Combining data from Years 1-4, Years 5&6, Years 7&8 and Years 9-11

The NDNS datasets for Years 1-4, Years 5&6, Years 7&8 and Years 9-11 can be combined for analysis of Years 1-11 but, to produce valid results, the four sets of weights should be re-scaled. This will ensure that the four sets of data are in their correct proportions i.e. 4:2:2:3. A different calculation is required for each weight (individual, nurse, blood etc).

Re-scaling is necessary because there were different numbers of participants in each dataset/year. The number of participants per dataset/year is as follows:

Year	**Number of participants**
1-4	6,828
5&6	2,546
7&8	2,723
9-11	3,558
Total	*15,655*

If the weights were combined into one variable without any further adjustment, Years 1-4 and Years 7&8 would have more weight *per year* than other years, while Years 9-11 would have less weight *per year*.

To re-scale the weights correctly, it is necessary to perform the following calculations:

i. Divide each weight variable by its sum (i.e. the sum of the weights);
ii. Multiply each by the combined sum of the four weights (15,655);
iii. Multiply the Years 1-4 weight by 4/11, Years 5&6 weight by 2/11, Years 7&8 weight by 2/11 and Years 9-11 weight by 3/11.

The resulting weights can then be combined into one variable.

Example: individual weights

For example, to create new individual weights for analysis of the combined dataset, the steps would be as follows:

i. Divide Years 1-4 weight by 6,828, the Years 5&6 weight by 2,546, the Years 7&8 weight by 2,723 and the Years 9-11 weight by 3,558;
ii. Multiply each weight by 15,655;
iii. Multiply the Years 1-4 weight by 4/11, the Years 5&6 weight by 2/11, the Years 7&8 weight by 2/11 and the Years 9-11 weight by 3/11;
iv. Combine the resulting weights into one variable.

To do all of this in in SPSS you could use the following syntax:

```
weight off.
compute wti_UKY1234r = wti_UKY1234 * 15655 / 6828 * (4/11).
compute wti_Y56r     = wti_Y56     * 15655 / 2546 * (2/11).
compute wti_Y78r     = wti_Y78     * 15655 / 2723 * (2/11).
compute wti_Y91011r      = wti_Y91011      * 15655 / 3558 * (3/11).
compute wti_UKY1to11 = sum (wti_UKY1234r, wti_Y56r, wti_Y78r,
wti_Y91011r).
```

Running these commands will result in a new set of weights (`wti_UKY1to22`) which can be used for analysing Years 1-9. The new weight should have a mean of 1 and be non-missing for all 13,350 cases in the combined dataset. You can check this by running descriptives on the weights:

```
desc wti_UKY1to11.
```

Notes:
1. The intermediate weights (`wti_UKY1234r, wti_Y56r`, `wti_Y78r, wti_Y91011r`) can be discarded/deleted.
2. In this example the sum of the weights is equal to the total sample size in each case. This will not hold for subgroups (see below).

Combining other weights

Analogous calculations can be performed for:

Nurse weights:	`wtn_UKY1234, wtn_UKY56, wtn_Y78, wtn_Y91011`
Blood weights:	`wtb_UKY1234, wtb_UKY56, wtb_Y78, wtb_Y91011`
RPAQ weights:	`wtr_UKY1234, wtr_UKY56, wtr_Y78, wtr_Y91011`
Spot- urine weights:	`wtsu_Y6, wtsu_Y78, wtsu_Y91011` {* see below}

* Note that spot-hour urine samples were only collected in Years 6-11 so when combining these weights, the correct proportions to use are 1/6; 2/6; 3/6.

Weights for combined sub-group analysis

The above explanation assumes that analysis will be performed for all cases i.e. all adults and children. If analysis of subgroups is required, analogous calculations should be performed on the combined dataset filtered to include only the subgroup of interest. This will produce bespoke weights for analysis of that particular subgroup (adults only for example).

One additional step is required but otherwise the procedure is the same:

i. Divide each weight variable by its sum (i.e. the sum of the weights);
ii. Multiply each by the total (combined) sum of the four weights;
iii. Multiply the Years 1-4 weight by 4/11, the Years 5&6 weight by 2/11, the Year 7&8 weight by 2/11 and the Years 9-11 weight by 3/11;
iv. Combine the resulting weights into one variable;
v. Re-scale this weight to have a mean of 1.

The additional step (v) ensures that the resulting weights have a mean of 1.

Example: individual weights (adults only)

To create new individual weights for analysis of adults (only) in the combined dataset, you can use the following syntax in SPSS.

```
select if age>=19.
weight off.
compute x=1.
aggregate outfile = * mode=addvar /break x /n1 = sum(wti_UKY1234)
/n2 = sum(wti_Y56) /n3 = sum(wti_Y78) /n4 = sum(wti_Y91011).
compute N = sum (n1, n2, n3, n4).
compute wti_UKY1234r = wti_UKY1234 * N / n1 * (4/11).
compute wti_Y56r     = wti_Y56     * N / n2 * (2/11).
compute wti_78r      = wti_Y78     * N / n3 * (2/11).
```

```
compute wti_91011r       = wti_Y91011      * N / n4 * (3/11).
compute wti_UKY1to11Ad = sum (wti_UKY1234r, wti_Y56r, wti_78r,
wti_91011r).
aggregate outfile = * mode=addvar /break x /mn=mean(wti_UKY1to11Ad).
compute WTI_UKY1to11Ad = WTI_UKY1to11Ad / mn.
exec.
```

The above syntax is generic i.e. it can be used to create bespoke weights for any subgroup; just change the criteria in the first line to select the relevant subgroup. (This includes the full combined dataset i.e. if the first line is excluded then the syntax will create appropriate weights for the full dataset.)

The resulting weights (`wti_UKY1to11Ad`) should have a mean of 1 and be non-missing for all cases in the combined dataset of adults. As above, you can check this by running descriptives on the weights:

```
desc wti_UKY1to11Ad.
```

Notes:

1. As above, the intermediate weights (`wti_UKY1234r, wti_Y56r, wti_Y78r, wti_Y91011r`) can be discarded/deleted.
2. If subgroup analysis is performed using weights produced for the whole dataset, Years 1-4, Years 5&6, Years 7&8 and Years 9-11 may not be in the correct proportion.