

Data Cleaning for Height and Weight – Younger and Mid-aged

Age Cohorts	Younger and Mid-age
Surveys	All
Derived Variable	Height
Definition	Best estimate of actual height given all survey responses
Source Items (Index Numbers)	Height reported at survey (WTSH-001)
Statistical form	Continuous variable
Index Number	(WTSH-074)
Age Cohorts	Younger, Mid-age and Older
Surveys	All
Derived Variable	BMI
Definition	Body Mass Index – a measure of adiposity which is independent of height; also known as the Quetelet Index
Source Items (Index Numbers)	Height: Estimated (WTSH-074) or reported (WTSH-001) Self-reported weight (WTSH-002)
Statistical form	Continuous variable
Index Number	(WTSH-041)
Prepared by	Jean Ball, Jess Ford, Anne Russell & Lauren Williams
Endorsed	24 April 2002 (updated 19 July 2006)

The earliest formal guidelines for cleaning anthropometric data were developed by Lauren Williams¹ as part of her doctoral research and are recorded below.

Height

At each survey, height is reported in either in centimetres, or in feet and inches and converted to centimetres; values were rounded to the nearest integer.

Between April 2002 and August 2006 limits were set on plausible values for height in each of the project cohorts (height less than 120cm: all cohorts; height greater than 200cm: Younger; height greater than 190cm: Mid-age and Older) and values outside these limits were set to missing. From August 2006, these values were used to trigger a data audit (see later).

Weight

At each survey, weight is reported either as stones and pounds, or as kilograms; stones and pounds are converted to kilograms. Weight in kilograms is rounded to the nearest 0.1 of a kilogram.

In April 2002 the following limits were set for plausible weights and the derived variable body mass index (BMI) in each of the project cohorts; values outside these limits were set to missing. The upper limit for weight among younger and mid-age women was set at the maximum reading (139.9kg) for the digital scales used in Australia's National Nutrition Survey (NNS) in 1995.²

	Younger	Mid-age	Older
Lower limit for weight (kg)	30	30	30
Upper limit for weight (kg)	140	140	120
<i>Lower limit for body mass index</i>	<i>14</i>	<i>14</i>	<i>14</i>
Upper limit for body mass index	55	55	50

In July 2006 this decision was reviewed in light of analytic issues arising from it. Specifically, records with missing data are excluded from a statistical analysis. This is undesirable and, in cases such as weight (and consequently BMI) where the level of missing data is high, this has become a major issue. The exclusion of extreme values as a data cleaning strategy exacerbates this problem. Interestingly the technique of data imputation, which is being applied to the problem of missing data, would use responses to other survey items to estimate weight (or BMI), even if weight had not been missing from surveys but was believed to be implausible. Application of imputation in this instance would introduce incorrect data values into the statistical analysis.

An examination of the patterns of reported weight and BMI across Surveys 1 to 3 for all Mid-age women with an extreme BMI value (<14 or >55) at 1 or more surveys showed that the pattern of weight was mostly consistent, suggesting the reports were true though extreme. A sample of readily accessible survey forms was checked to exclude an error in data-entry. Values were mostly correctly entered and, in one case, the correction a data-entry error removed the outlying value.

In addition, data published from longitudinal studies of free-living populations of women in which body weight was *measured* was reviewed to describe the true range of weight changes which women experience. An attempt was made to use these data to develop values for weight change between ALSWH surveys consistent with that observed when weight is measured rather than reported (Appendix 2). A wide range of values of real changes were reported and so the option of assigning weight (and BMI) to missing when an extreme change is reported was rejected.

Examples of extreme changes reported by members of the Younger cohort (Table 1) suggested data entry errors, e.g. the Survey 2 value for Case 1 may have been incorrectly entered as 30kg rather than 50kg.

Table 1 Some extreme weight changes in the Younger cohort

case	Reported weight at:			Weight Change for:			
	Survey 1	Survey 2	Survey 3	Surveys 1 to 2 (4 years)		Surveys 2 to 3 (3 years)	
				Actual	Percent	Actual	Percent
1	58	30	55	-28kg	48%	+25	183%
2	50	50	115	0kg	0%	+65kg	230%
3	64	63	134	-1kg	98%	+71kg	213%
4	Missing	33.6	100			+66kg	298%

So, in 2006 it was decided that:

- Extreme values should no longer be deleted
- Extreme values for weight, BMI, height and change in weight should be audited and unless a data-entry error is identified, the value should stand. Extremes are defined as:
 - Weight < 30kg (all cohorts)
 - Weight > 140kg (Younger and Mid-age) or > 120kg (Older)
 - BMI < 14 (all cohorts)
 - BMI > 55 (Younger and Mid-age) or > 50 (Older)
 - Height less than 120cm (all cohorts)
 - Height > 200cm (Younger) or > 190cm (Mid-age and Older)
 - Lose more than 11 kg in 3 years (3.7 kg in 1 year)
 - Gain more than 20 kg in 3 years (6.6 kg in 1 year)

Weight and pregnancy

For the first 3 surveys of the Younger cohort and the first survey of the Mid-age cohort, the value of weight is set to missing for women who reported being pregnant at the time of the survey. Current pregnancy was not asked of the Mid-age women after Survey 1. Starting with Survey 4, pregnant Younger women were asked to record their pre-pregnancy weight.

Historical Criteria for screening and cleaning of the anthropometric data of Mid-age cohort at Surveys 1 and 2

Lauren Williams¹ developed these guidelines for cleaning anthropometric data for Mid-age women. Acceptable ranges for height, weight and BMI were based on the distributions of these parameters among women aged 45-64 years in the National Nutrition Survey (NNS).³ In the NNS heights and weights were measured, unlike the self-report data from ALSWH. The plausibility of weight changes between surveys was assessed against reported weight changes in The Melbourne Mid-life study.⁴

Height (cm)

Acceptable range:² ≥ 150 and < 180

Rationale:²

- No heights were 180cm or above.
- Only 3.9% of heights were in the range of 140-149 cm.

Data Checking:

Survey 1

If Survey 1 height is outside the reference range or missing:

- o If Survey 2 height is inside the reference range, assign Survey 2 height to Survey 1.
- o If Survey 2 height is outside the reference range, assign both heights to missing.

Survey 2

Use the reverse of Survey 1 procedure.

Weight (kg)

Acceptable range:² > 35 and < 130

Rationale:²

- No weights were below 40 kg; the lower limit of the acceptable range was decreased to 35 kg to allow for underestimation in self-report weight.
- Only 0.2% of weights were 130 kg or more; 130 kg is the maximum value on bathroom scales.

Data Checking:

Surveys 1 and 2

If the difference in weight between Surveys 1 and 2 exceeds 30 kg, assign weight at both surveys to missing.

BMI¹

Acceptable range:² ≥ 15 and < 50

Rationale:²

- The 5th percentile for BMI was 20.5.
- The 95th percentile for BMI was 37.9.

Data Checking:

Weight and height were revised as shown above prior to calculation of BMI.

Surveys 1 and 2

If both weight and height are in the acceptable ranges, BMI is calculated.

If either weight or height is outside the acceptable ranges, BMI is assigned as missing.

Weight change from Survey 1 to Survey 2 (two years)

Acceptable change: ≤ 30 kg

Rationale: The Melbourne Mid-life study¹⁰ recorded weight loss over a five year period. The maximum observed for weight loss was 18.5 kg and for weight gain was 26.0 kg. The acceptable range of weight change was widened for the ALSWH.

Checking procedure:

Calculated weight changes exceeding 30 kg were set to missing.

References

1. Williams L. Factors affecting weight change in mid-aged women. PhD Thesis, University of Newcastle, 2003
2. McLennan W and Podger A. National Nutrition Survey Users' Guide 1995. (Page 23) *Australian Bureau of Statistics*, 1997, Catalogue Number 4801.0
3. Physical Measurements. 1998, AGPS: Canberra. p. 110-112, 115
4. Guthrie, JR, Dennerstein L, Dudley EC. Weight gain and the menopause: a 5-year prospective study. *Climacteric*, 1999. 2: p. 205-211