## Startup kit

# Checklist table 1
# File(s) required
# for data-linkage

For the purpose of data-linkage and producing the research file for your project, you must transmit certain data to the ISQ.

The table in this document sets out the instructions to follow and the checks to be made before transmitting the files containing these data to the ISQ.

### Important

- The files required for the production of the research file must be supplied to the ISQ via **two separate transfers**. These files must be sent using a secure exchange method.

  - The **first transfer** must include the file(s) required for data-linkage, meaning files that contain the identification data needed for matching with the relevant files.

  - The **second transfer** must include the file(s) containing the research variables, meaning files that contain the data to be included in the research file.

  The checklist table on the next page relates to the files sent in the **first transfer**.

- Files must not **under any circumstances** be sent via email.

- **You must include the completed checklist table below** when transmitting the files required for data-linkage.

**It is the research team's responsibility to ensure the quality and consistency of the information contained in the files transmitted to the ISQ. The ISQ will not perform this validation.**

# First transfer: File(s) required for data-linkage

**Description:** File(s) required for data-linkage contain the information that can directly identify an individual or organization (e.g. last name, first name, address, health insurance number, date of birth, sex, date of death, etc.) and any other information useful for data-linkage (including the index date, i.e. the date on which a reference event for a particular individual occurs. This could be, for example, the date of a first cancer diagnosis or the date of inclusion in the study). An individual's current or past addresses (full or partial), phone number, and the last and first name of their parents or partner could also be particularly useful for data-linkage. Each individual included in a match file must have a **unique research identifier** (anonymous number). This identifier must always appear in the files transmitted, because the ISQ uses it to create a **unique number specific to the individual** (NOINDIV) in order to link all the information associated with the same individual included in the files produced by the ISQ.

## ⚠ Checks to be completed by the research team

| | Yes | No | Comments |
|---|---|---|---|
| **Anonymous numbers and individuals** | | | |
| Only information on individuals who match the selection criteria defined in the research project are included. | | | |
| The number of records in each file transmitted to the ISQ is _____ , and these records are for _____ separate individuals. This is the number of individuals agreed upon with the ISQ. | | | |
| Each identification file contains an anonymous number uniquely representing each individual (this number must not be linked to a person's identity). Each anonymous number corresponds to a single individual. | | | |
| If several anonymous numbers could be linked to the same individual, please specify it in the "Comments" field. Otherwise, the ISQ could consider that they are separate individuals. | | | |
| The unique research identifier (anonymous number) makes it possible to identify the same individual within one or several files. | | | |
| If a file has more than one line for the same individual (e.g. several addresses), the individual's anonymous number is repeated on each line. Leave blank if this does not apply. | | | |

| | Yes | No | Comments |
|---|---|---|---|

**Index date**
An index date is the date on which a reference event for a particular individual occurs. It is used to extract data on an individual for a duration determined based on this event. The index date may be, for example, the date of inclusion in the study, the date of a given diagnosis, the date of a visit at a clinic, etc.

For your project, is there an index date to be taken into account?

If yes, do the file(s) transmitted contain one or more index dates?

If there are one or more index dates, please clearly define each one in the "Comments" field.

**Variables**

The file(s) contain one item of information per column.
*e.g. Last name = Column 1; First name = Column 2; etc.*

The names of the variables (column titles) do not exceed 30 characters and do not contain any special characters or spaces.

A description (label) is provided for each variable.

Each variable follows a format that is standardized throughout the file(s) (e.g. YYYY-MM-DD for dates).

All variables required for data-linkage that have been agreed upon with your consulting analyst are provided.

| | Yes | No | Comments |
|---|---|---|---|
| **File format** | | | |
| The file(s) are in one of the following formats: SAS, CSV, XLS or XLSX. Files in text format (.txt) are acceptable, but not preferred. If the computer used to compress the file(s) works with the macOS operating system, you must ensure that the compressed file created can be extracted and read under the Windows operating system. Certain types of files, such as DMG files, cannot be opened in Windows.. | | | |
| The file(s) are compressed into a password-protected .zip file (it is not necessary for the files themselves to be password-protected). | | | |
| The name of the compressed file includes the lead researcher's last name, the file number assigned by the ISQ (e.g. 40001A01), and the term "Appariement". *E.g. "TREMBLAY_40001A01_Appariement"* | | | |
| If the compressed file exceeds 2 GB, please specify it in the "Comments" field. | | | |