Appendix V Measuring physical activity in adults using the Recent Physical Activity Questionnaire (RPAQ)

Jennifer Mindell

V.1 Background

The principal purpose of estimating energy expenditure in a dietary survey is so that energy intake can be analysed together with energy expenditure, to assess the degree of under-reporting of food and drink intake. Physical activity is measured in the NDNS RP primarily as a proxy for energy expenditure, although this cannot be determined accurately except in a small sub-sample using doubly-labelled water (DLW). The purpose was not to measure the population's activity levels per se. Resting metabolism and physical activity account for 90% of energy use on average. Despite their many limitations, questionnaires have often been used in the past to assess energy expenditure as this is an inexpensive and convenient method.

V.2 Choice of instrument

In Year 1 of the NDNS RP, physical activity was measured among participants aged 16 years and over using a bespoke questionnaire that had been developed for and tested in the Comparison Study that preceded the mainstage of the NDNS RP.[1] This new questionnaire had been constructed in an attempt to estimate more accurately individual-level energy expenditure (EE). This produced data of a comparable validity to many other studies[2] but had the drawback of being very time-consuming. It therefore had the potential to impact adversely on participants' willingness to continue with other parts of the NDNS RP and was costly to administer but it did not provide the anticipated benefit of greater individual-level accuracy.

These results from the Comparison Study became available during Year 1 of the NDNS RP; a decision was therefore made to use the RPAQ (Recent Physical Activity Questionnaire) instead of the bespoke questionnaire from Year 2 onwards. The RPAQ produces slightly better results and is significantly shorter.[3] Results of the validation of RPAQ showed a significant correlation of r=0.43 and a root mean square error[4] (RMSE) value of 24.7 kJ/day/kg for physical activity energy expenditure.[3, 5] This compares favourably with the results of the Comparison Study questionnaires, which had r=0.30 (r^2=0.09) overall and RMSE 36.5kJ/d/kg for participants aged 16 and over for PAEE.

The RPAQ is a self-completion questionnaire designed to assess physical activity in four domains:

- home (watching television, using a computer, climbing stairs)
- work

- commuting (by car, public transport, cycling, and/or walking)
- leisure activities

The RPAQ was developed by the MRC Epidemiology Unit, Cambridge. This has recently been validated against DLW in 50 adults aged 20 years and older.[3] Results of the validation of RPAQ showed a significant correlation of r=0.43 and a RMSE value of 24.7 kJ/day/kg for PAEE.[3] Use of RPAQ in the NDNS RP also allows comparison with other large cohort studies, such as MRC Fenland Study (n>5,000).

For the NDNS RP, the RPAQ was modified by the inclusion of additional questions designed to capture sun exposure during activities outside, in relation to vitamin D synthesis.

V.3 Data collection using RPAQ (from Year 2 onwards)

At the diary pick-up visit, the interviewer gave participants aged 16 years and over the self-completion RPAQ. Participants answered the RPAQ while the interviewer was present, and it was taken away by the interviewer. The completed questionnaires were sent to NatCen for manual data entry, using a programme with soft and hard checks to reduce keying errors. Duplicate data entry was conducted with the questionnaire responses keyed twice and then edited.

V. 4 Processing the raw RPAQ responses

V.4.1 Introduction

The MRC Epidemiology Unit, Cambridge, provided a Stata dofile to impute missing data and calculate summary variables to estimate the total physical activity of each participant.[6]

V.4.2 Imputing missing variables

Because estimation of the summary variables requires a value for each of the types of activity included, it was important to impute missing values: failure to do so could have resulted in large numbers of participants not having physical activity summary variables even if each had only a single missing variable. Median values within the dataset were used for imputation of missing values.

The variables were imputed using the amalgamated data from Years 2, 3, and 4, as this creates the most robust foundation for the imputation.

- for each set of variables, a missing data variable was first created, with a value of 2 if there were no missing values for any of the relevant variables and a value of 1 if all the variables in a section had missing values

- individuals who reported being unemployed, or who were employed but recorded no work activities, were considered to have missing values for work-related variables

V.4.3 Deriving variables

- where answers were categorised using a series of ranges of duration or frequency, the mid-point for each range was assigned to participants reporting that category of response
- for frequency of commuting by each transport mode, a value of 0 was assigned for 'never or rarely', 0.25 for 'occasionally', 0.75 for 'usually' and 1 for 'always'. Where distances commuted were provided in km, these were converted to miles and truncated at 100miles overall, at 15 miles for cycle commutes, and 2.5 miles for walking
- for participants reporting multi-mode commutes, many assumptions were made.[7] Further amendments were also made when the total frequency of travel was >1.25 (when the frequency of each transport mode, assigned as in the previous bullet point, was summed), e.g. when someone reported always travelling by car and walking, the assumption was made that they drive most (99%) of the way
    - for driving or public transport and walking, the distance was estimated as 99% driven or by public transport and 1% walked
    - for public transport and cycling, the distance was estimated as 90% by public transport and 10% cycled
    - for cycling and walking, the distance was estimated as 95% cycled and 5% walked
    - for driving, public transport, and cycling, the distance was estimated as 47.5% each driven and by public transport and 5% cycled
    - for driving, public transport, and walking, the distance was estimated as 49.5% each driven and by public transport and 1% walked
    - for driving or public transport plus cycling and walking, the distance was estimated as 90% driven or by public transport, 9% cycled, and 1% walked
    - for use of all four modes, the distance was estimated as 45% each driven and by public transport, 9% cycled, and 1% walked
- assumed speeds by mode (using the existing syntax[6]) were: walk at 3mph, cycle at 10mph, car at 45mph and public transport at 30mph
- sports and leisure exercises were converted to frequencies per week, using

```
replace `var' = 0 if `var' <= 1 None
replace `var' = 1/4 if `var' == 2 Once in the last 4 weeks
replace `var' = 2.5/4 if `var' == 3 2 to 3 times in the last 4 weeks
replace `var' = 1 if `var' == 4 Once a week
replace `var' = 2.5 if `var' == 5 2-3 times a week
replace `var' = 4.5 if `var' == 6 4-5 times a week
```

replace `var' = 7 if `var' == 7 Every day

- over-estimation of duration of recreational activities were truncated
  - at 10 hours for golf, hunting, shooting, fishing, sailing, windsurfing, boating and do-it-yourself (DIY)
  - at 8 hours for cricket, mountain climbing, backpacking, walking, cycling (any type), dancing, heavy gardening or weeding
  - at 4 hours (for a single episode) for competitive or leisure swimming, competitive running, jogging, bowling, tennis, badminton, table tennis, horse-based activities, ice-skating, snooker, billiards, darts, playing musical instruments, and singing
  - at 3 hours for aerobics, other conditioning exercises, floor exercises, football, rugby, hockey, netball, basket ball, volley ball, rowing, combat sports, and mowing or watering the lawn
  - at 2 hours for exercise weights and squash
- where no frequency was reported but duration was recorded, the median frequency from those participating in that activity and reporting frequency was assigned
- where no duration was reported (but frequency was recorded) the median duration from those participating in the activity was assigned
- the mean time spent in each activity per day was then calculated by multiplying the episode duration by the frequency divided by 7, for each activity. All activities (including e.g. watching television, using a computer, i.e. not just physical activities) were then summed and truncated at 18 hours per day
- the number of hours of sleep per day was estimated as 6 hours if total duration of activities was >18 hours, as 8 hours if total duration was <16 hours, and as 24 hours minus total duration of activities if total duration was 16 to 18 hours.[8] Any remaining time (if total duration <16 hours) was called ‘unaccounted’ time
- adjusted durations for each domain (home, work, commuting, leisure) were generated considering maximum value of 18 for total activities, such that the total activity for these four domains plus the estimated sleep time always equalled 24 (minus ‘unaccounted time’)
- for each activity, MET scores (as per Ainsworth's PA Compendium[9]) were assigned, as in Table V.1

Table V.1. MET scores used when converting duration of each activity to energy expended

Domain	Activity	MET	Notes
Home	Watching TV	1.0	
	Using a computer	1.5	
	Using stairs	5.5	Halfway between 8 for going up and 3 for going down
Work	Sedentary occupation	1.5	Assigned 1 (median in Fenland dataset) where worktype is missing but time in work is reported
	Standing occupation	2.3	
	Manual work	3.5	
	Heavy manual work	5.5	
Commuting	Car	1.5	(driving 2.0, passenger 1.0)
	Public transport	1.0	
	Cycle	6.0	
	Walk	3.3	
Leisure	Watering lawn,	1.5	
	Snooker, billiards, darts	2.5	
	Singing, playing musical instruments	2.7	
	Exercise weights, bowling, hunting, shooting, fishing, sailing, windsurfing boating	3.0	
	Walking,	3.3	
	Cycling for pleasure, floor exercises, table tennis, horse-based activities	4.0	
	Weeding, pruning, DIY, dancing, golf	4.5	
	Aerobics (except high activity), cricket	5.0	
	Mowing lawn, conditioning exercises, netball, volley ball, basket ball	5.5	
	Leisure swimming heavy gardening, tennis, badminton,	6.0	

	Backpacking, mountain climbing, high activity aerobics, jogging, rowing, ice skating	7.0	
	Football, rugby, hockey	8.0	
	Competition swimming, cycle racing, Combat sports	10.0	
	Squash	12.0	
	Competition running	12.5	

From these, total time and total MET scores were obtained for each of the four domains and for all activities. The total time was subtracted from the total MET hours to yield the Activity METS score. A second activity MET hours variable was obtained by subtracting the estimated time spent sleeping from the total MET hours with 'unaccounted for' time. The difference between these two is the assignment of energy to 'unaccounted for' time; the first variable assigned 1.0 MET to all 'unaccounted for' time and the second assigned 1.3MET to 'unaccounted for' time if the person reports getting about actively. Estimating the total energy expended at each category of activity intensity was done using the categorisation in Table V.2.

Table V.2. Energy spent at different intensities[a]

Energy category	METs range	Activities included
Sedentary (SPA)	≤ 1.5 (excluding sleep)	Watching TV, Using a computer, Traveling by car or public transport, + Job if sedentary occupation
Light (LPA)	1.5001 – 2.99	Watering the garden, snooker, billiards, darts, singing, playing musical instruments, + Job if standing occupation
Moderate (MPA)	3.0 – 5.99	Using stairs, Walking, cycling, digging, mowing lawn, weeding, DIY, swimming, aerobics, weights, exercises, floor exercises, dancing, bowls, tennis, badminton, table tennis, golf, cricket, netball, hunting, shooting, fishing, horse-based activities, sailing, windsurfing, boating, leisure swimming + Job if manual work
Vigorous (VPA)	≥ 6.0	Swimming laps, competitive swimming, Backpacking, mountain climbing, High impact and Step aerobics, competitive

		running, cycling or swimming, racing or rough terrain cycling, jogging, squash, football, rugby, hockey, rowing, skating boxing, combat sports, + Job if heavy manual work

[a] Time spent in each of the activity intensity categories was estimated using this classification.

V.4.4 Variable names and labels

Total daily time spent in activity and/or sleeping (required as basis for deriving other variables)

TOTtime "Total reported duration (hours) of activity [hrs/d]"
TOTALtime "Total reported duration (hours) of activity + assumed sleep [hrs/d]"

Estimated daily total energy expended

TOTMETHRS "Total reported duration (hours) of activity times intensity (MET) [METhrs/d]"
TOTMETHRS_w_UNACCtime "Total reported plus unaccounted duration hours) times intensity (MET) [METhrs/d]"
ACTMETS "Total activity energy expenditure discounting resting [net METhrs/d]"

Daily energy expenditure by domain

HOME_METS "Home domain energy expenditure [METhrs/d]"
WORK_METS "Work domain energy expenditure [METhrs/d]"
LEIS_METS "Leisure domain energy expenditure [METhrs/d]"
COMMUTE_METS "Commute domain energy expenditure [METhrs/d]"

Daily energy expenditure by level of intensity

SED_INTENSITY "Sedentary behavior energy expenditure [METhrs/d]"
LIGHT_INTENSITY "Light intensity energy expenditure [METhrs/d]"
MODERATE_INTENSITY "Moderate intensity energy expenditure [METhrs/d]"
VIGOROUS_INTENSITY "Vigorous intensity energy expenditure [METhrs/d]"

Time spent per day at different intensity levels of activity

SEDtime "Time spent sedentary, excluding sleep [hrs/d]"
LIGHTtime "Time spent at light intensity activity [hrs/d]"
MODERATEtime "Time spent at moderate intensity activity [hrs/d]"
VIGOROUStime "Time spent at vigorous intensity activity [hrs/d]"

The last two variables were summed to provide the time per day spent in moderate of vigorous activity, the variable used in this report.

References

[1] Stephen A, Teucher B, Bluck L, Cole D, Fitt E, Mander A, Woodward R, Wright A, Bates B, Roberts C, Mackenzie H, Deverill C, Mindell J. National Diet and Nutrition Survey Rolling Programme, Comparison Study, Part 1. A comparison of results by dietary assessment method: repeat 24-hour recall and four-day estimated (unweighed) diet diary. Unpublished. 2008.

[2] In the Comparison Study, the correlation value against doubly labelled water (DLW) was 0.30. This is similar to the figures for a new short questionnaire for adults of r=0.38 for total activity, r=0.31 for strenuous intensity activities and r=0.29 for moderate intensity activities, compared with readings from an accelerometer (Orrell A *et al. Development and validation of a very brief questionnaire measure of physical activity in adults with coronary heart disease.* Eur J Cardiovasc Prev Rehabil. 2007;14:615-23); compared with an accelerometer, the OSWEQ and IPAQ each had lower estimates (mean error±95% PI) of MVPA MET.min·day^{-1} by 150.4±477.6 and 247.5±477.5, respectively (Taylor NJ *et al. Development and validation of the Online Self-reported Walking and Exercise Questionnaire.* J Phys Act Health. 2012. [Epub ahead of print]).

[3] Besson H, Brage S, Jakes RW, Ekelund E, Wareham NJ. Estimating physical activity energy expenditure, sedentary time and physical activity intensity by self-report in adults. *Am J Clin Nutr.* 2010;91:106-114.

[4] Root mean square error (RMSE) is the geometrical average distance from the estimate to the criterion, thus a measure of overall accuracy (lower is better).

[5] Kappa values of 0.21-0.40 are considered to represent fair agreement; values of 0.41-0.60 represent moderate agreement. See: Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;**33**:159-74.

[6] Scott R, den Hoed M, Brage S. *Syntax for deriving summary variables from the Recent Physical Activity Questionnaire, version 2.0.* Cambridge, UK: MRC Epidemiology Unit, 2011.

[7] It should be noted that mixed mode trips are not well quantified by RPA, but the fractions can be changed if need be. Ideally, this should be informed by external data or literature. The proportions stated in the text above are those used when processing the RAQ data for the Years 1-4 NDNS RP report.

[8] Questions were asked in the NDNS RP about time spent sleeping, so a more accurate estimate could be obtained by researchers using the archived data by using participant' individual responses instead of these estimates.

[9] Ainsworth BE, Haskell WL, Whitt MC, et al. Compendium of Physical Activities: an update of activity codes and MET intensities. *Med Sci Sport Exer.* 2000;32(suppl):S498-S516.