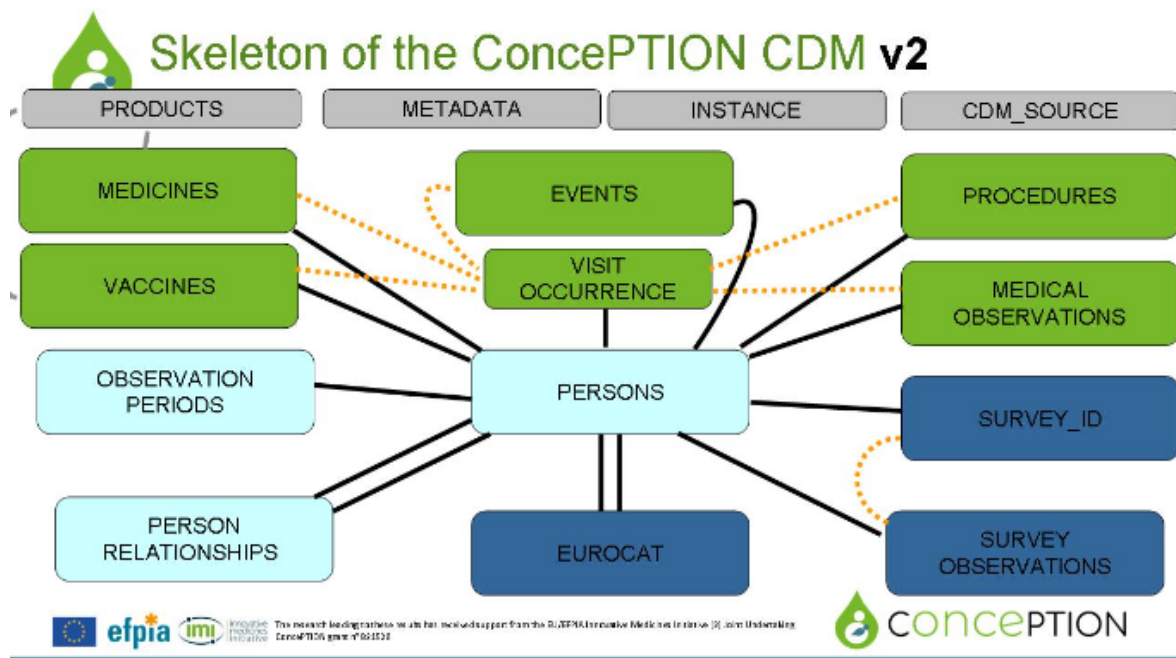


PRACTICAL 1

Martijn Koster, 6234119

Jurrian van de Kraats, 5961688

Tim Poorthuis, 0651478



1. Which reports are produced from running Level 1 and Level 2 scripts? Are all the reports present? Explain why yes or no.

The reports produced for running Level 1 checks:

- CDM Source and INSTANCE
- EVENTS
- MEDICINES
- METADATA
- OBSERVATION PERIODS
- PERSONS
- STEP 1 TO 3
- SURVEY OBSERVATIONS
- VACCINES

Reports that are not produced after running Level 1 checks:

- MEDICAL OBSERVATIONS
- PRODUCTS
- VISIT_OCCURENCE
- SURVEY_ID
- PERSON_RELATIONSHIPS

These reports are missing, because there are no csv files provided for these tables.

Reports produced after lvl 2 checks:

- Event dates before birth date
- Event dates after date of death
- Event dates outside observation periods
- Subjects observed in a table of interest without a corresponding record in the PERSONS table

Reports that are not produced after level 2 checks:

- Observations associated with a visit_occurrence_id which occur after the visit_end_date
- Observations associated with a visitr_occurrence_id which occur after the visit_end_date
- Observations associated with a visit_occurrence_id for which the associated person_id differs from that in the visit_occurrence table
- Subjects indicated in PERSON_RELATIONSHIPS as the parent of a child with birth_date less than 12 years prior to the recorded birth_date of the associated child

Visit occurrence is not present in the level 1 checks, hence they will not be available in the level 2 checks. The same goes for “person-relationships”.

2. Review reports METADATA, STEP 1 to 3, EVENTS from Level 1. Identify issues and explain the potential reasons.

METADATA:

The consistency check shown in the figure below. This shows that the following tables are present in the metadata, but not in the directory;

(INSTANCE, MEDICAL_OBSERVATIONS, PERSON_RELATIONSHIPS, PROCEDURES, PRODUCTS, VISIT_OCCURENCE).

2. Consistency check				
Comparison between METADATA and directory				
Check if there are any inconsistencies between the information in the METADATA table and the working directory. If a mandatory table is missing either in the directory or the METADATA table it will be colored red . You need to first provide that table or fix the information and then proceed with the next steps.				
Mandatory tables:				
a. METADATA				
b. CDM_SOURCE				
c. EVENTS (not if EUROCAT table is present)				
d. MEDICINES (not if EUROCAT table is present)				
e. PROCEDURES (not if EUROCAT table is present)				
f. VACCINES (not if EUROCAT table is present)				
g. MEDICAL_OBSERVATIONS (not if EUROCAT table is present)				
h. PERSONS				
i. OBSERVATION_PERIODS (not if EUROCAT table is present)				
j. PERSON_RELATIONSHIPS (not if EUROCAT table is present)				
k. PRODUCTS (not if EUROCAT table is present)				
Show <input type="text" value="10"/> entries		Search: <input type="text"/>		
	table_name	presence_directory	presence_metadata	comment
1	INSTANCE	false	true	The table is present in the METADATA table but not in the directory.
2	MEDICAL_OBSERVATIONS	false	true	The table is present in the METADATA table but not in the directory.
3	PERSON_RELATIONSHIPS	false	true	The table is present in the METADATA table but not in the directory.
4	PROCEDURES	false	true	The table is present in the METADATA table but not in the directory.
5	PRODUCTS	false	true	The table is present in the METADATA table but not in the directory.
6	VISIT_OCCURENCE	false	true	The table is present in the METADATA table but not in the directory.

STEP 1 TO 3:

Values check:

- An error occurred in the EVENTS_SDO.csv table and a warning in the EVENTS_SALM table.
- EVENTS_SDO table has 0, 'si', and 'no' values, where it should be 'yes' or 'no' in the 'present_on_admission' column
- EVENTS_SALM table has only 'yes' values on the present_on_admission column, this is OK

Explanations:

a. If `table_directory` and `values_tables` are empty but `values_metadata` is completed means that the table is not present in the working directory so there should be no data in the METADATA table. Fix the ETL if the variable should be present or remove this variable from the `list_of_values` in the METADATA table if it should be empty (red).

b. If `values_table` is completed but `values_metadata` is empty please fix the METADATA table accordingly (orange).

c. If both `values_table` and `values_metadata` are completed means that the information doesn't match. Correct the table or the METADATA accordingly (blue).

d. If `values_metadata` contain extra values compared with `values_table` will set the `error` column to "warning". Double check if the information in the table or METADATA is correct and fix the error if present (purple).

Show 10 entries

Search:

	table_name	table_directory	variable_name	values_table
21	MEDICINES	MEDICINES_2015_SPF.csv	meaning_of_drug_record	dispensing_in_commu
22	MEDICINES		origin_of_drug_record	
23	OBSERVATION_PERIODS	OBSERVATION_PERIODS.csv	op_meaning	region3_PC region7_PC region14_PC region15_ region2_PC region6_PC outpatient_pharmacy observed_in_hospital r
24	PERSONS	PERSONS.csv	quality	reliable
25	PERSONS	PERSONS.csv	sex_at_instance_creation	F M O U
26	PROCEDURES		meaning_of_procedure	
27	PROCEDURES		origin_of_procedure	
28	PROCEDURES		procedure_code_vocabulary	
29	SURVEY_ID	SURVEY_ID_RMR.csv	survey_meaning	death_registry
30	SURVEY_OBSERVATIONS	SURVEY_OBSERVATIONS_RMR.csv	so_meaning	death_registry

Showing 21 to 30 of 39 entries

Previous

1

2

3

4

Next

Tables	Potential reason	How to fix
CDM_SOURCE	Value_table is completed, but values_metadata is empty	Fix METADATA table accordingly
EVENTS	Both values_table and values_metadata information doesn't match	Correct the table, METADATA table accordingly
EVENTS	values_metadata contain extra values compared with values_table (WARNING)	double check if information in the table or METADATA is correct and fix the error if present
EVENTS	If table_directory and values_tables are empty but values_metadata is completed means that the table is not present in the working directory so there should be no data in the METADATA table.	Fix the ETL if the variable should be present or remove this variable from the list_of_values in the METADATA table if it should be empty
EVENTS	Value_table is completed, but values_metadata is empty	Fix METADATA table accordingly
MEDICAL_OBSERVATION	If table_directory and values_tables are empty but values_metadata is completed means that the table is not present in the working directory so there should be no data in the METADATA table.	Fix the ETL if the variable should be present or remove this variable from the list_of_values in the METADATA table if it should be empty
MEDICINES	values_metadata contain extra values compared with values_table (WARNING)	double check if information in the table or METADATA is correct and fix the error if present
MEDICINES	If table_directory and values_tables are empty but values_metadata is completed means that the table is not present in the working directory so there should be no data in the METADATA table.	Fix the ETL if the variable should be present or remove this variable from the list_of_values in the METADATA table if it should be empty
OBSERVATION_PERIODS	Value_table is completed, but values_metadata is empty	Fix METADATA table accordingly

PERSON	Value_table is completed, but values_metadata is empty	Fix METADATA table accordingly
PROCEDURES	If table_directory and values_tables are empty but values_metadata is completed means that the table is not present in the working directory so there should be no data in the METADATA table.	Fix the ETL if the variable should be present or remove this variable from the list_of_values in the METADATA table if it should be empty
SURVEY_ID	Value_table is completed, but values_metadata is empty	Fix METADATA table accordingly
SURVEY_OBSERVATION	Value_table is completed, but values_metadata is empty	Fix METADATA table accordingly
VACCINES	Value_table is completed, but values_metadata is empty	Fix METADATA table accordingly
VISIT_OCCURRENCE	If table_directory and values_tables are empty but values_metadata is completed means that the table is not present in the working directory so there should be no data in the METADATA table.	Fix the ETL if the variable should be present or remove this variable from the list_of_values in the METADATA table if it should be empty

Dates check:

EVENTS_SALM.csv

- Variable name = start_date_record, 21 variables where the years are in the future
- Does not give an error!

OBSERVATION_PERIODS.csv

- Variable name = op_start_date, 762 errors in the year, both before 1995 as in the future
- The 12 future dates do not give an error

OBSERVATION_PERIODS.csv

- Variable name = op_end_date, 1142 errors in year, of which 939 are in the future
- These are okay.

Dates check PERSONS table:

These have 715 errors in the month and 292 errors in the day. *These are probably producing errors, because 1 is used in months instead of 01 ,etc.*

They are expecting 2 characters instead of 1.

Missing Data:

Table_name	Variable_name	Calender_year	Percentage missing	Meaning	Potential reasons/solution
MEDICINES	<i>date_prescription</i>	2010-2016	100%	<i>dispensing_in_community_pharmacy</i>	<i>Data was not provided. Consult with people at concePTION to check what went wrong in transferring the data.</i>
	<i>disp_number_medicinal_product</i>	2010-2016	100%	<i>dispensing_in_community_pharmacy</i>	<i>Data was not provided. Consult with people at concePTION to check what went wrong in transferring the data.</i>
	<i>indication_code</i>	2010-2016	100%	<i>dispensing_in_community_pharmacy</i>	<i>Data was not provided. Consult with people at concePTION to check what went wrong in transferring the data.</i>
	<i>indication_code_vocabulary</i>	2010-2016	100%	<i>dispensing_in_community_pharmacy</i>	<i>Data was not provided. Consult with people at concePTION to check what went wrong in transferring the data.</i>
	<i>presc_duration_days</i>	2010-2016	100%	<i>dispensing_in_community_pharmacy</i>	<i>Data was not provided. Consult with people at concePTION to check what went wrong in transferring the data.</i>
	<i>presc_quantity_per_day</i>	2010-2016	100%	<i>dispensing_in_community_pharmacy</i>	<i>Data was not provided. Consult with people at concePTION to check what went wrong in transferring the data.</i>
	<i>presc_quantity_unit</i>	2010-2016	100%	<i>dispensing_in_community_pharmacy</i>	<i>Data was not provided. Consult with people at concePTION to check what went wrong in transferring the data.</i>
	<i>prescriber_speciality</i>	2010-2016	100%	<i>dispensing_in_community_pharmacy</i>	<i>Data was not provided. Consult with people at concePTION to check what went wrong in</i>

					<i>transferring the data.</i>
	<i>prescriber_speciality_vocabulary</i>	2010-2016	100%	<i>dispensing_in_community_pharmacy</i>	<i>Data was not provided. Consult with people at concePTION to check what went wrong in transferring the data.</i>
	<i>product_lot_number</i>	2010-2016	100%	<i>dispensing_in_community_pharmacy</i>	<i>Data was not provided. Consult with people at concePTION to check what went wrong in transferring the data.</i>
EVENTS	<i>event_free_text</i>	2004, 2008-2013, 2015-2016, 2018 (mental health) 1996-2020 (hospitaliza tion)	100%	<i>hospitalisation_primary</i> <i>hospitalisation_secondary</i> <i>access_to_mental_health_s ervice_primary</i>	<i>Data was not provided. Consult with people at concePTION to check what went wrong in transferring the data.</i>
	<i>text_linked_to_event_code</i>	2004, 2008-2013, 2015-2016, 2018 (mental health) 1996-2020 (hospitaliza tion)	100%	<i>hospitalisation_primary</i> <i>hospitalisation_secondary</i> <i>access_to_mental_health_s ervice_primary</i>	<i>Data was not provided. Consult with people at concePTION to check what went wrong in transferring the data.</i>
	<i>vislt_occurrence_id</i>	2004, 2008-2013, 2015-2016,	100%	<i>hospitalisation_primary</i> <i>hospitalisation_secondary</i>	<i>Data was not provided. Consult with people at concePTION to check what went wrong in transferring the data.</i>

		2018 (mental health) 1996-2020 (hospitaliza tion)		<i>access_to_mental_health_ service_primary</i>	
	<i>event_free_text</i>	2004, 2008-2013, 2015-2016, 2018 (mental health) 1996-2020 (hospitaliza tion)	100%	<i>hospitalisation_primary hospitalisation_secondary access_to_mental_health_ service_primary</i>	<i>Data was not provided. Consult with people at concePTION to check what went wrong in transferring the data.</i>
	<i>laterality_of_event</i>	2004, 2008-2013, 2015-2016, 2018 (mental health) 1996-2020 (hospitaliza tion)	100%	<i>hospitalisation_primary hospitalisation_secondary access_to_mental_health_ service_primary</i>	<i>Data was not provided. Consult with people at concePTION to check what went wrong in transferring the data.</i>
	<i>end_date_record</i>	2004, 2008-2013, 2015-2016, 2018 (mental	4.5%	<i>access_to_mental_health_ service_primary</i>	<i>Data was not provided. Consult with people at concePTION to check what went wrong in transferring the data.</i>

		health) 1996-2020 (hospitaliza tion)			
VACCINES	<i>visit_occurrence_id</i>	2012-2020	100%	<i>administration_of_vaccine_ unspecified</i>	<i>Data was not provided. Consult with people at concePTION to check what went wrong in transferring the data.</i>
	<i>vx_text</i>	2012-2020	100%	<i>administration_of_vaccine_ unspecified</i>	<i>Data was not provided. Consult with people at concePTION to check what went wrong in transferring the data.</i>
	<i>vx_type</i>	2012-2020	100%	<i>administration_of_vaccine_ unspecified</i>	<i>Data was not provided. Consult with people at concePTION to check what went wrong in transferring the data.</i>
SURVEY_ OBSERVET ATIONS	<i>so_unit</i>	2005-2015	77.8%	<i>death_registry</i>	<i>The potential reason for the missing data is probably the fact that people have not yet died and not registered in any form, and thus have no death_registry & so_unit</i>
	<i>so_source_value</i>	2005, 2010-2013	51.9%	<i>death_registry</i>	<i>The potential reason for the missing data is probably the fact that people have not yet died and not registered in any form, and thus have no death_registry & so_unit</i>
PERSONS	<i>day_of_death</i>	1950-2020	98.9%		<i>The potential reason for the missing data is probably the fact that people have not yet died, and thus have no day_of_death</i>
	<i>month_of_death</i>	1950-2020	98.9%		<i>The potential reason for the missing data is probably the fact that people have not yet died, and thus have no month_of_death</i>

	<i>year_of_death</i>	1950-2020	98.9%		<i>The potential reason for the missing data is probably the fact that people have not yet died, and thus have no year_of_death</i>
	<i>counrty_of_birth</i>	1950-2020	100%		<i>Data was not provided. Consult with people at concePTION to check what went wrong in transferring the data.</i>
	<i>race</i>	1950-2020	100%		<i>Data was not provided. Consult with people at concePTION to check what went wrong in transferring the data.</i>
CDM_SOURCE	<i>data_dictionary_link</i>		-		<i>Data was not provided. Consult with people at concePTION to check what went wrong in transferring the data.</i>
	<i>etl_link</i>		-		<i>Data was not provided. Consult with people at concePTION to check what went wrong in transferring the data.</i>

EVENTS from level 1:

- Duplicated rows:

There are 9 duplicated rows in the data. Take caution when interpreting counts.

- Counts stratified by meaning (other variables)

There are 6 rows with a zero value for both count and total. Those will not be displayed in the tables or graphs. This happens when a variable is completely missing for a particular meaning category

Counts stratified by meaning and year (2 or more categories)

If years before 1995 or years in the future are present, they will be colored red.

Show 10 entries

Search:

	meaning	vocabulary	year	count	total
y	access_to_mental_health_service_primary	ICD10	4713	<5	21
y	access_to_mental_health_service_primary	ICD9CM	4713	20	21
	access_to_mental_health_service_primary	access_to_mental_health_service_primary	4713	21	21
	access_to_mental_health_service_primary	SALM	4713	21	21
	access_to_mental_health_service_primary	yes	4713	21	21
y	access_to_mental_health_service_primary	ICD10	2020	<5	37
y	access_to_mental_health_service_primary	ICD9CM	2020	<5	37
	access_to_mental_health_service_primary	access_to_mental_health_service_primary	2020	<5	37
	access_to_mental_health_service_primary	SALM	2020	<5	37
	access_to_mental_health_service_primary	yes	2020	<5	37

Showing 1 to 10 of 294 entries

Previous12345...30Next

- The table above shows that there are 5 observations in the future (year = 4713), which is impossible.

- Counts stratified by meaning and year (other variables)

There are 136 rows with a zero value for both count and total. This happens when a variable is completely missing for a particular meaning category and year.

- Counts stratified by meaning and year

There is 1 row with a zero value for both count and total. Those will not be displayed in the tables or graphs. This happens when a variable is completely missing for a particular meaning category and year.

3. If the METADATA table is not correct, fix it and save it as a csv file.

The tables are not present in the working directory. That is why there are errors in the metadata. Hence, the metadata table will be changed by putting the missing tables on 'no'. This is a short-term solution, because some of the tables who are not in the directory are mandatory.

4. Review Report 2.1 from Level 2. Describe issues the scripts find in the data.

2.1 - Date values before birth					
The table below shows shows the number of date values taking place before the person's date of birth, by table, date variable, and meaning:					
Show 10 entries			Search:		
TABLE	v.DATES	Meaning	Issues	Total	Percentage
VACCINES	vx_admin_date	administration_of_vaccine_unspecified	36	1026	3.509
VACCINES	vx_record_date	administration_of_vaccine_unspecified	36	1026	3.509
EVENTS	start_date_record	hospitalisation_primary	55	1735	3.17
EVENTS	start_date_record	hospitalisation_secondary	22	1340	1.642
MEDICINES	date_dispensing	dispensing_in_community_pharmacy	132	17712	0.745
EVENTS	start_date_record	access_to_mental_health_service_primary	0	145	0
SURVEY_ID	survey_date	death_registry	0	6	0
SURVEY_OBSERVATIONS	so_date	death_registry	0	54	0
MEDICINES	date_prescription	dispensing_in_community_pharmacy	0	0	
Showing 1 to 9 of 9 entries			Previous	1	Next

There are no issues, because all percentages are below 5%

5. Give a potential reason when an observation happening before birth date is not an issue.

There are several potential possible reasons why an observation happening before a person's birth date might not be an issue:

- *Prenatal factors:* The observation relates to a factor or event that occurred during the prenatal period such as maternal exposure to a substance or a genetic trait inherited from the parents.
- *Family history:* The observation might relate to a family history or genealogy, such as the presence of a certain trait or disease in previous generations.
- *Genetic predisposition:* The observation might relate to a genetic trait or predisposition that the individual inherited from their parents, and which was present from conception or before birth.
- *Developmental Trajectories:* The observation might relate to a developmental trajectory or pattern that begins before birth and continues into later life,