



EUROPEAN COMMISSION
EUROSTAT

Directorate F: Social statistics
Unit F-5: Education, health and social protection

Luxembourg, 31/01/2023

Doc ESTAT/F5/ESAW/2023/01

Guidelines for ESAW data transmission 2021

TABLE OF CONTENTS

1.	Introduction	3
2.	Format and structure of data file	3
3.	Format of variables.....	3
4.	Weights.....	4
5.	Classifications	6
6.	Fatal road traffic accidents and fatal accidents on board of any other means of transport	6
7.	RTA variable	7
8.	Transmission of data files	7
9.	Deadline for data transmission.....	7
10.	Reference populations	8

1. Introduction

The transmission of data on accidents at work has to follow a common procedure and to include data structured in a common way. This request implies the respect of guidelines with regard to the following aspects:

- Format and structure of the data file
- Format of the variables
- Weights
- Classifications
- Road traffic accidents
- Transmission of data files
- Deadline for data transmission
- Reference populations

2. Format and structure of data file

Data have to be transmitted in the form of a **.csv file** using Eurostat's official tool for data transmission, see section 8. of this document. This file should have the following structure:

1. The header (first row) has to include all variables with the correct label and in the correct sequence:

CASE;NACE;ISCO;AGE;SEX;INJ;BODY;GEO;DATE;TIME;SIZE;NATI;EMPST;SEV;WST;WEN;WPR;SPA;MA_SPA;DEV;MA_DEV;CMI;MA_CMI;RTA;WGT;WGT_P3

2. The variables/codes have to be semicolon (;) delimited.
3. All variables of all cases have to be reported, no blank is accepted for any data cell (there has to be at least a code for 'unknown').
4. No special characters, e.g. '-' or '##', are accepted.
5. Data coming from different sources have to be summarised in one single file per country.
6. The case number has to be unique; the first 4 digits have to correspond to the year of reference, e.g. 20210000001, 20210000002, etc.
7. Unknown data have to be reported using a specific code for each variable (see section 3).

3. Format of variables

There are **26 variables** in ESAW data files and each must have a specified string length and data type (alphanumeric or numeric). In addition, the code for 'unknown' data is defined for each variable (see table below).

Table 1: ESAW variables and codes:

Position	Variable name	Length	Data type	Code for 'unknown'
1	CASE	11	alphanumeric	-
2	NACE	4	alphanumeric	UNK
3	ISCO	2	alphanumeric	UN
4	AGE	2	alphanumeric	99
5	SEX	1	alphanumeric	9
6	INJ	3	alphanumeric	000
7	BODY	2	alphanumeric	00
8	GEO	5	alphanumeric	UNK
9	DATE	8	alphanumeric (YYYYMMDD)	00000000
10	TIME	2	alphanumeric	99
11	SIZE	1	alphanumeric	9
12	NATI	1	alphanumeric	0
13	EMPST	3	alphanumeric	000
14	SEV	3	alphanumeric	000
15	WST	1	alphanumeric	0
16	WEN	3	alphanumeric	000
17	WPR	2	alphanumeric	00
18	SPA	2	alphanumeric	00
19	MA_SPA	4	alphanumeric	0000
20	DEV	2	alphanumeric	00
21	MA_DEV	4	alphanumeric	0000
22	CMI	2	alphanumeric	00
23	MA_CMI	4	alphanumeric	0000
24	RTA	1	alphanumeric	U
25	WGT	-	numeric	-
26	WGT_P3	-	numeric	-

Alphanumeric means a text string (not a number), even if the text string contains only digit symbols, which is often the case for ESAW data. An example of a text string of digits is the 4-digit NACE Rev. 2 code '2711' (Manufacture of electric motors, generators and transformers). For the text strings, among the others, leading zeros (0) are important and should not be left out. At the same time there should be no decimal points, thousands separators etc.

Numeric means integer or floating point data in the form of a decimal number. In ESAW, this data type is used only for the weight variables such as WGT and WGT_P3. They have no leading zeros except before the decimal point such as in '0.6'. A typical value is '1.0'.

Phase III variables “Causes and Circumstances” (grey shaded on table): three out of nine variables on causes and circumstances must be provided.

4. Weights

Traditionally, the input files include two columns for weights: WGT and WGT_P3. When the data transmitted are ‘pure’-administrative data, weight variables equal to 1.0. If the columns containing the two variables WGT and WGT_P3 are same, the column for WGT_P3 could be omitted.

Weights have always to be reported using a numeric type, e.g. WGT = 1.0, WGT_P3=1.0.

1. WGT (weight) could be used to correct the data for under-reporting and/or in case that the data file includes only a sample of all accidents at work in a country, e.g. if a survey is used. This weight will be used to calculate the number of accidents as well as their

breakdowns by sex, age, etc. for **all variables** related to the **main characteristics of the accidents** (phase I and II variables):

- For **Non-fatal** accidents, WGT is always of numeric type. If **NO** weight is specified for a non-fatal accident, it is assumed to be "1.0".

Example of reporting a non-fatal accident:
CASE;NACE;ISCO;AGE;SEX;INJ;BODY; (...);WGT
20090000001;6820;91;45;1;032;61; (...);2.1

WGT of numeric type

- For **Fatal** accidents - WGT equals always "1.0"

Example of reporting a **fatal accident**:
CASE;NACE;ISCO;AGE;SEX;INJ;BODY;...;SEV; (...);WGT
20090000001;6820;91;45;1;032;61;...;998;...; (...);1.0

WGT should be equal to '1.0'
for fatal accidents.

2. WGT_P3 (weight) could be used to correct the data for under-reporting and/or in case that the data file includes only a sample of all accidents at work in a country, e.g. if a survey is used in **exceptional situations when administrative data are not available**. This weight will be used to calculate the number of accidents as well as their breakdowns by sex, age, etc. for **all variables** related to the **causes and circumstances of the accidents** (phase III variables). Similar to WGT the variable WGT_P3 should be of numeric type, and WGT_P3 for fatal accidents should be equal to "1.0". According to ESAW regulation three out of nine variables on causes and circumstances must be available, i.e. **for each row** in the data file it is expected to have values different from 'Unknown' codes for at least three out of nine phase III variables.
3. WGT and WGT_P3 should not be equal to "0.0", if this is the case additional information should be provided on the reasoning.
4. It is expected that the total number of accidents calculated with WGT and WGT_P3 would be consistent.

5. Classifications

The codes used for the variables have to follow the classifications as described in the ESAW summary methodology (latest version), including any new methodological guidance and any revisions of the existing classifications as requested. The following statistical classifications are requested NACE Rev.2, NUTS 2021, ISCO-08, phase III variables codes of the ESAW summary methodology etc.

6. Fatal road traffic accidents and fatal accidents on board of any other means of transport

‘Accident at work’ is defined in ESAW methodology as a discrete occurrence in the course of work which leads to physical or mental harm. The phrase ‘in the course of work’ means ‘*while engaged in an occupational activity or during the time spent at work*’.

A **fatal accident** is defined as an accident that leads to the death of a victim within one year of the accident.

According to the ESAW Regulation 349/2011 of 11 April 2011, for these specific cases, **it is mandatory** to report to Eurostat (see annex III – metadata of the Regulation) the following numbers:

- The total number of **fatal** road traffic accidents for persons employed outside the NACE Rev.2 sector H "Transportation
- The total number of **fatal** accidents on board of any other means of transport during a journey in the course of work for persons employed outside the NACE Rev.2 sector H "Transportation".

These **two numbers** must be filled-in separately in the ESS Metadata handler (<https://webgate.ec.europa.eu/estat/spe/metaconv/home.htm>) as requested by the existing template.

Example:

COVERAGE OF ACCIDENTS OUTSIDE PREMISES (Annual update expected)

Accidents in the course of work	
1. Commuting accidents	N ¹
2. Accidents in a public place or in a mean of transport during a journey in the course of work	Y
2.1 Road traffic accidents in the course of work (public highways, car parks, internal ways inside the premises of the enterprise)	Y
2.2 Number of fatal road traffic accidents during a journey in the course of work for persons employed outside the NACE Rev. 2 sector H Transportation	53
2.3 Other accidents (slips, falls, aggressions, etc.) in a public place (pavement, staircases, etc.) or in the arrival and starting points (station, port, airport, etc.) of any mean of transport, during a journey in the course of work	Y
2.4 Accidents on board of any means of transport (underground railway, tram, train, boat, plane, etc.) used during a journey in the course of work	Y
2.5 Number of fatal accidents on board of any other means of transport during a journey in the course of work for persons employed outside the NACE Rev. 2 sector H Transportation	14
3. Accidents occurred within the premises of another company than that which employs the victim, or in a private individual, in the course of work	Y
4. Accidents having only a medical origin, in the course of work	N ¹

7. RTA variable

In the data file (csv), it is optional to submit information on whether or not an accident at work happened during a journey in the course of the work or not. For this purpose, the variable RTA has been created, which is, as mentioned, above optional. However, when available, it is **strongly encouraged** to submit this information in the data file.

The variable RTA is alphanumeric, string length is 1. The codes available for this variable:

- 'Y': accident happening during a journey in the course of the work, on the road or on board of any other means of transport
- 'N': accident not happening during a journey in the course of the work
- 'U': unknown if the accident happening was during a journey in the course of the work or not

NB: Commuting accidents (accidents which occur during the normal journey to/or from home and the place of work) **should not be reported** in the data file. They are excluded from ESAW accidents at work following the definition of an accident at work in the ESAW Summary methodology.

8. Transmission of data files

Data have to be transmitted to the Single Entry Point of Eurostat (via 'EDAMIS') through the following link https://webgate.ec.europa.eu/edamis/login/index.cfm?TargetUrl=disp_login.

The files to be transmitted are: data file (.csv format) + filled-in ref. population template (.xls format) via eDAMIS. The metadata file via ESS – MH tool.

Transmission of the three files should be done in the same time.

9. Deadline for data transmission

The deadline for data transmission is 18 months after the end of the reference year (30 June of year N+2, where N is the reference year).

By the legal deadline the data transmission has to include:

- the main data input file (.csv) for the reference year N
- the filled-in reference population template for the reference year N
- the metadata information should be updated in the ESS Metadata handler tool for the reference year N, including the two mandatory numbers on fatal RTA.

10. Reference populations

Reference populations are mentioned in the ESAW implementing Regulation and in the ESAW summary methodology (section 7.1). They are the denominator for the calculations of incidence rates and standardised incidence rates (the nominator is the number of accidents). This means that they are as important for ESAW incidence and standardised incidence rates as the number of accidents. They are broken down by sex, economic activity (NACE 1-digit sections and 2-digit divisions) and employment status (employees, self-employed, family workers and others). In order to achieve a high level of comparability of ESAW incidence rates, it is important that reference populations have a similar level of quality as the number of accidents.

Countries may either (1) provide their own specific national reference population taken from relevant national sources, or (2) let Eurostat use the EU-LFS based reference populations. Some countries which provide their own specific reference populations deliver only a part of the required data, for example only data on employees, only for men, only from one insurance system or only for a subset of NACE sections/divisions. Other countries provide specific national reference populations if there is a significant number of workers commuting on a daily basis to the country.

If a country sends its own specific reference population to Eurostat but it cannot cover all NACE sections (1-digit) and divisions (2-digit) or employment types, which are nevertheless covered by accident data, Eurostat may use LFS reference populations for the missing NACE sections/divisions or employment types.

For example, if a country does not provide data on self-employed persons, the LFS data on self-employed may be used (always provided that accident data covers self-employed and that the LFS is comparable with the rest of the specific national reference population). In such a case, the calculation of incidence rates is based on two different data sources: (i) country specific and (ii) from LFS. This mixing may not be ideal but in some justified and comparable cases, it may not be possible to avoid it. **Eurostat encourages all countries to provide the reference population that is appropriate for the statistical population covered by the accidents.**

Excel template for reference populations

According to the reference population template, the information requested is the number of **persons employed**. In the context of the EU- LFS, an employed person is a person aged 15 and over who during the reference week performed work - even if just for one hour a week - for pay, profit or family gain. Since the ESAW data might refer to students and other employment categories, more complex measures of employment like; measuring the number of hours worked or by converting, the number of hours into full-time equivalent units (FTEs) are not accepted, as they will be conflicting with the ESAW methodology.

Reference populations will be validated in a similar way as the data file. If specific national reference populations are sent to Eurostat, they should be accompanied with relevant comments to explain their origin and the difference with LFS reference populations (see “Comments” field of the reference population template). In addition, the reference population template should be consistent with the filled-in information in the national metadata section “3. Statistical presentation”.

In general, all cells in reference population templates should be filled. Into all cells which relate to (parts of) NACE sections/divisions or types of employment for which no accident data is collected (i.e. non-coverage), zeros ("0") **should be filled, in order to avoid that Eurostat uses LFS** reference populations in such cases, which would result in too high reference populations and too low incidence rates.

On the other hand, if national ESAW data covers certain NACE sections/divisions and employment types but reference populations are not available at national level and LFS reference populations should be used, **the relevant cells should be filled with "LFS"**.

Thus, the following coding has to be used in the template for reference populations (as mentioned on top of the template):

- No data cell in the reference population template should be left blank; it should either contain a number, "0" or "LFS".
- If some NACE sections/divisions or types of employment are not covered by ESAW accident data, the relevant cells of the template should be filled with zeros ("0") (and not left blank).
- The cells covered by ESAW accident data should either be filled with specific national reference populations or with "LFS" to indicate that Eurostat should use LFS reference populations.

Important points to consider:

- *The reference population template is requested from all ESAW participants, including the countries using 'pure' LFS as reference population.*
- *Don't leave cells empty.*
- *Please provide the reference population template filled-in with together with microdata and metadata.*
- *The LFS data are an alternative when the Member State cannot provide its own figures*

In general, Eurostat prefers not mixing two data sources, **ideally is to have the appropriate national reference population provided with the data.** The use of LFS reference populations is done ONLY if justified from the statistical point of view (i.e. if it reflects reality best) or the unavailability of the ESAW reference population from national source.

However, there are notable exceptions to the use of LFS reference populations in ESAW:

1. There exists a complete register of workers in a country or a complete database of insured workers which match the reference population of the notification system of accidents at work;
2. The LFS reference population is not reliable, in particular not for smaller NACE divisions or employment types and better data is available from a national high quality data source; this is sometimes the case for smaller countries and smaller LFS samples. However, in this case, it should be reproducible why the specific national reference population has higher quality and from which sources it is exactly derived;
3. An insurance company or another significant part of the working population (of a certain NACE section/division or employment type) is not included in the accident notification system and the LFS reference population does not have a breakdown for it;
4. The LFS reference population is a resident concept, i.e. it can count only workers living in a country. All workers who live abroad and commute into a country will not be included in the LFS reference population. If the number of cross-border workers is significant, the LFS

reference population would underestimate the number of workers (and the incidence rates would be overestimated).

On the other hand, the EU-LFS reference population may, in some countries, be the best data source and has the important advantage for EU data that it is comparable between Member States. In the following cases LFS reference population may be of better quality:

1. The relevant LFS is based on a significantly large, high quality sample, also concerning the smaller NACE divisions and other employment types than employees; all NACE sectors and employment types are covered;
2. Same as 1. above: even if certain NACE sectors or employment types (such as self-employed) are not included in the national accident notification system, Eurostat prefers that Member States indicate which specific NACE section (1-digit), division (2-digits), group (3-digit) or class (4-digit) or employment type is excluded or included (by adding "0" or "LFS" in the relevant parts of the Excel template). Eurostat will then use the LFS reference populations for all covered sectors and employment types (LFSs) and will not use the others (0s), however this should rather represent an exception than the rule.

If a Member State, sends only isolated cases of accidents for some NACE sections/divisions or employment types with a significant number of workers, Eurostat will contact countries bilaterally to clarify which reference population should be used for calculating incidence rates.