CLEANING VARIABLES FROM THE RPAQ

DATA ENTRY CLEANING

Section A

Getting about:

Problem	Recommended action
More than one column selected in a row	Enter the lowest category i.e. favouring the direction of "car" since the categories are sorted as following: Car< Public transport<Cycle<Walk

TV, DVD or Video viewing:

Problem	Recommended action
More than one column selected in a row	Enter the category in the middle or the highest category when no middle (This is the highest in this case instead of the lowest category because it is a sedentary activity)

Computer use at home:

Problem	Recommended action
More than one column selected in a row	Enter the category in the middle or the highest category when no middle (This is the highest in this case instead of the lowest category because it is a sedentary activity)

Flights of stairs:

Problem	Recommended action
More than one column selected in a row	Enter the category in the middle or the lowest category when no middle

Section B

How many hours per week did you work?

Problem	Recommended action
Range entered	Enter the average of the range

Type of work

Problem	Recommended action
More than one type of work selected	Enter the category in the middle or the lowest category when no middle

How many kilometers from home to work:

Problem	Recommended action
Range entered	Enter the average of the range

How many miles from home to work:

Problem	Recommended action
Range entered	Enter the average of the range

How many times a week did you travel from home to work:

Problem	Recommended action
Range entered	Enter the average of the range

How did you normally travel to work:

Problem	Recommended action
More than one column selected in a row	Enter the middle of the two columns selected or the lowest category when no middle (i.e. favouring the direction of 'never or rarely')

Section C

Problem	Recommended action
More than one column selected in a row	Enter the middle of the two columns selected or the lowest category when no middle
Range entered in 'times per episode' box	Enter the average of the range

Derivation of some preliminary variables:

The following section explains some of the cleaning that are <u>already</u> implemented in the data processing syntax.

Variable	Computation of the variable
MISSINGA	If nothing is ticked in section A then MISSINGA=1, otherwise MISSINGA=2
MISSINGJOB	If people do not tick anything in the 1st fold of section B then MISSINGJOB=1, otherwise MISSINGJOB=2
EMPLOYED	If people report having been in employment or they partly completed the 1st fold of section B then EMPLOYED=1, otherwise EMPLOYED=2
DURATIONJOB	The average of weekly hours worked per week for participants reporting hours of work per week.
MISSINGCOMMUTING	If people do not tick anything in the 2nd fold of section B then MISSINGCOMMUTING=1, otherwise MISSINGCOMMUTING=2
DISTWORKMILES	DISTWORKMILES=0 if people do not report any distance between home and work in miles or in kilometers DISTWORKMILES= distance in miles if people report only the distance between home and work in miles DISTWORKMILES= distance in kilometers converted to miles if people report only the distance between home and work in kilometers DISTWORKMILES = Average of distance in miles and distance in kilometers converted to miles if people report both the distance between home and work in miles and in kilometers
MISSINGC	If nothing is ticked in section C then MISSINGC=1, otherwise MISSINGC=2
MISSING	If MISSINGA=1 and MISSINGJOB=1 and MISSINGCOMMUT=1 and MISSINGC=1 then MISSING=1, otherwise MISSING=2

Section A

Getting about

Problem	Recommended action
Missing data	If MissingA=1 then kept as missing Otherwise car assigned by default

TV, DVD or Video Viewing

Problem	Recommended action
No tick in a row	If MissingA=1 then kept as missing Otherwise value assigned as in table 1

Computer use at home

Problem	Recommended action
No tick in a row	If MissingA=1 then kept as missing Otherwise value assigned as in table 1

TV, Video Viewing or Computer using **(Modified RPAQ)**

Problem	Recommended action
No tick in a row	If MissingA=1 then kept as missing Otherwise value assigned as in table 1

Table 1:

TV, DVD or Video Viewing	Assigned values
Weekdays before 6 pm	Less than 1 hour a day (or the most frequently reported in the population, if the study sample is big enough i.e. n >25)
Weekdays after 6 pm	1 to 2 hours a day(or the most frequently reported in the population, if the study sample is big enough i.e. n >25)
Weekend before 6 pm	Less than 1 hour a day (or the most frequently reported in the population, if the study sample is big enough i.e. n >25)
Weekend after 6 pm	2 to 3 hours a day(or the most frequently reported in the population, if the study sample is big enough i.e. n >25)
Computer use at home	
Weekdays before 6 pm	Less than 1 hour a day (or the most frequently reported in the population, if the study sample is big enough i.e. n >25)
Weekdays after 6 pm	Less than 1 hour a day (or the most frequently reported in the population, if the study sample is big enough i.e. n >25)
Weekend before 6 pm	Less than 1 hour a day (or the most frequently reported in the population, if the study sample is big enough i.e. n >25)
Weekend after 6 pm	Less than 1 hour a day (or the most frequently reported in the population, if the study sample is big enough i.e. n >25)
TV, Video Viewing or Computer using (Modified RPAQ)	
Weekdays before 6 pm	1 to 2 hours a day(or the most frequently reported in the population, if the study sample is big enough i.e. n >25)
Weekdays after 6 pm	1 to 2 hours a day(or the most frequently reported in the population, if the study sample is big enough i.e. n >25)
Weekend before 6 pm	1 to 2 hours a day(or the most frequently reported in the population, if the study sample is big enough i.e. n >25)
Weekend after 6 pm	2 to 3 hours a day(or the most frequently reported in the population, if the study sample is big enough i.e. n >25)

Stair Climbing at Home

Problem	Recommended action
No tick in a row	If MissingA=1 then kept as missing Otherwise Considered as zero

Section B

Occupation Details

Problem	Recommended action
DURATIONJOB >70	Truncated to 70
DURATIONJOB is missing	If EMPLOYED =1 then DURATIONJOB assigned is 24 (or the median of reported hours per week calculated from people reporting hours per week) If EMPLOYED =2 then DURATIONJOB considered as zero

Type of work

Problem	Recommended action
No tick	"Sedentary occupation" assigned by default

Travel to and from work
Active and passive commuting have been distinguished since these 2 activities can be combined (e. g. going to the train station by bike and then taking the train. In this case, the distance home – work represents neither the distance travelled by bike nor this travelled by train)

Problem	Recommended action
People go to work by foot only (and never by bike) and they report a distance between home and work > 5 ml or they do not report a distance between home and work	The distance assigned for active commuting is 1 (or assign the median in the population, if enough people in this category i.e. n >25)
People go to work by foot and by bike and they report a distance between home and work > 10 ml or they do not report a distance between home and work	The distance assigned for active commuting is 2.5 (or assign the median in the population, if enough people in this category i.e. n >25)
People go to work by bike (and never by foot) and they report a distance between home and work > 15 ml or they do not report a distance between home and work	The distance assigned for active commuting is 3 (or assign the median in the population, if enough people in this category i.e. n >25)
People go to work by foot only (and never by bike) and they report a number of journeys between home and work > 14 or they do not report the number of journeys between home and work	The number of journeys assigned for active commuting is 5 (or assign the most frequently reported in the population, if enough people in this category i.e. n >25)
People go to work by foot and by bike and they report a number of journeys between home and work > 14 or they do not report the number of journeys between home and work	The number of journeys assigned for active commuting is 5 (or assign the most frequently reported in the population, if enough people in this category i.e. n >25)
People go to work by bike (and never by foot) and they report a the number of journeys between home and work > 14 or they do not report the number of journeys between home and work	The number of journeys assigned for active commuting is 5 (or assign the most frequently reported in the population, if enough people in this category i.e. n >25)
People go to work by car and they report a distance between home and work > 150 ml or they do not report a distance between home and work	The distance assigned for passive commuting is 5 (or assign the median in the population, if enough people in this category i.e. n >25)
People go to work by public transport and they report a distance between home and work > 150 ml or they do not report a distance between home and work	The distance assigned for passive commuting is 5 (or assign the median in the population, if enough people in this category i.e. n >25)
People go to work by car and they report a number of journeys	The number of journeys assigned for passive commuting is 5 (or assign the most frequently reported in the population,

between home and work > 14 or they do not report the number of journeys between home and work	if enough people in this category i.e. n >25)
People go to work by public transport and they report a number of journeys between home and work > 14 or they do not report the number of journeys between home and work	The number of journeys assigned for passive commuting is 5 (or assign the most frequently reported in the population, if enough people in this category i.e. n >25)
'Always/usually' walking and 'always/usually' cycling both selected	The coefficients of frequency are normalized to 1 by dividing the reported frequency coefficient by the summation of the frequency coefficients reported by the subject

Section C

Problem	Recommended action
No tick in a row	If Missingc=1 then kept as missing Otherwise Considered as zero
Frequency given but duration per episode missing or excessive (see Annex 1 for allowable values)	Values assigned as in the table 5 'Duration to be assigned' (or assign the medians in the population, if enough people in the corresponding categories i.e. n >25)
Duration per episode given but frequency missing	Values assigned as in the table 5 'Frequency to be assigned' (or assign the most frequently reported in the population, if enough people in the corresponding categories i.e. n >25)
Both frequency and duration per episode are missing	If MissingC=1 then kept as missing Otherwise Considered as zero