



# APAC Community-wide ETL Project

Sprint 5 Review

ETL Team



# Sprint Overview

## To do

condition\_era

drug\_era

## Nearing completion

condition\_occurrence

observation

drug\_exposure

measurement

## In review

procedure\_occurrence

note

## Done

person

care\_site

provider

death

visit\_occurrence

visit\_detail

specimen

observation\_period

device\_exposure

\*\*\*Location table: Lack of source data for mapping



# Sprint Overview

- Populated the second batch of the [STCM](#) table and other clinical tables.
- Schema unified for development
- Vocab team to get back on the pending mappings
- Some starts for the 2nd draft of the concept IDs field.
- Clinical events tables may have concerns about the date-time range not falling within the criteria.
  - such as in Visit Occurrence, Condition Occurrence, or Observation.
- QA team [providing ongoing feedback](#) from DQD; ongoing fixes to improve data quality.



# Sprint Reflections

- We have made a decision regarding the Clinical tables:
  - **Current Structure:** Based on `session_id`, suffixed with `00` or `0n` for duplicates.
  - **Step 1:** Check if the clinical event falls within the 1-week threshold in the `visit_occurrence` table → reference the IDs.
  - **Step 2:** If Step 1 is **No** → create a new `visit_occurrence` record.
- Multiple target concept ids for drug exposure source concept codes – clarified combination of ingredients and translate into multiple records in `drug_exposure`
- Any other assumptions made during development may be added [here](#) while pending clarifications to move forward.
- Extend the project duration by one more sprint to resolve pending tasks.



**Thank you!**



# Quality Assurance Team





# Sprint Status

- Documented SQL errors and conformance failures of DQD results
- Failed checks related to \*\_concept\_id columns are known issues, related to the in-progress work
- Executed checks twice with and without filters
- Filtered DQD results pass 96%



# Unfiltered Results

DataQualityDashboard Version: 2.6.1

Results generated at 2024-10-15 09:14:09 in 4 mins

	Verification				Validation				Total			
	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass
Plausibility	505	<b>8</b>	513	98%	291	0	291	100%	796	<b>8</b>	804	99%
Conformance	870	<b>56</b>	926	94%	137	<b>8</b>	145	94%	1007	<b>64</b>	1071	94%
Completeness	437	<b>15</b>	452	97%	15	<b>2</b>	17	88%	452	<b>17</b>	469	96%
Total	1812	<b>79</b>	1891	96%	443	<b>10</b>	453	98%	2255	<b>89</b>	2344	<b>96%</b>

1160 out of 2255 passed checks are Not Applicable, due to empty tables or fields.

39 out of 89 failed checks are SQL errors.

Corrected pass percentage for NA and Errors: 96% (1095/1145).





# Filtered Results

## Tables Excluded

- COHORT, COHORT\_DEFINITION
- CONCEPT, VOCABULARY, CONCEPT\_ANCESTOR, CONCEPT\_RELATIONSHIP, CONCEPT\_CLASS, CONCEPT\_SYNONYM, RELATIONSHIP, DOMAIN
- MEASUREMENT, DRUG\_EXPOSURE, CONDITION\_ERA, DRUG\_ERA, PROCEDURE\_OCCURRENCE

## DATA QUALITY ASSESSMENT

### PASAR

DataQualityDashboard Version: 2.6.1  
Results generated at 2024-10-15 13:36:37 in 2 mins

	Verification				Validation				Total			
	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass
Plausibility	225	2	227	99%	207	0	207	100%	432	2	434	100%
Conformance	690	24	714	97%	111	0	111	100%	801	24	825	97%
Completeness	347	9	356	97%	11	0	11	100%	358	9	367	98%
<b>Total</b>	<b>1262</b>	<b>35</b>	<b>1297</b>	<b>97%</b>	<b>329</b>	<b>0</b>	<b>329</b>	<b>100%</b>	<b>1591</b>	<b>35</b>	<b>1626</b>	<b>98%</b>

770 out of 1591 passed checks are Not Applicable, due to empty tables or fields.

4 out of 35 failed checks are SQL errors.

Corrected pass percentage for NA and Errors: 96% (821/852).

Desc	Count
Sql Errors	4
Known errors related to _concept_id column	21
Others	10
<b>Total</b>	<b>35</b>



## Next steps

- Review comments provided by the teams on the existing failed checks
- To document and analyze remaining failed checks related to completeness and plausibility



**Thank you!**