

Weights – Older Cohort

General weight calculation (see details in younger and mid aged section of data dictionary supplement)

1. Weight reported in kgs as required
2. Weight reported in stones and pounds - convert to kg
3. Weight reported in pounds only - convert to kg
4. All weight variables are missing - weight is missing

All weights >120kg or <30kg were considered to be extreme. As for heights, in Survey 1 to Survey 3 extreme values were deleted, but from 2006 extreme values will be accepted unless a data error has been detected. This applies to past and future surveys, which means that raw data for Survey 1-3 was used in order to find out whether a previously deleted extreme value will be recovered based on similar values in the other surveys.

The problem with weights was to find an acceptable range of losing or gaining weight between two surveys. Previously a weight loss of 11kg and a weight gain of 20kg in 3 years was allowed. Although the weight in older people is probably more stable than in the other cohorts, allowing 11kg loss in three years would exclude extreme values which might provide important information, eg people with cancer can lose this in 6months or less. Additionally, for heavy people gaining another 10kgs is not as unusual as for someone with normal weight. Therefore it has been decided to use percentage weight loss/gain rather than a set value.

A weight change up to +/- 30% between consecutive surveys was considered to be acceptable, which increased to -32% loss between non-consecutive surveys. This decision is based on values from Table 8 and the UK cancer help website, which suggests losing 10% in 6 months needs investigating, particularly if the weight loss continues

(<http://www.cancerhelp.org.uk/help/default.asp?page=10428>, 21/3/2007, 21:45hrs.)

Correcting extreme values for weight in Survey 4

Our data showed 65 women with a weight >120kg and 47 women with a weight <30kg at Survey 4 (N=112).

In order to find out whether the value for weight in Survey 4 is acceptable or not, the mean of the previous surveys was calculated, and this mean value was used for calculating the percentage gain or loss to Survey 4. For this comparison a change of 35% was considered to be within limits as it was a comparison between the mean of heights from Survey 1-3 to height in Survey 4.

```
weightcheck='empty';
meanS1S2S3=MEAN(of wt1kg wt2kg wt3kg);
perdiff4321=((wt4kg/meanS1S2S3)*100)-100;

IF perdiff4321>35 or perdiff4321<-35 THEN weightcheck='wr';ELSE
weightcheck='ok';
```

If the weight in Survey 4 was within the limits, it was considered to be a true extreme value, if not the following checks were undertaken to find out whether stones were entered into pounds or pounds in kgs etc.

Table 6 Possible mistakes and their interpretation and proportion of weights recovered.

Rule	Kgs	Stones	Pounds	Action	weights recovered
1	X (>120)	.	.	a)do kgs*0.4536 b)do kgs-100 c)do kgs/10 The new weight is the one that results in the smallest difference between weights, or is set to missing if the weight check test indicates it is not within the limits.	26 of 28
2	X (<30)	.	.	a)do kgs*14*0.4536 b)do kgs*10 The new weight is the one that results in the smallest difference between weights, or is set to missing if the weight check test indicates it is not within the limits.	16 of 18
3	.	X	X	Do newkg= (((o4q24stones/10)*14)+o4q24pounds)*0.4536 This value is compared with previous surveys, and is set to missing if the weight check test indicates it is not within the limits.	4 of 4
4	.	X	.	AND the calculated value for kgs is >120 then do a)kgs=stones, they meant kgs. b)eg 129stones=>12stones9pounds Again, both values are compared with previous surveys and The new weight is the one that results in the smallest difference between weights, or is set to missing if the weight check test indicates it is not within the limits.	29 of 30
5	.	X	.	AND the calculated value for kgs is <30, then set weight for o4 to missing.	0 of 3
6	.	.	X	AND the calculated value for kgs is <30 then do If 0<x<18 then they meant stones, if x>=18 then they meant kgs. The new value is compared with previous surveys, and kept unless it was beyond the limits.	21 of 23
7	X(>120)	X	X	Stones and pounds were taken instead to derive kgs, and as above compared with previous surveys.	2 of 2

101 out of 112 extreme values were recovered, the other 11 were set to missing.

Comparing weight change and filling in missing values/substitute wrong values in all surveys

Next, the change in weight between surveys was calculated. Between two consecutive surveys a percent weight change of 30% was accepted and if there were one or more surveys between the two values, the possible weight loss increased to 32%. The following table shows all possible combinations of consecutive surveys and the action that was undertaken. S21 stands for the percentage weight change between S1 and S2, S32 stands for the percentage weight change between S2 and S3, and S43 stands for the percentage weight change between S3 and S4. An empty field indicates that the percentage weight change could not be calculated (one survey missing), + or – 30 stands for a weight change greater or less than 30% respectively, ok means the weight change was within limits.

Table 7 Combination of percentage weight change between consecutive surveys and their interpretation.

Between surveys			N	Action
S21	S32	S43		
			2951	Keep as it is except*
		+30	3	Set Survey 3 and Survey 4 missing
		-30	3	Set Survey 3 and Survey 4 missing
		ok	319	Keep as it is
	+30		2	Set Survey 2 and Survey 3 missing
		ok	3	Set Survey 2 missing
	-30		1	Set Survey 2 and Survey 3 missing
		ok	50	Keep as it is
		-30	1	Set Survey 4 missing
		ok	116	Keep as it is
+30			40	Set Survey 1 and Survey 2 missing
	-30		11	Set Survey 2 missing
		ok	39	Set Survey 2 missing
		ok	5	Set Survey 1 missing
		ok	11	Set Survey 1 missing
-30			41	Set Survey 1 and Survey 2 missing
		ok	5	Set Survey 1 and Survey 2 missing
	+30		16	Set Survey 2 missing
		-30	2	All missing
		ok	51	Set Survey 2 missing
		ok	3	Set Survey 1 missing
		ok	5	Set Survey 1 missing
ok			2280	Keep as it is
	+30		9	Set Survey 3 missing
		-30	8	Set Survey 3 missing
		ok	19	Keep as it is
	-30		12	Set Survey 3 missing
		+30	13	Set Survey 3 missing
		ok	3	Keep as it is
		ok	1634	Keep as it is
		+30	47	Set Survey 4 missing
		-30	25	Set Survey 4 missing
		ok	5212	Keep as it is

*The following cases had to be distinguished:

1. S1 NE(not equal) . AND S2=. AND S3=. AND S4 NE .
2. S1 NE . AND S2=. AND S3 NE . AND S4=.
3. S1=. AND S2 NE . AND S3=. AND S4 NE .

If the percentage weight change in these cases was less than a 32% loss or more than a 30% gain, the values were set to missing, otherwise they were kept as they were.

Table 8 Distribution of percentage weight change between surveys BEFORE data cleaning.

Variable	N	Min	1st Pctl	5th Pctl	Mean	95th Pctl	99th Pctl	Max	Median
perdiff21	9478	-98.43	-35.90	-13.02	2.297	10.24	33.33	1110	0
perdiff32	7298	-90.88	-32.24	-13.85	5.476	11.59	57.63	6456	0
perdiff43	5885	-55.00	-27.35	-14.24	-1.188	10.64	32.09	139.0	-0.065
perdiff41	6398	-58.47	-31.43	-18.75	-2.403	12.36	25.00	105.3	-1.786
perdiff31	7635	-58.90	-27.03	-15.71	-1.328	11.43	22.20	109.8	0
perdiff42	6134	-91.60	-37.44	-17.76	3.118	13.29	52.76	6144	-1.213

Table 9 Distribution of percentage weight change between surveys AFTER data cleaning.

Variable	N	Min	1st Pctl	5th Pctl	Mean	95th Pctl	99th Pctl	Max	Median
perdiff21	9262	-29.63	-20.00	-11.81	-0.903	9.013	17.24	29.64	0
perdiff32	7131	-32.23	-20.67	-12.59	-0.658	10.24	20.31	51.42	0
perdiff43	5783	-29.98	-23.08	-13.41	-1.496	9.872	19.34	29.89	-0.167
perdiff41	6292	-45.80	-27.82	-18.17	-2.455	11.83	20.37	58.75	-1.822
perdiff31	7557	-39.69	-24.21	-15.38	-1.374	11.01	18.75	58.75	0
perdiff42	5957	-45.18	-25.64	-16.67	-1.829	11.58	21.43	69.69	-1.271

This clearly shows that the data cleaning was effective.