

Startup kit

Checklist table 2

File(s) containing the research variables

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 second transfer.

- Files must not **under any circumstances** be sent via email.
- **You must include the completed checklist table below** when transmitting your files containing the research variables.

Content of the file(s) containing the research variables

- The file(s) must only contain the variables required for the analyses to be carried out as part of the research project.
- Any information that can directly identify any individual, business, organization or association cannot be deposited in the ISQ's secure environment due to the ISQ's legal obligations.

ACT RESPECTING THE INSTITUT DE LA STATISTIQUE DU QUÉBEC

Section [25](#). The Chief Statistician, assistant chief statisticians, public servants and any other person whose services are used by the Chief Statistician in the exercise of the Chief Statistician's functions shall not disclose or cause to be disclosed, by any means whatsoever, any information obtained under this Act if disclosure would allow information to be associated with a specific person, enterprise, body or association.

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.

Second transfer: File(s) containing the research variables

Description: File(s) containing the research variables contain the information on the individuals to be included in the research file. These file(s) must not contain any information that can directly identify an individual or an organization (e.g. last name, first name, address, health insurance number, phone number, institution or business name, etc.). Each individual in these files must be identified by 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
Variables			
The file(s) only contain information relating to the individuals included in the files sent in your first transfer.			
For each separate individual, the research variables are associated with an anonymous number. This anonymous number is the one used for the same individual in the file(s) required for data-linkage.			
In the “Comments” field, specify: <ul style="list-style-type: none"> The number of separate individuals in the file(s) required for data-linkage. The maximum number of separate individuals in the file(s) containing the research variables. 			
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 the file(s) contain more than one line for the same individual (e.g. several events), the individual's anonymous number is repeated on each line. Leave blank if this does not apply.			

Continued on page 4

	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			
<ul style="list-style-type: none"> Do the file(s) contain any real numbers, such as institution, file or questionnaire numbers? Ideally, the file(s) should not contain any such data. If they are necessary, enter the name of each variable that contains real numbers as well as the name of the file(s) where they are found in the "Comments" field. 			
The file(s) do not contain any information that could directly identify:			
<ul style="list-style-type: none"> an individual (e.g. last name, first name, address, health insurance number, phone number, etc.), whether this person is a participant in a study or anyone else (doctor, interviewer, teacher, etc.); a business (business name, phone number, mailing address or email), an organization (name or real number of the institution) or an association. 			
Important! Pay particular attention to any type of comment in the files that may contain identifying information. If applicable, please remove any information that is not required for the analysis.			
The file(s) contain one item of information per column. <i>E.g. Date 1 = column 1; Date 2 = column 2; etc.</i>			
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.			

Continued on page 5

	Yes	No	Comments
Each variable follows a format that is standardized throughout the file(s) (e.g. YYYY-MM-DD for dates).			
All variables to be integrated into the research file that have been agreed upon with your consulting analyst are provided.			
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. 40001A ⁰¹), and the term "Variables_recherche". E.g. "TREMBLAY_40001A01_ Variables_recherche"			
If the compressed file exceeds 2 GB, please specify it in the "Comments" field.			