

Weight Variables for the SDA Version of GSS

Composite Weight for All Years: COMPWT

The SDA version of the GSS provides a **composite weight (COMPWT)** that can be used for all the years of the GSS. It incorporates the various weights provided in the GSS for various years. **This is the default weight for the analysis programs in SDA.**

This composite weight adjusts for:

- The number of adults in the household
- Black oversampling in 1982 and 1987
- Problems with form randomization in 1978-1985
- Differential non-response in 2004 and later
- Post-stratification to Census variables in 2021

There is also a small adjustment factor, so that the number of weighted cases equals the number of unweighted cases in the data file.

This document explains how the composite weight was constructed out of the various weights included in the GSS dataset, depending on the year of the interview.

Weight Variables for 1972-2002

- **WTSSALL (With adjustment for number of adults)**
The weight variable WTSSALL includes an adjustment for number of adults in the household. (Since only one adult was selected to be interviewed, adults in households containing other adults have less chance to be selected than adults living alone.)
- **OVERSAMP**
In 1982 and 1987 there was an oversample of blacks. As a result, there are "too many" blacks in the sample for those years. The weight OVERSAMP weights down the blacks in those years and weights up the non-blacks, while leaving the total weighted number of cases in those

years equal to the unweighted number of cases. All of the cases in other years have a value of 1.0 on OVERSAMP.

The weight WTSSALL by itself does NOT adjust for this oversampling of blacks in 1982 and 1987. The weighted numbers of cases for those years still have "too many" black cases compared to non-black cases. However, COMPWT incorporates the OVERSAMP weight variable to handle this oversampling.

- **FORMWT**

In the years 1978-1985, there were some problems with the randomization of various forms of the questionnaire. The weight FORMWT adjusts for those problems and is recommended for use when variables that appear on only one form in those years are analyzed. All of the cases in other years now (in Release 3b) have missing data codes on FORMWT. So this weight variable can only be used for the years 1978-1985.

The weight WTSSALL by itself does NOT adjust for this problem, although it does include adjustments for form administration in 2004 and later. But COMPWT incorporates FORMWT to handle those adjustments for cases prior to 2004.

Weight Variables for 2004-2018

- **WTSSNR (With adjustment for Non-response and Number of adults)**
The weight variable WTSSNR is the same as WTSSALL (which adjusts for the number of adults in the selected household) **except that it also includes an adjustment for differential nonresponse in the various geographic areas in the sample for 2004 and later.**

In 2004-2018, the initial non-responders were divided at random into two groups. One group was set aside, and no further attempts were made to interview those cases. The other group was followed up intensively, in order to complete more cases. The eventual responders from the subsampled group (indicated by the variable PHASE) should therefore be weighted up, to compensate for the non-responders that were not interviewed.

The cases in the dataset for years prior to 2004 do not have this adjustment, because there was not enough information available to adjust for differential nonresponse in those years.

(Note that cases from 2004-2018 use the weight WTSSNR instead of the weight WTSSALL used for those cases in previous SDA versions of the GSS. This change was done in order to be consistent with the new 2021 weight (WTSSNRPS), which also adjusted for differential non-response. The correlation between WTSSALL and WTSSNR for the years 2004-2018 is .96. Therefore, this slight change in weight variables should have a negligible effect on any previous analysis results.)

New Weight Variables for the 2021 GSS (and earlier)

- Beginning in Release 1 for 2021, the GSS includes new weights that adjust for the number of adults in selected households and for differential non-response between Census Divisions. **The new weights are also post-stratified** to match the distributions of several Census variables, using a procedure called “raking”. (See the NORC GSS codebook for details. That codebook can be accessed from the title page of the SDA online codebook.)
- There are two such weights – WTSSPS and WTSSNRPS. Because the post-stratification included the nine Census Divisions (also used for the differential non-response adjustment), these two weight variables ended up being exactly the same in Release 3. Consequently, we could have used either of the two weights for constructing COMPWT for 2021. **We used WTSSNRPS because the name includes the characters ‘NR’, to make explicit that the weight includes a non-response adjustment.**
- As of Release 3, NORC extended these new weights back a few years. WTSSPS was extended back to include all the years since 2000. WTSSNRPS was extended back to include all the years since 2004. **However COMPWT was not changed for years prior to 2021, so that results for those years in Release 3b would remain the same as in Release 1 on the SDA site.**

The following COMPUTE command was used to generate COMPWT:

```
if (year eq 1982) $wt = wtssall * oversamp * formwt
else if (year eq 1987) $wt = wtssall * oversamp
else if (year le 1977) $wt = wtssall
else if (year le 1985) $wt = wtssall * formwt
else if (year le 2002) $wt = wtssall
else if (year le 2018) $wt = wtssnr
else if (year ge 2021) $wt = wtssnrps
endif
compwt = $wt * 68846 / 68884.819
```

(The factor at the end was a small adjustment to make the total number of weighted cases equal to the number of unweighted cases.)

(3/26/2023)