

# Installation Guide of Tool for Analysis of Interim Datasets ESS11

## Requirements and Implementation

Roberto Briceno-Rosas<sup>1</sup>, May Doušak<sup>2</sup>, Paulette Flore<sup>3</sup>, Joost Kappelhof<sup>3</sup>

<sup>1</sup> GESIS – Leibniz Institute for the Social Sciences

<sup>2</sup> University of Ljubljana

<sup>3</sup> SCP – The Netherlands Institute for Social Research

Introduction .....	1
General description of the Tool for Analysis of Interim Dataset.....	2
Requirements.....	2
Installation procedure.....	3
Implementation.....	3
Testing.....	4
Troubleshooting & FAQ.....	4

## Introduction

In ESS Round 11, National Teams of participating countries are expected to analyse an interim dataset from the main questionnaire to check for data quality and detect interviewer-related data quality issues arising during the interviewing process with the support of the CST. The objective of this measure is – in case of systematic interviewer-related problems during the interview process – to detect these problems early and if possible intervene and correct them while the data collection process is still on-going (as defined in the ESS Specification, see below).

*“After approximately one third of the planned interviews have been completed, the Survey Agency is required to provide an interim data file to the NC for initial checking of interview quality. The NC will be provided with syntax by the CST and asked to use the syntax to flag potential problematic interviews and/or interviewers and discuss the results with the CST. This information should be used in discussion with the Survey Agency with respect to ongoing quality control procedures. Note that the interim data file should contain a numeric interviewer identifier to enable the Survey Agency to identify any interviewers flagged.”<sup>1</sup>*

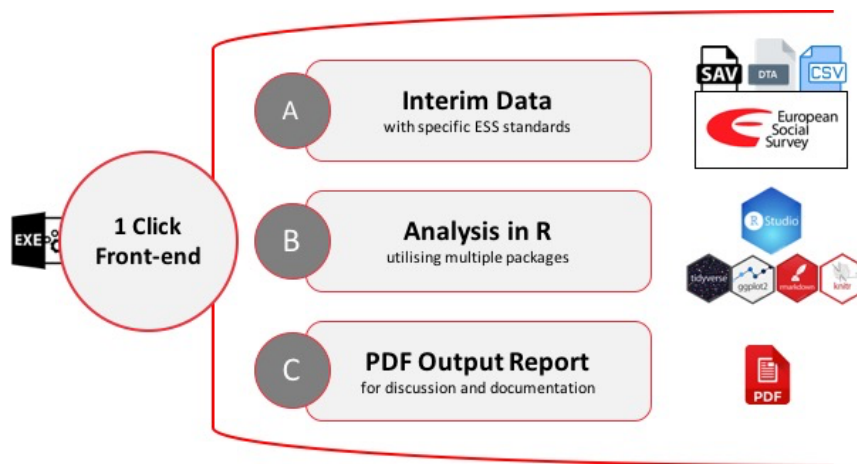
As mentioned in the specifications, the CST has developed a tool for the analysis of the ESS interim datasets to facilitate this process. This paper informs about the requirements for using the tool and how it should be installed and implemented.

---

<sup>1</sup> More information available in the *Round 11 Survey Specification for ESS ERIC Member, Observer and Guest Countries*

## General description of the Tool for Analysis of Interim Dataset

The tool allows for standardised analysis of the ESS interim datasets in a user-friendly manner. It takes the ESS interim dataset, runs a predefined syntax over data, and presents the results in textual and graphical format. As shown in the figure below, the tool has three key components which are embedded in a Windows application: (a) the ESS interim datasets, (b) the syntax with the code for the analysis, (c) a PDF report with the results. This allows keeping the requirements and effort for the National Team to a minimum. The Windows application provides a front-end that allows users to run the tool in an intuitive manner. The syntax for the analysis has been written in R, which is a free software environment for statistical computing and graphics, and National Teams can adapt it if they wish to. The software also allows for more transparency about the indicators and for easier collaboration.



## Requirements

### Data requirements

- *Main Dataset:* Interim Dataset from the Main Questionnaire with structure as specified in ESS11 Data Protocol, including the specified variables names and formats (see ESS11 Data Protocol in myESS<sup>2</sup>). Please note that the structure of the dataset is fundamental for the tool to work. We recommend setting up the dataset to the specified format from the start of fieldwork. The amount of time required to convert the interim dataset to specified format at a later stage can be quite substantial and in turn will be a net loss in the effectiveness of implementing this measure in a timely manner.
- *Interviewer Dataset:* A dataset with the case identification number (IDNO) and its respective interviewer identification number (ITNUM).
- Both datasets need to be on one of following formats: SAV, DTA, SAS, or CSV.

### System requirements

- Windows 7 or newer
- .NET 4 or newer
- 2 GB free disk space

<sup>2</sup> Please note that the final Data Protocol should be used. Alerts with changes to the current version are planned and will inform NCs of any changes to the current Data Protocol.

## Installation procedure

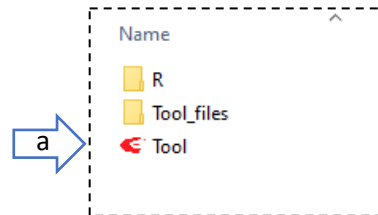
The tool resides in its folder and works without installation.

- a) User should simply unzip the downloaded file anywhere on disk.
- b) A new folder called "Tool" will be created on destination.

Depending on the speed of computer, unzipping might take a few minutes.

## Implementation

- a) Open the newly created folder "Tool" and run "Tool" file with the ESS logo (EXE file):



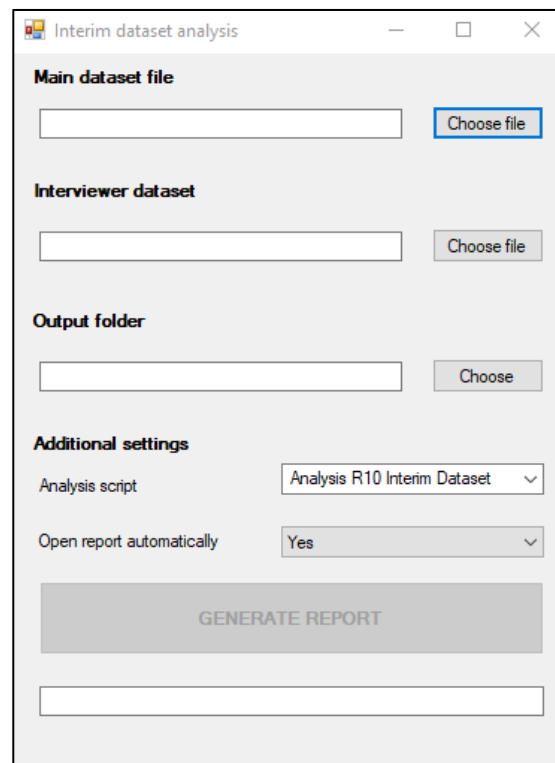
- b) The application opens. The interface of the tool is shown on the right.

- c) Please provide the *Main dataset* and the *Interviewer dataset* files, which should comply with the requirements previously mentioned. Both datasets need to be on the same file format (SAV, DTA, SAS, or CSV files)

- d) Please choose output folder where the report PDF will be saved.

- e) Select the correct script for analysing the selected datasets. For R11, select "Analysis R11 Interim Dataset"

- f) Click "Generate". The tool will indicate that is generating the report below.  
*Depending on your computer, report generation might take anywhere from 5 to 30 minutes. In some cases, it might take up to an hour.*



- g) Retrieve the PDF Report from the designated folder. Please take the time to review the report and discuss any issues with the survey agency and the CST. Share the report with your Country Contact.

- h) Note that an annex folder will also be copied. It contains more detailed results for each indicator as CSV files to help investigate the issues. You can also find the csv files in the folder Tool\_files\Annex within the tool.

## Testing

Testing the functionality of the tool: You can test whether the tool runs on your computer, whether the installation was correct before implementing it on R11 data, and get a feeling of how the tool works. For this purpose, the tool includes a dummy R11 dataset stored within the folder Tool\_files/DATA\_R11). You can run an analysis over this dataset using the script “Example Report R11”. You can also test the functionality of the tool with single country published data from R10. These files have been prepared and stored within the folder Tool\_files/DATA\_R10. Please run this analysis using the script “Example Report R10” from the dropdown menu in step E of the implementation.

Testing the consistency of the tool with the country-specific data structure: You can test the data structure consistency with early data from the field or use a dummy dataset with random values. If possible, you can run the analysis before you reach 1/3 of the expected interviews. The idea is to solve any issues early and facilitate the process once the mandatory number of interviews has been reached. It should be noted that if the dataset has very few cases, some of the analyses or figures might not work properly. If pre-test data was collected via the same CAPI system as the main survey, you might be able to use the pre-test data to run a preliminary test of the compatibility of the tool.

## Troubleshooting & FAQ

### **Unzipping the tool file takes a long time**

To save time, it is recommended to extract the content of the zip file directly into the desired folder on the computer's drive from the interface of any unzipping programme (instead of double clicking). The reason is that the tool folder includes all the necessary files to analyse the data without installing any programs on your computer. Therefore, it contains thousands of files with the necessary code for running the analysis and producing the report (including the R statistical software and LaTeX document preparation system).

### **Do I need to save the data on my local computer?**

A secure storage of the data on the local drive of your computer is the easiest way to run the analysis. If you need to remotely analyse the interim data stored on a server, it is recommended to save the tool on the server and run the analysis on the server's computer. Ideally, the tool and the data are stored on the same computer, either both on the server or on the local computer to avoid longer analysis time and possible error in the exchange process (between the server and the local computer).

### **Tool is stuck in the process of analysing the data**

Depending on your computer specifications and the size of the data, the compiling the report can take from 5 min to half an hour. In some extreme cases, it can take up to a full hour to complete the analysis. However, the tool might become unresponsive in some cases. If so, please force quit the tool's programme and restart the process. Please make sure that the data and the tool are stored on the same drive to reduce (see server issues on the previous FAQ). If the problem is not solved, please contact the CST.

### **What to do if I get an error while trying to produce the pdf report?**

There are generally two reasons why the tool might fail to compile the report: (a) either errors caused by the interaction of the tool and the computer or (b) errors caused by interactions of the data and analytical script. Errors of interaction between the tool and the computer are usually related to the folder names and paths. Errors of interaction between the data and analytical script are usually related to data format, variable names, and variable content.

To try to solve the most common error, please try the following steps:

1. Please make sure that the folder paths where the interim datasets and the tool are stored, as well as the path of the output folder for the report) do not contain any spaces (substitute spaces for \_) or any non-English letters (like ä, æ, ß, ñ, ø, ç, etc).
2. Please check that the data structure of the interim dataset meets the specification of the ESS11 Data Protocol. The data does not need to be perfect, but it needs to contain the expected variable with the respective variable names and format, including the time stamps.
3. Run the script “Inspect Data Structure” included in the tool. This will produce a simple PDF report describing the variables in the dataset. Please share this with your Country Contact and the CST to receive appropriate assistance in trying to solve the issues.

### **Results saved as CSV files in the Annex folder do not coincide with the results shown in the report**

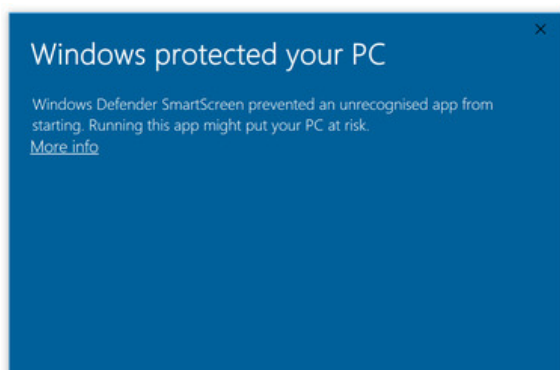
Please make sure that the csv file date creation date coincides with the creation date of the report. Detailed results of the analysis are saved in csv file format in the annex folder every time you run the analysis. Running the analysis again should overwrite old results. However, if an issue during the compiling process hinders the ability of the tool to overwrite the files (e.g. file open, error in the data, etc.), the newest version might not be saved.

### **The report shows some errors, e.g. incomplete or empty figures or tables**

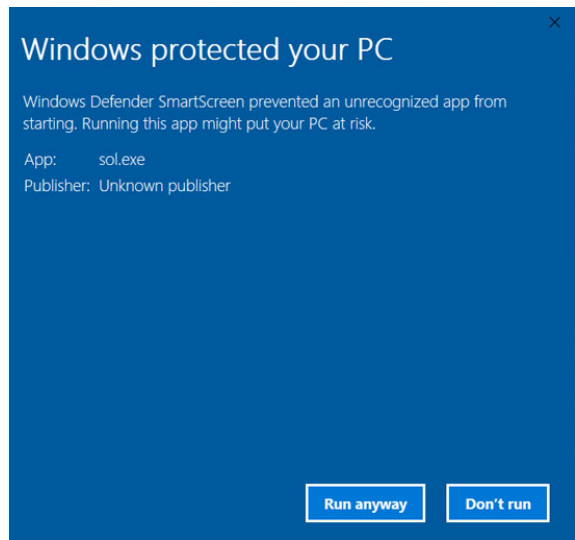
There are a variety of reasons for such errors. Usually these are either consistency errors (consistency between data structure and analytical syntax) or content requirement error (the values do not meet the necessary standards for the indicator to work). Consistency errors need to be investigated on a case-by-case level and should be corrected. Content requirement errors do not need to be corrected. The latter occurs when an indicator imposes some requirements over the values of specific variables to work properly, for example, a minimum number of interviews per interviewer, the removal of very extreme values, etc. Note that content requirements are noted in the text, figure, and tables of the report. Checking the csv files can help determine whether the issue is due to consistency issues or content requirements.

### **Windows Smart Screen when first running the tool**

First time you run it, Windows Smart Screen protection might pop up:



Should this happen, please click “More Info” and press the button “Run Anyway”:



### **Running the Tool in another operating systems like MacOS and Linux**

Currently, the CST only provide support for running the tool in Windows. However, it is possible to run the code and produce the PDF in MacOS and Linux. Some basic knowledge in R is required. For this purpose, users can extract the syntax (RMD code) and related files in order to run the code in their computers using R. Please note that users will need to install the necessary software (R, R Studio, and MikTex) and assigned the respective path for the datasets manually.

### **Can I show someone from the CST my screen via a virtual meeting call to receive assistance running the tool.**

The national teams are expected to run the tool following the guidance from the CST without the need of direct assistance. In case that troubleshooting and written exchange via MyESS is not sufficient to solve the issues, a virtual meeting might be arranged. However, screen sharing while handling interim datasets should not be taken lightly. There are steps that can be taking to adjust the DPA between the CST and the national teams to allow remote assistance via call while assuring data protection. If you wish to receive this type of assistance, please contact the CST.

### **Can I run a part of the analysis only? I am having problems with one part of the report or some variable like the timestamps**

To run individual sections of the analysis separately, you will need to manually modify the "Analysis\_InterimD\_R11-sav.Rmd" files in the tool\_files folder. However, due to co-dependency in the analytical procedures of script, excluding parts of the code might lead to errors. If you need support doing so, please contact your Country Contact or the CST team for assistance.

### **Who to contact in case of problem**

In case of other problems or questions, please contact the CST via myESS or your Country Contact.