

Updated: September 21st, 2022

Log of Suspected Issues under Investigation in the GSS Cumulative File, Panel Datafiles, and Data Explorer

Curation of GSS survey data is an ongoing activity, and we rely on the user community to help us report possible issues with the data. Users are encouraged to report potential data issues to gss@norc.org. Making the data available to the community in a timely manner is a foundational principle for our work. For example, the GSS team holds no data embargos.

The GSS team corrects suspected errors as soon as possible but we do not have the resources to send out a notification each time one is confirmed and corrected. Therefore, we plan to release updated data about twice a year (Spring and Summer). However, if you have a question about a specific question and you believe you may have discovered an error or want to check in on one you notified us about, please get in touch with us directly so we can discuss it.

Current known errors that are scheduled to be fixed in the cumulative datafile, panel datafiles, and single-year datafiles are listed below, along with notes about anticipated corrections for the data.

This list will be constantly updated. Once the issues have been addressed and updated data has been released, the item will be removed from this list, and corrections will be included in release notes. Details on the actual updates will be provided in the GSS data documentation.

FORM, BALLOT

Years affected: 2002, 2010, 2012, 2016, and 2018

As reported previously in this log (2/25/22; 5/6/22), the GSS team was alerted to an unexplained gender imbalance among respondents who answered specific questions. The NORC team investigated this imbalance and verified that an error occurred in how respondents were assigned to one of the six versions of the GSS questionnaire (3 ballots by 2 forms). This error was caused by an unintended relationship between the age order of adults in the household and the questionnaire assignment. It resulted in age imbalances in households with multiple eligible-adult respondents, and as a byproduct in sex imbalances. Trends and basic relationships for the core GSS time-series questions that are asked on all three ballots and both forms are not impacted by the assignment error. Our preliminary analysis indicates that impact of the sex and age imbalances are mostly mitigated for many questions because they appear on multiple questionnaires. We are continuing to investigate any other potential analytic implications and working to develop a solution to minimize the differences. We plan to release a final methodological report by November 2022 and associated analytic tools that account for the gender imbalance in analyses.

Expected release: November 2022

PISTOL, RIFLE, SHOTGUN

Years affected: 2021

Datasets affected: 1972-2021 GSS Cumulative Datafile, GSS Data Explorer

Due to an automated processing error, only open-ended cases coded off of GUNSPEC were included in PISTOL, RIFLE, and SHOTGUN in 2021, resulting in what appear to be very low responses to PISTOL, RIFLE, and SHOTGUN. The GSS Team has verified that no information was lost due to this error, and responses will be corrected with the next data release.

Expected release: Late Summer 2022

SEXNOW

Years affected: 2018

Datasets affected: 1972-2021 GSS Cumulative Datafile, GSS Data Explorer

Label values are reversed for SEXNOW in 2018, with male respondents appearing as female and vice versa. SEXNOW1, the current version of the variable, is unaffected.

Expected release: Fall 2022

PRFTIMPV, INEQUAL6

Years affected: 2021

Datasets affected: 1972-2021 GSS Cumulative Datafile, GSS Data Explorer

The variable PRFTIMPV, first used in 2021, is identical to INEQUAL6, used in 1987 and 1996. For ease of future analysis, PRFTIMPV will be discontinued, and instead INEQUAL6 will appear in the 2021 data.

Expected release: Fall 2022

RINCOME

Years affected: 2021

Datasets affected: 1972-2021 GSS Cumulative Datafile, GSS Data Explorer, 2016-2020 GSS Panel

For 2020 panel data and 2021 cross-section data, RINCOME and RINCOM16 do not match each other. RINCOM16 is asked of respondents, while RINCOME is recoded to fit historical values. Due to a technical error in the recode, cases were assigned incorrectly to RINCOME. RINCOM16 is accurate and should be used instead of RINCOM when analyzing 2020 and 2021 data. RINCOME may be discontinued in a future release.

Expected release: Fall 2022

PAYDOC, PAYCLERK, PAYEXEC, PAYUNSKL, PAYCABNT, GIVDOC, GIVCLERK, GIVEXEC, GIVUNSKL, GIVCABNT

Years affected: 2021

Datasets affected: 1972-2021 GSS Cumulative Datafile, GSS Data Explorer

In the first release of the 2021 data, the PAY- and GIV- variables were released with topcoding of 100,000+; any respondent who gave either an expected or appropriate income above \$100,000 were collapsed. However, this topcoding collapsed an abnormally high number of responses for some of the variables in the series. The whole series is being re-examined, and a higher topcode will be assigned to some or all of the variables. At a minimum, both PAYEXEC and GIVEXEC will have their topcodes extended.

Expected release: Fall 2022

Some older variables, including VIOLWHTS, VIOLJEWS, VIOLBLKS, VIOLASNS, VIOLHSPS, VIOLSO, among others

Years affected: 1990, 2000

Datasets affected: 1972-2021 GSS Cumulative Datafile, not prior datafiles

When the 2021 cumulative datafile was initially prepared, some GSS variables which had not been asked in the last several rounds (prior to 2016) had the wrong labels applied to them, because of a procedural error. The error includes mislabeled variables and labels applied to variables with slight wording variation (e.g. "DK" vs "Don't Know"). At least 300 variables which have not been asked since before 2016 have 3 or more incorrect labels in the more recent cumulative dataset that includes the 2021 data. Datasets produced prior to the year to 2021 are not affected by the error. Included among impacted variables are the VIOL-series listed in the header for this issue, which were the first variables where this issue was brought to our attention. The GSS team has systematically reviewed and reassigned labels applied to the cumulative file produced for the 2021 data, and will be reassigning labels to all affected variables.

Expected release: Fall 2022

Exports using the GSSDE-generated .dat file

Years affected: All years

Datasets affected: GSS Data Explorer

While generating .dat files and associated syntax files for export, the automated syntax files do not properly align labels and values with columns of the created .dat file, causing the syntax files to incorrectly process the associated .dat file. As a result, variables in the .dat file do not get properly assigned variable labels, value labels, or missing values. This issue affects both SPSS and Stata exports from the GSS Data Explorer.

Expected release: Fall 2022

WANT MORE INFORMATION?

Call Us
(888) 837-8988

Email Us
GSS@NORC.org

Visit Us Online
<http://gss.norc.org>