3.8; 9.9]	−9.6 [−15.2; −4.0]
Mean bias (%)	–	13.7	−19.3
Limits of agreement (kJ·day^{-1}·kg^{-1})	–	−24; 38	−66; 47
RMSE (kJ·day^{-1}·kg^{-1})	15.3	16.8	29.8
RMSE (%)	30.8	33.8	60.0
Spearman correlation (rho)	.38	.38	.38
Range (kJ·day^{-1}·kg^{-1})	44; 71	51; 76	8; 160

Note: Combined explained variance is the product of the explained variance of the two bridge equations. It is not the explained variance (r^2) of the indirect model.
Abbreviations: ACCHR = combined heart rate and movement sensing; BBVS = Biobank Validation Study; DLW = doubly labelled water; PAEE = physical activity energy expenditure; RMSE = root mean square error; RPAQ = Recent Physical Activity Questionnaire; SE = standard error.

criterion, and that gains in precision are dependent upon the variance explained by the contributing bridge equations. These findings reinforce the necessity to carefully consider the inferential equivalence of harmonized data and "the truth" according to the scientific context and the purpose of the analyses being performed (Fortier et al., 2010).

For analyses such as exposure-disease associations which primarily rely on relative validity, harmonization using indirect models is beneficial in that it increases precision while retaining the correlation between the original data and estimates of the latent truth from the gold standard criterion (when all bridge equations are linear). Our results for RPAQ-derived PAEE demonstrate that even when data pre-exist in a seemingly compatible format, improvements in precision are possible following harmonization.

Furthermore, even if no harmonization need be conducted, the newly derived indirect model could be used to perform measurement error correction by regression calibration (Keogh & White, 2014).

Using indirect models revealed some unexpected advantages. When harmonizing Cambridge Index categorical data to a continuous estimate of PAEE, the direct model assigned a higher PAEE value to level 1 than both level 2 and 3, whereas the indirect model was more logically ordered by mean PAEE in kJ·kg^{-1}·day^{-1} using the relationships in the Fenland sample or the published EPIC data. The illogical ordering in the small sample likely contributed to the null PAEE-BMI association for directly harmonized PAEE, whereas the two logically ordered indirectly harmonized PAEE estimates both showed inverse relationships. Given that validation

Table 5   Model Coefficients and Performance for Predicting DLW-Based PAEE from RPAQ Cambridge Index

	Direct BBVS	Indirect via ACCHR	Indirect via ACCHR$_{EUROPE}$
Coefficients			
Cambridge Index Inactive β (SE)	0	0	0
Cambridge Index Moderately inactive β (SE)	−7.9 (6.3)	3.8 (1.6)	3.1 (1.2)
Cambridge Index Moderately active β (SE)	−4.8 (6.4)	8.7 (2.4)	6.1 (1.7)
Cambridge Index Active β (SE)	6.9 (6.5)	16.1 (3.7)	9.9 (2.4)
α (SE)	52.3 (5.8)	49.3 (13.7)	43.9 (12.6)
Combined explained variance	−	.054	.045
Performance vs DLW PAEE			
Mean bias (kJ·day^{-1}·kg^{-1}) [95%CI]	−	7.8 [4.7; 10.9]	−.2 [−3.3; 2.8]
Mean bias (%)	−	15.7	−.4
Limits of agreement (kJ·day^{-1}·kg^{-1})	−	−23; 39	−31; 31
RMSE (kJ·day^{-1}·kg^{-1})	15.0	17.2	15.5
RMSE (%)	30.1	34.6	31.2
Spearman correlation (rho)	.34	.27	.27
Range (kJ·day^{-1}·kg^{-1})	44; 59	49; 65	44; 54

Note: Combined explained variance is the product of the explained variance of the two bridge equations. It is not the explained variance (r^2) of the indirect model.
Abbreviations: ACCHR = combined heart rate and movement sensing; ACCHR$_{EUROPE}$ = combined heart rate and movement sensing from European population; BBVS = Biobank Validation Study; DLW = doubly labelled water; PAEE = physical activity energy expenditure; RMSE = root mean square error; RPAQ = Recent Physical Activity Questionnaire; SE = standard error.

Table 6   Model Coefficients and Performance for Predicting DLW-based PAEE from ACC$_{WRIST}$

	Direct BBVS	Indirect via ACC$_{TRUNK}$	Indirect via ACCHR
Coefficients			
β (SE)	1.05 (.12)	.79 (.17)	.81 (.14)
α (SE)	4.2 (5.3)	10.9 (5.6)	15.5 (7.6)
Combined explained variance	−	.196	.302
Performance vs DLW PAEE			
Mean bias (kJ·day^{-1}·kg^{-1}) [95%CI]	−	−4.8 [−7.2; −2.3]	.9 [−1.6; 3.4]
Mean bias (%)	−	−9.5	1.8
Limits of agreement (kJ·day^{-1}·kg^{-1})	−24; 24	−29; 20	−24; 25
RMSE (kJ·day^{-1}·kg^{-1})	11.9	13.1	12.2
RMSE (%)	24.0	26.4	24.5
Spearman correlation (rho)	.69	.69	.69
Range (kJ·day^{-1}·kg^{-1})	31; 94	31; 78	37; 85

Note: Combined explained variance is the product of the explained variance of the two bridge equations. It is not the explained variance (r^2) of the indirect model.
Abbreviations: ACC$_{TRUNK}$ = mean daily trunk acceleration; ACC$_{WRIST}$ = mean daily high pass filter vector magnitude wrist acceleration signal; ACCHR = combined heart rate and movement sensing; BBVS = Biobank Validation Study; DLW = doubly labelled water; PAEE = physical activity energy expenditure; RMSE = root mean square error; SE = standard error.

studies using comparisons with a gold-standard criterion are often relatively small, mapping of categorical data from questionnaires to continuous metrics may be unduly influenced by incorrectly classified individuals with extreme values. In a larger study such as a prospective cohort like Fenland, the influence of incorrectly classified individuals on category means is diminished resulting in more logically ordered group means. Thus in certain scenarios and under some feasibility constraints, it is possible that indirect models using bridge equations sourced from larger studies with silver-standard assessment are actually preferable to harmonization using a direct model from a smaller study using the gold standard. Although indirect harmonization is a potential solution in the absence of direct models, these findings also suggest investigators should make use of all available information—both direct and indirect—in a form of network harmonization analogous to network meta-analysis.

We observed that beta coefficients of indirect models were attenuated compared to those from direct models, and that the level of attenuation is related to the variance explained by the bridge equations being combined. We calculated the theoretical combined explained variance which summarises the r^2 values of the two contributing bridge equations derived in separate studies; indirect models with a higher value tended to have less attenuated beta coefficients, greater precision and less narrowed range of resulting
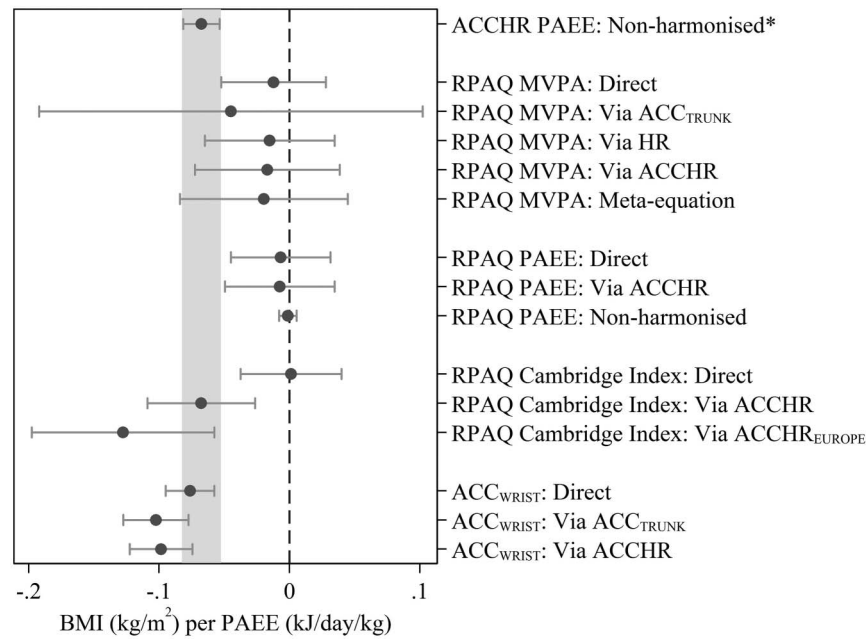
Figure 3 — Association (beta coefficients and 95% confidence intervals) between PAEE and BMI, by exposure estimation method (Fenland Study, $n = 1695$ subsample with wrist acceleration). *Note*: Associations are adjusted for age and sex. *Abbreviations*: ACC_{TRUNK} = mean daily trunk acceleration; ACCHR = combined ACC_{TRUNK} and heart rate sensing; $ACCHR_{EUROPE}$ = combined heart rate and movement sensing from European population; BMI = body mass index; ACC_{WRIST} = mean daily high-pass filtered vector magnitude wrist acceleration; HR = heart rate; PAEE = physical activity energy expenditure. *In the absence of doubly labelled water-assessed PAEE in a large cohort, a silver standard (ACCHR) has been used for comparison for the cross-sectional association.

values. Shrinking of the range of harmonized values towards the mean when compared to the original data and the criterion will have implications for dose-response analyses which aim to assess the shape of any relationship across the full exposure range. It was noticeable that the indirect model for harmonizing MVPA to PAEE via ACC_{TRUNK} had the most attenuated beta coefficient resulting in a very narrow range of PAEE values and the widest confidence intervals for the applied example of studying the PAEE association with BMI. A key finding was that associations between PAEE and BMI were similar for data harmonized using both direct and indirect models; the two techniques may therefore be inferentially equivalent but depending upon the validity of the original method, neither may result in data inferentially equivalent to the true level of exposure. The utility of all harmonization equations is influenced by the quality of the data being harmonized and the pre-existing error of the methods used in the bridge equations.

For analyses which depend upon absolute agreement between harmonized data and the latent truth, even greater caution is required as most values from indirect harmonization were biased at group-level. The bias was typically <15% and did not appear to be related to the combined explained variance. The biases observed may therefore reflect differences in the populations from which the bridge equations and the data undergoing harmonization are sourced. Perhaps most strikingly, the criterion measures of PAEE in the BBVS and Brage et al. (2015) differed on average by 22%. The Brage et al. (2015) participants were younger and had lower BMI so these are likely to be true differences, even though some of the bias observed may be due to minor differences in the criterion method used to estimate PAEE in the two studies, e.g. different RMR protocols. We examined the effect of using alternative bridge equations from potentially incompatible measurement tools and populations by deriving two indirect models for harmonizing the RPAQ Cambridge

Index to PAEE. The indirect model using a Bridge Equation AC derived from the Fenland Study resulted in biased group-level estimates, whereas the 'non-ideal' Bridge Equation AC derived from published data using the EPIC-PAQ in the less active European population did not. It is possible that the lower PAEE in this latter population counteracted any bias resulting from the higher PAEE in the second Bridge Equation from Brage et al. (2015). The contrast may also be partially attributable to differences between RPAQ and EPIC-PAQ. Any incompatibility of the populations and assessment methods from which the two bridge equations are derived must therefore be considered alongside the data being harmonized.

Further work is necessary to examine the suitability of combining bridge equations from different populations and the generalizability of the resulting indirect models, but our findings suggest that harmonization is sensitive to differences in the true level of exposure between participants in the bridge equation studies and those in the dataset being harmonized.

One potential solution to population differences could be the addition of covariates to indirect models, or, if the sample is large enough, stratification of results by covariates. However, this would require some existing published bridge equations to be re-derived, and limit future harmonization to datasets with compatible covariates. Moreover, adding covariates to the prediction likely requires that these are always adjusted for in subsequent association analyses, as associations may otherwise be driven by the covariates (i.e., confounded). In addition and on a more practical level, it is often the case that regression equations are not published, and this limits harmonization efforts despite some of the necessary fieldwork having been conducted. Populating the network of bridge equations between methods, and making this discoverable, is therefore an important task. We have added beta and alpha coefficients alongside their standard errors of the bridge equations

used in the present study to the online repository Diet, Anthropometry, and Physical Activity Measurement Toolkit (www.measurement-toolkit.org) which enables sharing of such results and which should increase the visibility and use of population-specific harmonization models.

A potential limitation of the current work is that only simple linear bridge equations were combined and that the participants in the studies used were all from the same region of the UK; the generalizability of the technique to other populations therefore remains unclear. While it was a strength of this work that we conducted indirect harmonization with several different combinations of methods, there are many more methods in use and the viability of this process with other combinations, particularly when both bridge equations have weak r^2 is difficult to judge. In this study we were able to assess the validity of indirect harmonization using the equivalent direct relationship from a study with gold-standard measures. This will not be possible in most other scenarios as direct models would be unavailable—the exact problem that indirect harmonization tries to resolve. In these circumstances it is not possible to assess the precision or mean bias of indirectly harmonized estimates; the variance explained by the two bridge equations provides an indication, but additional research is required to examine this formally.

Conclusion

In summary, indirect models can harmonize data to compatible format when direct models are not available, and can therefore improve the inclusivity and resolution of data in analyses integrating information from different sources. This network harmonization approach has greater validity when the original data and bridge equations are stronger (i.e., more variance explained). Further work is required to examine the sources of bias, address difficulties when generalizing population-specific equations, and increase both the number and discoverability of bridge equations in the network.

Acknowledgments

References

Ainsworth, B.E., Haskell, W.L., Herrmann, S.D., Meckes, N., Bassett Jr., D.R., Tudor-Locke, C., . . . Leon, A.S. (2011). 2011 Compendium of Physical Activities: a second update of codes and MET values. *Medicine & Science in Sports & Exercise, 43*(8), 1575–1581. PubMed ID: 21681120 doi:10.1249/MSS.0b013e31821ece12

Atkin, A.J., Biddle, S.J.H., Broyles, S.T., Chinapaw, M., Ekelund, U., Esliger, D.W., . . . van Sluijs, E.M.F. (2017). Harmonising data on the correlates of physical activity and sedentary behaviour in young people: methods and lessons learnt from the international Children's Accelerometry database (ICAD). *International Journal of Behavioral Nutrition and Physical Activity, 14*(1), 174. PubMed ID: 29262830 doi:10.1186/s12966-017-0631-7

Aune, D., Norat, T., Leitzmann, M., Tonstad, S., & Vatten, L.J. (2015). Physical activity and the risk of type 2 diabetes: a systematic review and dose-response meta-analysis. *European Journal of Epidemiology, 30*(7), 529–542. PubMed ID: 26092138 doi:10.1007/s10654-015-0056-z

Besson, H., Brage, S., Jakes, R.W., Ekelund, U., & Wareham, N.J. (2010). Estimating physical activity energy expenditure, sedentary time, and physical activity intensity by self-report in adults. *The American Journal of Clinical Nutrition, 91*(1), 106–114. PubMed ID: 19889820 doi:10.3945/ajcn.2009.28432

Brage, S., Brage, N., Franks, P.W., Ekelund, U., & Wareham, N.J. (2005). Reliability and validity of the combined heart rate and movement sensor Actiheart. *European Journal of Clinical Nutrition, 59*(4), 561–570. PubMed ID: 15714212 doi:10.1038/sj.ejcn.1602118

Brage, S., Brage, N., Franks, P.W., Ekelund, U., Wong, M.Y., Andersen, L.B., . . . Wareham, N.J. (2004). Branched equation modeling of simultaneous accelerometry and heart rate monitoring improves estimate of directly measured physical activity energy expenditure. *Journal of Applied Physiology, 96*(1), 343–351. PubMed ID: 12972441 doi:10.1152/japplphysiol.00703.2003

Brage, S., Ekelund, U., Brage, N., Hennings, M.A., Froberg, K., Franks, P.W., & Wareham, N.J. (2007). Hierarchy of individual calibration levels for heart rate and accelerometry to measure physical activity. *Journal of Applied Physiology, 103*(2), 682–692. PubMed ID: 17463305 doi:10.1152/japplphysiol.00092.2006

Brage, S., Westgate, K., Franks, P.W., Stegle, O., Wright, A., Ekelund, U., & Wareham, N.J. (2015). Estimation of free-living energy expenditure by heart rate and movement sensing: a doubly-labelled water study. *PLoS ONE, 10*(9), e0137206. PubMed ID: 26349056 doi:10.1371/journal.pone.0137206

Elia, M., & Livesey, G. (1988). Theory and validity of indirect calorimetry during net lipid synthesis. *The American Journal of Clinical Nutrition.* doi: 10.1093/ajcn/47.4.591

Fortier, I., Burton, P.R., Robson, P.J., Ferretti, V., Little, J., L'Heureux, F., . . . Hudson, T.J. (2010). Quality, quantity and harmony: the DataSHaPER approach to integrating data across bioclinical studies. *International Journal of Epidemiology, 39*(5), 1383–1393. PubMed ID: 20813861 doi:10.1093/ije/dyq139

Goldberg, G.R., Prentice, A.M., Davies, H.L., & Murgatroyd, P.R. (1988). Overnight and basal metabolic rates in men and women. *European Journal of Clinical Nutrition, 42*(2), 137–144. PubMed ID: 3378547

Golubic, R., May, A.M., Benjaminsen Borch, K., Overvad, K., Charles, M.-A., Diaz, M.J.T., . . . Brage, S. (2014). Validity of electronically administered recent physical activity questionnaire (RPAQ) in ten European countries. *PLoS ONE, 9*(3), e92829. PubMed ID: 24667343 doi:10.1371/journal.pone.0092829

Granda, P., & Blasczyk, E. (2011). *Data harmonization. Guidelines for best practice in cross-cultural surveys.* 3rd ed. Retrieved from http://ccsg.isr.umich.edu/index.php/chapters/data-harmonization-chapter

Haugen, H.A., Melanson, E.L., Tran, Z.V, Kearney, J.T., & Hill, J.O. (2003). Variability of measured resting metabolic rate. *The American*

Journal of Clinical Nutrition, 78(6), 1141–1144. PubMed ID: 14668276 doi:10.1093/ajcn/78.6.1141

Henry, C.J. (2005). Basal metabolic rate studies in humans: measurement and development of new equations. *Public Health Nutrition, 8*(7A), 1133–1152. PubMed ID: 16277825 doi:10.1079/PHN2005801

Jequier, E. (2002). Pathways to obesity. *International Journal of Obesity and Related Metabolic Disorders, 26*(Suppl. 2), S12–S17. doi:10.1038/sj.ijo.0802123

Keogh, R.H., & White, I.R. (2014). A toolkit for measurement error correction, with a focus on nutritional epidemiology. *Statistics in Medicine, 33*(12), 2137–2155. PubMed ID: 24497385 doi:10.1002/sim.6095

Kilpelainen, T.O., Qi, L., Brage, S., Sharp, S.J., Sonestedt, E., Demerath, E., . . . Loos, R.J. (2011). Physical activity attenuates the influence of FTO variants on obesity risk: a meta-analysis of 218, 166 adults and 19, 268 children. *PLoS Med, 8*(11), e1001116. PubMed ID: 22069379 doi:10.1371/journal.pmed.1001116

Lu, G., & Ades, A.E. (2004). Combination of direct and indirect evidence in mixed treatment comparisons. *Statistics in Medicine, 23*(20), 3105–3124. PubMed ID: 15449338 doi:10.1002/sim.1875

Nielsen, S., Hensrud, D.D., Romanski, S., Levine, J.A., Burguera, B., & Jensen, M.D. (2000). Body composition and resting energy expenditure in humans: role of fat, fat-free mass and extracellular fluid. *International Journal of Obesity and Related Metabolic Disorders, 24*(9), 1153–1157. PubMed ID: 11033984 doi:10.1038/sj.ijo.0801317

Peters, T., Brage, S., Westgate, K., Franks, P.W., Gradmark, A., Tormo Diaz, M.J., . . . Wareham, N. (2012). Validity of a short questionnaire to assess physical activity in 10 European countries. *European Journal of Epidemiology, 27*(1), 15–25. PubMed ID: 22089423 doi:10.1007/s10654-011-9625-y

Schoeller, D.A., Ravussin, E., Schutz, Y., Acheson, K.J., Baertschi, P., & Jequier, E. (1986). Energy expenditure by doubly labeled water: validation in humans and proposed calculation. *American Journal of Physiology, 250*(5), R823–R830. doi:10.1152/ajpregu.1986.250.5.R823

Spurr, G.B., Prentice, A.M., Murgatroyd, P.R., Goldberg, G.R., Reina, J.C., & Christman, N.T. (1988). Energy expenditure from minute-by-minute heart-rate recording: comparison with indirect calorimetry. *The American Journal of Clinical Nutrition, 48*(3), 552–559. PubMed ID: 3414570 doi:10.1093/ajcn/48.3.552

van Hees, V.T., Gorzelniak, L., Dean Leon, E.C., Eder, M., Pias, M., Taherian, S., . . . Brage, S. (2013). Separating movement and gravity components in an acceleration signal and implications for the assessment of human daily physical activity. *PLoS ONE, 8*(4), e61691. PubMed ID: 23626718 doi:10.1371/journal.pone.0061691

Wareham, N.J., Jakes, R.W., Rennie, K.L., Schuit, J., Mitchell, J., Hennings, S., & Day, N.E. (2003). 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 Nutrition, 6*(4), 407–413. PubMed ID: 12795830 doi:10.1079/PHN2002439

Watson, L.P., Raymond-Barker, P., Moran, C., Schoenmakers, N., Mitchell, C., Bluck, L., . . . Murgatroyd, P.R. (2014). An approach to quantifying abnormalities in energy expenditure and lean mass in metabolic disease. *European Journal of Clinical Nutrition, 68*(2), 234–240. PubMed ID: 24281313 doi:10.1038/ejcn.2013.237

White, T., Westgate, K., Hollidge, S., Venables, M., Olivier, P., Wareham, N., & Brage, S. (2019). Estimating energy expenditure from wrist and thigh accelerometry in free-living adults: a doubly labelled water study. *International Journal of Obesity* . epub ahead of print. PubMed ID:30940917 doi:10.1038/s41366-019-0352-x

White, T., Westgate, K., Wareham, N.J., & Brage, S. (2016). Estimation of physical activity energy expenditure during free-living from wrist accelerometry in UK adults. *PLoS ONE, 11*(12), e0167472. PubMed ID: 27936024 doi:10.1371/journal.pone.0167472