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.

**SAS formats**

<p>| Years affected: | All years |</p>
<table>
<thead>
<tr>
<th>Datasets affected:</th>
<th>All data files</th>
</tr>
</thead>
<tbody>
<tr>
<td>Due to the format limitation in SAS, the Release 3 of the GSS 1972-2021 data uses two format catalogs (format.sas7bcat) instead of one (please refer to Appendix B in the <a href="#">2021 GSS Codebook</a> for more information). This two-format setup is causing inconvenience to SAS users when they use the GSS cumulative data. The GSS team is working on a solution to simplify the data use for SAS users.</td>
<td></td>
</tr>
<tr>
<td><strong>Expected release:</strong></td>
<td>Fall 2023</td>
</tr>
</tbody>
</table>

**Household enumeration form variables, XNORCSIZ, SRCBELT, SIZE**

<p>| Years affected: | 2021 GSS |</p>
<table>
<thead>
<tr>
<th>Datasets affected:</th>
<th>All data files</th>
</tr>
</thead>
<tbody>
<tr>
<td>Due to issues raised by the shift from F2F to web data collection in 2021, as well as changing data cleaning processes, variables from the household enumeration form (OLD1-14, GENDER1-14, MAR1-14, and RELATE1-14, as well as its derivatives) and geocoded variables (XNORCSIZ, SRCBELT, SIZE) are not yet available for 2021. Please note that due to the nature of the web-only data collection, AWAY1-14 and WHERE1-14 will never be available for 2021, as they were not asked.</td>
<td></td>
</tr>
<tr>
<td><strong>Expected release:</strong></td>
<td>Fall 2023</td>
</tr>
</tbody>
</table>

**Exports using the GSSDE-generated .dat file (I)**

<p>| Years affected: | All years |</p>
<table>
<thead>
<tr>
<th>Datasets affected:</th>
<th>GSS Data Explorer</th>
</tr>
</thead>
<tbody>
<tr>
<td>GSS Data Explorer utilizes negative values to denote missing values (e.g., don’t know, inapplicable), but many statistical software packages (e.g., SAS, Stata, SPSS) do not recognize these values as missing. Users may need to recode these negative values to missing, according to their chosen package’s specifications, to exclude these cases from analysis.</td>
<td></td>
</tr>
<tr>
<td>In addition, the associated syntax files generated for Stata try to assign decimal values a label, even when the value is just a</td>
<td></td>
</tr>
</tbody>
</table>
number. Stata does not allow decimal values to be assigned labels and the associated syntax will produce an error. The GSS team is actively working to find a fix for these issues.

**Expected release:** Summer 2023

**Grid Experiment and FORM assignment**

**Years affected:** 2021

**Datasets affected:** All data files

Phone mode cases on the grid experiments are assigned to the non-grid version of the experiment, regardless of the assignment of FORM.

**Expected release:** Fall 2023

**Exports using the GSSDE-generated .dat file (II)**

**Years affected:** All years

**Datasets affected:** GSS Data Explorer

The GSS Data Explorer occasionally experiences off-by-one alignment errors with .dat file exports, which can cause issues reading in variables and data from the affected files.

The GSS team is actively working to find a fix for these issues.

**Expected release:** Summer 2023

**FILEVERSION Label**

**Years affected:** 2022

**Datasets affected:** All data files

FILEVERSION for the 1972-2022 R1 GSS Cross-section file is unlabeled. FILEVERSION for the 1972-2022 R1 datafile is set to 7222.1, indicating that it is the first release of the 1972-2022 cross-section file. Both the variable and the label will be updated with future releases.

**Expected release:** Fall 2023

**WLTHWHTS, WLTHBLKS, WLTHHSPS, WORKWHTS, WORKBLKS, WORKHSPS, INTLWHTS, INTLBLKS, INTLHSPS**

**Years affected:** 2018, 2021 (2021 and 2022 datafiles only)

**Datasets affected:** All data files

The affected variables have frequencies which are not consistent between released codebooks and the datafile, as well as retain unused punches such as 98 and 99, as well as other labeling issues. This issue was discovered by the Roper Center for Public Opinion Research.

**Expected release:** Fall 2023

**DECAUTO, DECBED, DECGIFT**

**Years affected:** 1996 (2021 and 2022 datafiles only)

**Datasets affected:** All data files

The affected variables have switched the labels “IAP” and “Not applicable - have not faced this decision.” This issue was discovered by the Roper Center for Public Opinion Research.

**Expected release:** Fall 2023

**SCINEWS3**

**Years affected:** 2021 (2021 and 2022 datafiles only)
Datasets affected: All data files
The affected variable has frequencies which are not consistent between released codebooks and the datafile, as well as retain unused punches such as 98 and 99, as well as other labeling issues. This issue was discovered by the Roper Center for Public Opinion Research.

**Expected release:** Fall 2023

**ETH1, ETH2, ETH3**

*Years affected:* All years (2021 and 2022 datafiles only)

**Datasets affected:** All data files

The affected variables have incorrect labels, substituting "uncodeable" for "American only". This issue was discovered by the Roper Center for Public Opinion Research.

**Expected release:** Fall 2023

**PASEI10EDUC, MASEI10EDUC, SPSEI10EDUC, COSEI10EDUC**

*Years affected:* All years (2021 and 2022 datafiles only)

**Datasets affected:** All data files

The affected variables have incorrect labels, removing the "Don't know" and "No Answer" labels assigned to 99.8 and 99.9, respectively. This issue was discovered by the Roper Center for Public Opinion Research.

**Expected release:** Fall 2023

**OTHER**

*Years affected:* 2021, 2022

**Datasets affected:** All data files

The label for value 219, "Messianic Jew" is missing. This issue was discovered by the Roper Center for Public Opinion Research.

**Expected release:** Fall 2023

**WWWHR**

*Years affected:* 2012

**Datasets affected:** All data files

One response is above the maximum possible threshold for hours in a week, and will be recoded. This issue was discovered by the Roper Center for Public Opinion Research.

**Expected release:** Fall 2023

**RECYCLE**

*Years affected:* 2021

**Datasets affected:** All data files

Punch 5 should have the label "Recycling not available where I live." This issue was discovered by the Roper Center for Public Opinion Research.

**Expected release:** Fall 2023

**VETYRS**

*Years affected:* All years

**Datasets affected:** All data files

The label for "Some, don't know how long" contains a typo. This issue was discovered by the Roper Center for Public Opinion Research.
Research.

**Expected release**: Fall 2023

---

**LETINASN, LETINHSP**

**Years affected**: 2022

**Datasets affected**: All data files

The questions and response options for LETINASN and LETINHSP are not consistent with the versions asked in 2000. The GSS team is working to generate new variables for the 2022 versions of the questions.

**Expected release**: Fall 2023

---

**CONRINC**

**Years affected**: 2010

**Datasets affected**: All data files

When RINCOM06 is any value but the top value (25), the corresponding value for CONRINC is missing. The GSS team is actively working to correct this issue.

**Expected release**: Fall 2023

---

**CONINC, CONRINC, REALINC, REALRINC**

**Years affected**: Multiple years

**Datasets affected**: All data files


The values assigned to the top-coded categories may not be consistent for certain of these variables for certain years. This affects 1994, 2010, 2018, 2021, and 2022.

The GSS team is actively working to correct these issues.

**Expected release**: Fall 2023

---

**RINCOM16**

**Years affected**: 2016, 2018, 2021, 2022

**Datasets affected**: All data files

In 2021 and 2022, a value of 0 has the corresponding label of “No income”. In previous years (2016 and 2018) and with other RINCOM variables (e.g., RINCOM06), these participants were coded as missing (.) with the label “Inapplicable”. The GSS team is working to make this more consistent.

**Expected release**: Fall 2023

---

**NEA and GSS Next Follow-on Studies**

**Years affected**: 2022

**Datasets affected**: All data files

Due to a programming error, a number of GSS baseline respondents were incorrectly flagged in the final data as being eligible (or ineligible) to be in the NEA Arts Participation and GSS Next follow-on studies. This resulted in an undercount of 271 eligible baseline respondents for the NEA follow-on and an overcount of 72 for the GSS Next follow-on. Given eligibility is related to the creation of follow-on study weights (WTSSNR_NEA, WTSSNR_NEXT, WTSSNRPS_NEA, WTSSNRPS_NEXT), follow-on study weights need to be recalculated for both studies along with revised retention and response rates.

**Expected release**: Fall 2023