

## Data management of the DQES

Managing data from the DQES has evolved from the time of the pilot for Mid-age Survey 3 through to the main 4<sup>th</sup> Survey of the Younger cohort. Changes have occurred in the procedures used at the CCV and the ALSWH has adapted its data management to accommodate this.

The CCV are unable to assign nutrient scores if:

- q1 or q2 is missing,
- the type for certain foods is missing. E.g. milk, bread and cheese,
- there are irreconcilable inconsistencies in certain responses. E.g. responses specify that the person does not eat cheese but also specifies a type of cheese that is usually eaten.

For surveys of the type normally conducted by the CCV, these difficulties are resolved by contacting participants; however the sizes of the ALSWH cohorts prohibit this. So, in order to minimise the number of records with missing nutrient data, the ALSWH Data Manager, with the endorsement of Amanda Patterson (Dietitian), has developed rules to reconcile some of these inconsistencies.

The full range of data cleaning applied to the DQES is described below; the extent to which data were re-coded is indicated for Mid-age Survey 3. A number of these procedures have subsequently been adopted by the CCV.

### 1. Re-coding according to CCV Guidelines

- assign the value of 1 (never or the null value) for all missing items other than q3a-e (type of milk), q5 a-f (type of bread), q7 a-f (type of fat spread) and q10 a-g (type of cheese)
- re-code all items concerning alcohol intake to 'never' for those who don't drink; note that the last 2 items concerning alcohol intake (q15 and q16) are not re-coded.

### 2. Standard ALSWH re-coding procedures applied to Type and Usual Amount of certain foods

#### Bread

Inconsistency:	'I don't eat bread' is true and the Type of bread usually eaten is given
Re-coding rule:	Re-code 'I don't eat bread' to be false
SAS commands to re-code:	if q5a=1 and (q5b=1 or q5c=1 or q5d=1 or q5e=1 or q5f=1) then q5a=0
Number re-coded:	5

## Cheese

Inconsistency:	'I don't use cheese' is true and the Type of cheese usually eaten is given
Re-coding rule:	Re-code 'I don't use cheese' to be false
SAS commands to re-code:	if q10a=1 and (q10b=1 or q10c=1 or q10d=1 or q10e=1 or q10f=1 or q10g=1) then q10a=0
Number re-coded:	8
Inconsistency:	Frequency of eating cheese is 'Never' and all Types of cheese are given as <i>not</i> usually eaten
Re-coding rule:	Re-code 'I don't use cheese' to be true
SAS commands to re-code:	if q15Ba=1 and q10a=q10b=q10c=q10d=q10e=q10f=q10g=0 then q10a=1
Number re-coded:	3
Inconsistency:	Frequency of eating cheese is missing, 'I don't eat cheese' is true and all Types of cheese are given as <i>not</i> usually eaten
Re-coding rule:	Re-code Frequency of cheese consumption to 'Never'
SAS commands to re-code:	if q15Ba=. and q10a=1 and q10b=q10c=q10d=q10e=q10f=q10g=0 then q15Ba=1
Number re-coded:	15
Inconsistency:	Frequency of eating cheese is missing, 'I don't eat cheese' is false and all types of cheese are given as <i>not</i> usually eaten
Re-coding rule:	Re-code Frequency of cheese consumption to 'Never' and 'I don't eat cheese' to be true
SAS commands to re-code:	if q15Ba=. and q10a=q10b=q10c=q10d=q10e=q10f=q10g=0 then q10a=1 and q15Ba=1

## Fat spread

Inconsistency:	'I don't usually use any fat spread' is true and the Type of spread usually eaten is given
Re-coding rule:	Re-code 'I don't usually use any fat spread' to be false
SAS commands to re-code:	if q7a=1 and (q7b=1 or q7c=1 or q7d=1 or q7e=1 or q7f=1) then q7a=0
Number re-coded:	25

## **Milk**

Inconsistency: Type of milk = 'None' is true and the Type of milk usually used is given

Re-coding rule: Re-code Type of milk = 'None' to be false

SAS commands to re-code: if q3a=1 and (q3b=1 or q3c=1 or q3d=1 or q3e=1) then q3a=0

Number re-coded: 9

Inconsistency: Type of milk is not specified and the Usual Amount = 'None' is true

Re-coding rule: Re-code Type of milk = 'None' to be true

SAS commands to re-code: If q4=1 and q3a=q3b=q3c=q3d=q3e=0 then q3a=1

Number re-coded: 4

Inconsistency: Type of milk = 'None' is true and the Usual Amount consumed is missing

Re-coding rule: Re-code Usual Amount = 'None' to be true

SAS commands to re-code: if q3a=1 and q4=. Then q4=1

Number re-coded: 4

### **3. Re-coding to overcome inconsistencies which prevent calculation of nutrients**

#### a) The frequency of consumption is reported but the usual type of food eaten is not

'Standard' food types' have been adopted in these instances. No data on the 'normal' or 'average' types of bread, milk or cheese consumed in Australia were available from the National Nutrition Survey. On a marketing website, Rosemary Stanton stated that the most popular bread was white bread, the most popular milk was full cream milk and that the most popular cheese was cheddar. The values have used these as the 'standard' when the type of product was missing.

## **Bread**

Usual type: White bread (q5c=1)

SAS commands to re-code: if q5b=q5c=q5d=q5e=q5f=0 and 2<=m3q55<=7 then do ;  
q5c=1 ;  
if q5a=1 then q5a=0 ; end ;

Number re-coded: 29

### **Cheese**

Usual type: Firm cheese, e.g. cheddar, edam (q10c=1)  
SAS commands to re-code: if 2<= q15bBa<=10 and q10a=1 then do ;  
q10a=0 ;  
q10c=1 ; end ;  
if 2<= q15bBa<=10 and  
q10a=q10b=q10c=q10d=q10e=q10f=q10g=0 then q10c=1 ;  
Number re-coded: 215

### **Milk**

Usual type: Full cream milk (q3b=1)  
SAS commands to re-code: if 2<=q4<=5 and q3a=1 and q3b=q3c=q3d=q3e=0 and then  
do ;  
q3a=0 ;  
q3b=1 ; end ;  
if 2<=q4<=5 and q3a=q3b=q3c=q3d=q3e=0 then q3b=1 ;  
Number re-coded: 58

### b) Type is given but the usual amount consumed is not

#### **Milk**

Usual Amount: Less than 250ml per day (q4=2)  
SAS commands to re-code: if (q3b=1 or q3c=1 or q3d=1 or q3e=1) and q4=1 then q4=2 ;  
Number re-coded: 122

### c) Usual amount is given as 'I don't eat fruit' and some fruits were shown as being consumed in the food frequency grid.

SAS code to identify inconsistent records: if q1=1 and  
(2<=q15Ca<=10 or 2<=q15Cb<=10 or 2<=q15Cc<=10 or  
2<=q15Cd<=10 or 2<=q15Ce<=10 or 2<=q15Cf<=10 or  
2<=q15Cg<=10 or 2<=q15Ch<=10 or 2<=q15Ci<=10 or  
2<=q15Cj<=10 or 2<=q15Ck<=10 or 2<=q15Cl<=10 or  
2<=q15Cm<=10) then inconsfr=1 ;

Responses to the frequency grid (q15Ca to q15Cm) were used to estimate the total number of pieces of fruit consumed in the last 12 months. First, a weighted sum of the response codes for all fruits listed on the grid was calculated, where the weighting is an estimate of that fruit's contribution to the total. The weights are shown below.

Response Code	Frequency of consumption	Weight
1	Never	0
2	Less than once month	6
3	1-3 times per month	24
4	1 time per week	52
5	2 times per week	104
6	3-4 times per week	182
7	5-6 times per week	286
8	1 time per day	365
9	2 times per day	730
10	3 or more times per day	1 095

The weighted sum is then used to assign appropriate values to item 1 'How many pieces of fresh fruit do you usually eat per day?', as shown.

#### Estimate of the total

number of pieces of fruit eaten	Response Code	Response category
0	1	I don't eat fruit
<= 312	2	Less than 1 piece of fruit per day
<= 547	3	1 piece of fruit per day
<= 912	4	2 pieces of fruit per day
<= 1 277	5	3 pieces of fruit per day
> 1 277	6	4 or more pieces of fruit per day

SAS commands to re-code:

```
array Qq15amt(13) q15Ca q15Cb q15Cc q15Cd q15Ce q15Cf
q15Cg q15Ch q15Ci q15Cj q15Ck q15Cl q15Cm;
addq15amt = 0 ;
do i=1 to 13 ;
  if qq15amt(i)=1 then addq15amt = addq15amt + 0 ;
  if qq15amt(i)=2 then addq15amt = addq15amt + 6 ;
  if qq15amt(i)=3 then addq15amt = addq15amt + 24 ;
  if qq15amt(i)=4 then addq15amt = addq15amt + 52 ;
  if qq15amt(i)=5 then addq15amt = addq15amt + 104 ;
  if qq15amt(i)=6 then addq15amt = addq15amt + 182 ;
  if qq15amt(i)=7 then addq15amt = addq15amt + 286 ;
  if qq15amt(i)=8 then addq15amt = addq15amt + 365 ;
  if qq15amt(i)=9 then addq15amt = addq15amt + 730 ;
  if qq15amt(i)=10 then addq15amt = addq15amt + 1095 ;
end ;

if inconsfr=1 then do ;
  if addq15amt <=312 then q1=2 ;
  else if addq15amt <=547 then do q1=3 ;
  else if addq15amt <=912 q1=4 ;
  else if addq15amt <=1277 q1=5 ;
  else if 1277<addq15amt<5000 then q1=6 ;
```

Number re-coded: 219

#### 4. Inconsistencies which prevent calculation of nutrients which were not re-coded

a) Number of pieces of fruit eaten per day selected and no types of fruit selected  
 SAS code to identify inconsistent records: if 2<=q1<=6 and q15Ca=1 and q15Cb=1 and q15Cc=1 and q15Cd=1 and q15Ce=1 and q15Cf=1 and q15Cg=1 and q15Ch=1 and q15Ci=1 and q15Cj=1 and q15Ck=1 and q15Cl=1 and q15Cm=1.

Observations (n): 97

b) Vegetables per day selected and no types of vegetable selected

SAS code to identify if 2<=q2<=7 and

inconsistent records: q15Da=q15Db=q15Dc=q15Dd=q15De=q15Df=q15Dg=q15Dh  
=q15Di=q15Dj=q15Dk=q15Dl=q15Dm=q15Dn=q15Do=q15Dp  
=q15Dq=q15Dr=q15Ds=q15Dt=q15Du=q15Dv=q15Dw=q15D  
x=q15Dy=1

Observations (n): 34

c) Cheese type selected and cheese never eaten

SAS code to identify if (q10b=1 or q10c=1 or q10d=1 or q10e=1 or q10f=1 or

inconsistent records: q10g=1) and q15Ba=1

Observations (n): 32