

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

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

Re-scaling is necessary because there were more participants *per year* in years 1-4 than in years 5&6 and years 7&8. In total, there are 6,828 participants in the Year 1-4 data but only 2,546 participants in Year 5&6 and 2,723 in Year 7&8 data. The ratio of these totals is (approx.) 2.68:1 and 2.51:1. Therefore, if the weights were combined into one variable without any further adjustment, years 1-4 would have more weight *per year* than years 5&6 and years 7&8.

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 the three of them by the total (combined) sum of the three weights;
- iii) Multiply the Year 1-4 weight by 1/2, Year 5&6 weight by 1/4 and the Year 7&8 weight by 1/4.

You can then combine the resulting weights 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 Year 1-4 weight by 6828, the Year 5&6 weight by 2546 and the Year 7&8 weight by 2723 ;
- ii) Multiply both weights by (6828 + 2546 + 2723);
- iii) Multiply the Year 1-4 weight by 1/2, the Year 5&6 weight by 1/4, and the Year 7&8 weight by 1/4;
- iv) Combine the resulting weights into one variable.

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

```
compute WTI_UKY1234r = WTI_UKY1234 * (6828 + 2546 + 2723) / 6828 *
(1/2) .
compute WTI_Y56r = WTI_Y56 * (6828 + 2546 + 2723) / 2546 * (1/4) .
compute wti_Y78r = wti_Y78 * (6828 + 2546 + 2723) / 2723 * (1/4) .
compute WTI_UKY1to8 = sum (WTI_UKY1234r , WTI_Y56r, wti_Y78r ) .
```

Running these commands will result in a new set of weights (WTI_UKY1to8) for analysing Years 1-8. They should have a mean of 1 and be non-missing for all 12,097 cases in the combined dataset. You can check this by running descriptives on the weights:

```
desc WTI_UKY1to8.
```

Notes:

1. The intermediate weights (WTI_UKY1234r, WTI_Y56r, wti_Y78r) 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 and wtn_Y78)

Blood weights (WTB_UKY1234, WTB_UKY56 and wtb_Y78)

RPAQ weights (WTR_UKY1234, WTR_UKY56 and wtr_Y78)

Spot- urine weights (wtsu_Y6 and wtsu_Y78) {* see note below}

* Note that spot-hour urine samples were only collected in Years 6-8 so when combining these weights, the correct proportions to use are 1/3 and 2/3 (rather than 1/2 and 1/4).

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 then 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 both by the total (combined) sum of the three weights;
- iii) Multiply the Year 1-4 weight by 1/2, the Year 5&6 weight by 1/4, and the Year 7&8 weight by 1/4;
- 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 could use the following syntax in SPSS:

```
select if age>=19.
desc WTI_UKY1234 WTI_Y56 wti_Y78 / stat = sum.  {*}
compute WTI_UKY1234r = WTI_UKY1234 * (5391.3596 + 2012.2877 +
2152.5750
) / 5391.3596 * (1/2).
compute WTI_Y56r    = WTI_Y56      * (5391.3596 + 2012.2877 + 2152.5750
) / 2012.2877 * (1/4).
compute WTI_78r     = wti_Y78      * (5391.3596 + 2012.2877 + 2152.5750
) / 2152.5750 * (1/4).
compute WTI_UKY1to8Ad = sum (WTI_UKY1234r , WTI_Y56r, WTI_78r).
desc WTI_UKY1to8Ad.                                     {*}
compute WTI_UKY1to8Ad = WTI_UKY1to8Ad / 1.552595.
```

* these lines are optional but required to produce the figures in the following commands.

The resulting weights (WTI_UKY1to8Ad) should have a mean of 1 and be non-missing for all 4,738 cases in the combined dataset of adults. As above, you can check this by running descriptives on the weights:

```
desc WTI_UKY1to8Ad.
```

Notes:

1. Again, the intermediate weights (`WTI_UKY1234r`, `WTI_Y56r`, `wti_Y78r`) can be discarded/deleted.
2. If subgroup analysis is performed using weights produced for the whole dataset, years 1-4, 5&6 and 7&8 may not be in the correct proportion.