2.5 Data Validation Routines

Validation routines are checks on data consistency designed to ensure data quality. These checks validate *(i)* that values fall within a usual range (e.g. weight in relation to the gestational age) or *(ii)* the logic links between variables.

2.5.1 DMS Validation Routine

A validation routine is implemented in the EUROCAT Data Management Software (DMS) and is automatically executed upon saving a data entry, whether done manually or via bulk import.

The validation checks result in three categories of messages:

- ***blocking errors***: do not allow the concerned cases to be saved.
- ***logical errors***: are inconsistencies between values in the case. The concerned cases can be saved even if the values aren't corrected. Logical errors are presented separately for core and non-core variables
- ***warnings***: draw attention to unusual values or minor inconsistencies to improve data quality. The concerned cases can be saved. Warnings are presented separately for core and non-core variables.

Table 1 below summarises the checks that are performed.

2.5.2 Registry Validation Routines

All babies must have a local identification number, date of birth, sex, number of babies delivered, number of malformed cases in multiple set, type of birth, civil registration status, birthweight, length of gestation, survival status, age of mother, when malformation was discovered, first surgery and at least one malformation or syndrome code. Before sending data to the Central Registry any cases with this "core" information lacking should be reviewed to find out if it is possible to complete the missing data.

Registries should make sure of the following:

- The local identity number (variable NUMLOC) within a registry is unique for each case.
- There are no duplicate cases. DMS identifies potential duplicates in the following way:
    - Cases with the same values for 4 key variables (date of birth, sex, birthweight +/- 100 g, maternal age) should be checked as possible duplicates. If value= unknown for any of the 4 matching variables, then the case is not matched on that specific variable.
    - If case is a twin or higher order multiple birth (NBRBABY=2, 3, 4, 5, 6 or 7), then no duplicate check should be carried out.
    - It is not sufficient to rely on matching the name of the baby or mother for finding duplicates
- Values of EUROCAT variables should fall within the accepted range (see Chapter 2.2.2). This is checked in the DMS when data entries are saved.

Table 1. List of validation checks performed by the EUROCAT Data Management Software

VARIABLE	CORE / NON-CORE VARIABLES	CHECK	DISPLAYED MESSAGE	TYPE
NUMLOC	CORE	Is empty	Local identification number cannot be missing.	BLOCKING ERROR
BIRTH_DATE	CORE	Is empty	Infant date of birth cannot be missing. Please provide at least the year of birth.	BLOCKING ERROR
SEX	CORE	Is empty	Infant sex is missing. Please check and enter 9 if unknown.	WARNING
NBRMALF	CORE	Is empty while NBRBABY>1 and NBRBABY<8	Number of malformed is missing. Please check and enter 9 if unknown.	WARNING
NBRMALF	CORE	Is higher than NBRBABY	Number of malformed in the pregnancy is higher than number of babies in the pregnancy. Please correct, potentially looking at siblings' numlocs in variable SIB1, 2, or 3.	LOGICAL ERROR
FIRSTPRE	NON-CORE	FIRSTPRE=8 or 10 while WHENDISC=6	The case is recorded as diagnosed prenatally (WHENDISC=6) while no positive first prenatal test (FIRSTPRE) is reported. Please check.	LOGICAL ERROR
SURVIVAL	CORE	Is empty	Survival beyond one week is missing. Use 9 if unknown.	WARNING
SURVIVAL	CORE	SURVIVAL=1 or 3 while TYPE=2, 3 or 4	Survival beyond one week of age must be 'No' for stillbirths, spontaneous abortions and TOPFA.	LOGICAL ERROR

VARIABLE	CORE / NON-CORE VARIABLES	CHECK	DISPLAYED MESSAGE	TYPE
DEATH_DATE	NON-CORE	DEATH_DATE is empty while TYPE=1 and PM=1, 2 or 4	Date of death expected for livebirth with PM.	WARNING
SIB1/2/3	NON-CORE	SIB1/2/3 are all empty while PREVSIB=1 and SIBANOM=1, 2 or 3	Affected siblings' local ID is missing. Please check if it can be found.	WARNING
TYPE	CORE	TYPE>1 while SURVIVAL=1	The type of birth should be *Livebirth* when the case survived beyond 1 week (SURVIVAL = 1).	LOGICAL ERROR
TYPE	CORE	TYPE>1 while SURVIVAL=3	The type of birth should be *Livebirth* when the case was alive at discharge (SURVIVAL = 1).	LOGICAL ERROR
WEIGHT	CORE	Is empty	Birth weight is missing. Please use 9999 for not known.	WARNING
GESTLENGTH	CORE	Is empty	Length of gestation is missing. Please use 99 for not known.	WARNING
AGEMO	CORE	Is empty	Maternal age is missing. Please use 99 for not known.	WARNING
AGEMO	CORE	Is unknown while it can be calculated with DATEMO and BIRTH_DATE	Age of the mother is declared as unknown (99) but can be calculated (BIRTH_DATE - DATEMO).	LOGICAL ERROR

VARIABLE	CORE / NON-CORE VARIABLES	CHECK	DISPLAYED MESSAGE	TYPE
AGEMO	CORE	Age entered does not match the age calculated with DATEMO and BIRTH_DATE	Age of the mother entered doesn't match the calculated age. Please check.	LOGICAL ERROR
WEIGHT	CORE	Check WEIGHT in relation with GESTLENGTH where gestational age has not been confirmed (gaconfirm=0)	Maximum birthweights for gestational age are usually: 20-22 weeks 750 g 23-25 weeks 1000 g 26-27 weeks 1500 g 28-31 weeks 2000 g 32-35 weeks 3500 g 36-37 weeks 4000 g 38+ weeks 6000 g Message: Birthweight is above the usual max birthweight for gestational age *XX* to *YY* weeks (*WWW*g). Please check birthweight and gestational age.	WARNING
SP_SYNDROME	CORE	SP_SYNDROME is empty while SYNDROME is provided	Please enter syndrome's description in SP_SYNDROM.	WARNING
SYNDROME	CORE	Not an ICD10 code	ICD10 codes allowed only.	WARNING
SYNDROME	CORE	Not a valid ICD10 code	ICD syndrome code is too short – 4th digit required (if not Q02*)	WARNING
MALF01-8	CORE	Not an ICD10 code	ICD 10 codes allowed only.	WARNING

VARIABLE	CORE / NON-CORE VARIABLES	CHECK	DISPLAYED MESSAGE	TYPE
MALFO1-8	CORE	Not a valid ICD10 code	ICD malfo code is too short – 4th digit required (if not Q02*).	WARNING
SYNDROME MALFO1-8	CORE	Consistency of coded malformations Check if Q03* and Q05* are both coded	Remove Q03* when Q05* is already coded.	WARNING
SYNDROME MALFO1-8	CORE	Consistency of coded malformations Check if Q35* and (Q36* or Q37*) are both coded	Remove Q35* when Q36* or Q37* are already coded.	WARNING
SYNDROME MALFO1-8	CORE	Consistency of coded malformations Check if Single ventricle (Q204*, 7453) and hypoplastic heart (Q234*, Q226*) are both coded	Single ventricle excludes hypoplastic heart.	WARNING
GESTLENGTH	CORE	Check extreme GESTLENGTH where gestational age has not been confirmed (gaconfirm=0)	Gestational age is outside the range 10 to 43 included. Please check.	WARNING
WEIGHT	CORE	Check extreme WEIGHT where weight has not been confirmed (bwconfirm=0)	Birth weight is greater than 6000g or birth weight less than 400g. Please check. (for live births and still births only)	WARNING

VARIABLE	CORE / NON-CORE VARIABLES	CHECK	DISPLAYED MESSAGE	TYPE
AGEMO	CORE	Check extreme AGEMO where age of the mother has not been confirmed (dobmconfirm=0)	Mother age is outside the range 15 to 50. Please check.	WARNING
AGEDISC	CORE	AGEDISC is empty while WHENDISC=6	Gestational age at discovery is expected when the anomaly has been diagnosed prenatally. Code 99 if unknown.	WARNING
TYPE	CORE	TYPE different from 3 or 4 while WHENDISC=7	Type of birth is expected to be a spontaneous abortion (TYPE=3) or TOPFA (TYPE=4) when the anomaly has been discovered at abortion (WHENDISC = 7).	LOGICAL ERROR
WHENDISC	CORE	WHENDISC different from 6 (prenatal detection in live fetus) and 9 (not known) while TYPE=4	For TYPE=4 (TOPFA), 'When discovered' should usually be 'Prenatal diagnosis'.	LOGICAL ERROR
GESTLENGTH AGEDISC	CORE	Consistency between GESTLENGTH and AGEDISC for TOPFA	The gestational age when the anomaly was discovered should be earlier than the gestational age at time of TOPFA.	LOGICAL ERROR
DEATH_DATE	NON-CORE	DEATH_DATE is null, SURVIVAL is 2, TYPE=1	Missing date of death.	WARNING
DEATH_DATE	NON-CORE	Consistency between DEATH_DATE=222222, and SURVIVAL is 2 for TYPE=1	Inconsistency between the variable SURVIVAL (declared as 2="No") and the variable DEATH_DATE (declared as "Known to be alive at 1 year") for a live birth. Please check.	LOGICAL ERROR
DEATH_DATE	NON-CORE	Consistency between DEATH_DATE more than 7 days after BIRTH_DATE and SURVIVAL=2 for TYPE=1	Inconsistency between the variable SURVIVAL (declared as 2="No") and the variable DEATH_DATE (more than 7 days after the birth) for a live birth. Please check.	LOGICAL ERROR

2.5.3 Registry Data Quality Evaluation

Before sending data to the JRC-EUROCAT Central Registry, local registries should produce some frequency tables to evaluate the quality of the information.

- A high frequency of unknown values for any variable should prompt an investigation of how the recording of the variable can be improved, and the registry should communicate with Central Registry concerning how the variable can be used in analyses of data, especially if there is selection bias in the distribution of known values.
- It may be useful to check that all malformation codes which have been used only once are valid.
- A high frequency of poorly specified malformation codes should prompt investigation.
- The number of cases where “total previous pregnancies” has been coded “0” should correspond approximately to the number of cases expected from the proportion of primiparous mothers in the population.
- Cross-tabulation of maternal age and number of previous pregnancies should show a distribution roughly corresponding to the distribution in the total birth population.

Last update: 16.04.2025