HSC Business Services Organisation
Honest Broker Service

RESEARCHER OUTPUT CHECKING PROCEDURES

VERSION	1.3
CLASSIFICATION	Open
OWNER	Alan Harbinson
DATE ISSUED	11th February 2026
STATUS	Final
REFERENCE	HBS Researcher Output Checking Procedures

Researcher Output Checking

The process of checking outputs assures that the 'Safe Outputs' part of the 5 Safes framework employed by the Honest Broker Service (HBS) is upheld, allowing the results of analysis performed by researchers to be released in a form that respects the privacy and confidentiality of HSC Service Users.

Poorly presented outputs that have not been properly checked by researchers can result in unnecessary work for HBS staff. This is time that could be spent working on outputs or project data for other projects. Researchers should adhere to the rules set out in this guidance. The general rules should be adhered to at all times and specific rules applied where appropriate.

It is the researcher's responsibility to not only apply Statistical Disclosure Control (SDC) but also to make sure their outputs are clearly laid out and easily understood. If an output does not adhere to SDC rules or is not easily understood it will be rejected and will need to be resubmitted by the requestor after applying disclosure controls.

The rules adopted by HBS are industry standard but if in doubt please speak to a member of HBS Staff before submitting your request.

General Rules

- Any hidden data within an output will be investigated as a possible breach of process (Examples could include hidden tabs, hidden cells or text colour changed to match the background or data copied into unusual cell ranges such as at the extreme edges of a worksheet).

- Deliberate attempts to export hidden data will lead to sanctions being applied to the researcher and potentially on their institution.

- Only submit an output request for results you actually need. Submitting a large amount of tables or leaving in test results because they were produced alongside the results you need is very time consuming.

- If there is more than one table or piece of analysis within an output, check for disclosure by differencing between the counts to make sure that small numbers cannot be worked out. This can apply to both frequency tables and statistical analysis in which counts are reported in the outputs.

- If you need to submit a frequency table to show how SDC has been applied but don't need the table to be released this should be left in a separate file in your research folder. The name and location of this file should be noted on the output request form.

- Wherever possible if you need to make a graph you should clear the counts needed to make the graph and create this outside the secure environment.

- Graphs are permitted however should be based on a flat file with no embedded data contained within them. The underlying data must be provided in frequency tables so that they can be checked by output checkers.

- Do not resubmit tables/results you have already had cleared or add results to a previously cleared file. You should only be submitting new results for clearance.

- Note that current practice allows for a maximum of 20 working days for clearance of all final outputs, with the exception of conference/journal abstracts for which five working days are required. Safe Haven Outputs will also require 5 working days for clearance.

Specific Rules

Frequency/Magnitude tables

- Make sure you clearly state what cohort the table relates to
- Make sure variables are named in a meaningful way
- Make sure there are no cell counts of less than 10 (unweighted). Cell counts relating to missing data is exempt from this rule.

- Zeros counts are only allowed if they are structural zeros. Structural zeros occur when a count can only take the value of zero e.g. the number of COVID-19 Vaccinations administered in 2018, as the vaccine was not developed and approved for use until December 2020
- Makes sure small counts can't be calculated from totals in the table or from other tables in the output. If the total is suppressed, included in a previous table or output either within this output or one which has been previously cleared. Make sure there is no risk of class disclosure contained in the table
- Make sure the figures are not dominated by any individual entity in the data.
- Frequency tables are not needed for weighted or standard statistics but zero counts should be removed.

Descriptive statistics

- These typically include but are not limited to mean, median, variance, standard deviation, skewness, ratios, percentiles etc.
- Make sure you clearly state what cohort the statistics relate to
- Make sure frequency tables have been provided, showing that the underlying count of individuals for any descriptive statistic is not less than 10.

Higher moments of distribution

- These typically include but are not limited to variance, skewness, kurtosis etc.
- Make sure you have shown all modelled outputs have at least 10 degrees of freedom and at least 10 individuals have been used to produce the model.

Graphs

- Make sure frequency tables have been provided, showing points/bars on a graph relate to a value shared by at least 10 individuals.
- Make sure both axes are clearly labelled with appropriate scales.
- Make sure the graph has been submitted as a 'fixed' picture, with no data attached.
- Make sure you show any step change in Kaplan-Meier curve represents at least 10 individuals.

Regressions Coefficients

- Make sure you show the model has greater than 10 degrees of freedom
- Make sure the intercept has been removed or a frequency table has been provided for univariate regressions.
- Make sure Residuals have been removed.

Summary and test statistics

- These typically include but are not limited to R^2 and variations, estimated variance, information criteria (e.g. AIC, BIC), individual and group tests and statistics (t, F, chi-square, Wald, Hausman etc.)
- Make sure you show the model has greater than 10 degrees of freedom.

Factor analysis

- Make sure you show the factor is related to at least two different observed variables.

Correlation coefficients

- Make sure you show there is a minimum of 10 unweighted units underlying each correlation coefficient and that the correlation coefficient is not equal to -1, 0, +1.

File Types

We have the capacity to check a wide variety of file types in the safe setting. These include:

Data files, such as:

- .csv, .tsv (generic)
- .xls, .xlsx (Microsoft Excel)
- .sav (SPSS)
- .dta (STATA)
- .Rdata, .Rda, .Rds (R) – when they contain a single dataframe/table

Log files, such as:

- .log (generic)
- .smcl (STATA)
- .spo, .spv (SPSS)

Image files, such as:

- .bmp, .gif, .jpeg, .png (generic)
- .gph (STATA)

Document and presentation files, such as:

- .doc, .docx (Microsoft Word)
- .ppt, .pptx (Microsoft PowerPoint)

Code files, such as:

- .do (STATA)
- .sps (SPSS)
- .R (R)
- .py, .py3 (Python)
- .ipynb (Jupyter Notebook)

Files associated with coding packages, such as:

- .ado, .mata, .pkg, .sthlp, .toc (STATA)
- .tar.gz, .tar (R)

Miscellaneous other files, including:

- .txt
- .pdf
- .TeX

There are some file types which we do not have the capacity to fully check. **These file types may not be cleared for any outputs from the safe setting**, regardless of their apparent content:

- Markdown files – with .html extension if in Hypertext Markup Language, .markdown, .md, .markdn or .mdown extensions if in Markdown language, with .Rmd extension if in R or with .Rnw extension if in LaTeX.
- R project files – with .Rproj extension.
- Shapefiles – with .shp, .shx or .dbf extensions or, more occasionally, with .prj, .sbn, .sbx, .fbn, .fbx, .ain, .aih, .ixs, .mxs, .atx, .shp.xml, .cpg or .qix extensions.
- JavaScript Object Notation files – with .json extension.
- R presentation files – with .RPres extension.
- R data files – with .Rdata, .Rda, .Rds extensions – that contain data in formats other than dataframes/tables

Researchers may convert their data back into these formats using cleared outputs outside the safe setting, and then submit them as Final Outputs for review.

If in doubt refer to Handbook on Statistical Disclosure Control for Outputs and Data without Boundaries Guidelines for Output Checking.

Handbook on Statistical Disclosure Control for Outputs

Data without Boundaries Guidelines for Output Checking