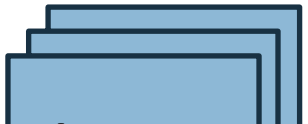
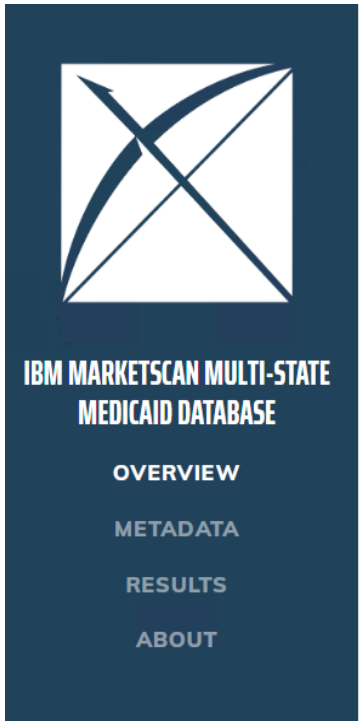
New DQD Features!

Updates since the 2019 OHDSI Symposium

Data Quality Dashboard





DATA QUALITY ASSESSMENT

IBM MARKETSCAN MULTI-STATE MEDICAID DATABASE

DataQualityDashboard Version: 1.0.0
Results generated at 2020-08-24 15:44:34 in 3 hours

	Verification				Validation				Total			
	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass	Pass	Fail	Total	% Pass
Plausibility	1849	6	1855	100%	281	6	287	98%	2130	12	2142	99%
Conformance	550	13	563	98%	80	0	80	100%	630	13	643	98%
Completeness	322	5	327	98%	12	0	12	100%	334	5	339	99%
Total	2721	24	2745	99%	373	6	379	98%	3094	30	3124	**99%**

https://github.com/OHDSI/DataQualityDashboard/

Exciting New Additions

- Metadata capture
- DQD for Cohorts
- Changes to the user interface
- Lots of helpful guides

Metadata Capture

SYNTHEA

DataQualityDashboard Version: 1.0.0
Results generated at 2020-09-09 09:37:37 in 14 mins

Column visibility | CSV

Show 5 entries

Search:

STATUS	TABLE	CATEGORY	SUBCATEGORY	LEVEL	NOTES	DESCRIPTION	% RECORDS
FAIL	MEASUREMENT	Completeness	None	FIELD	Exists	The number and percent of records with a value of 0 in the standard concept field unit_concept_id in the MEASUREMENT table. (Threshold=25%).	100.00%

Name:	standardConceptRecordCompleteness
Description:	The number and percent of records with a value of 0 in the standard concept field unit_concept_id in the MEASUREMENT table. (Threshold=25%).
Notes:	There are 20% of records in the CPRD MEASUREMENT table where the unit_source_value is NULL and the value_as_number is not NULL. Threshold moved to 25% to account for units that cannot be mapped. This issue of mapping will be addressed in wrike task https://www.wrike.com/open.htm?id=527764013.

DQD for Cohorts

Requested by contributors with data from multiple sites in one DB

This can also be applied to clinical cohorts

DQD for Cohorts

IBM CCAE

Column visibility | CSV

Show All entries

Search: 3002009

STATUS	TABLE	CATEGORY	SUBCATEGORY	LEVEL	NOTES	DESCRIPTION	% RECORDS
		Plausibility		CONCEPT			
PASS	MEASUREMENT	Plausibility	Atemporal	CONCEPT	None	For the combination of CONCEPT_ID 3002009 (Glucose [Mass/volume] in Serum or Plasma --12 hours fasting) and UNIT_CONCEPT_ID 8840 (milligram per deciliter), the number and percent of records that have a value higher than 150.0. (Threshold=5%).	2.42%

Showing 1 to 1 of 1 entries (filtered from 3,091 total entries)

Previous 1 Next

DQD for Cohorts

IBM CCAE Type 1 Diabetic cohort

Column visibility | CSV

Show All entries

Search: glucose

STATUS	TABLE	CATEGORY	SUBCATEGORY	LEVEL	NOTES	DESCRIPTION	% RECORDS
FAIL	MEASUREMENT	Plausibility	Atemporal	CONCEPT	None	For the combination of CONCEPT_ID 3002009 (Glucose [Mass/volume] in Serum or Plasma --12 hours fasting) and UNIT_CONCEPT_ID 8840 (milligram per deciliter), the number and percent of records that have a value higher than 150.0. (Threshold=5%).	33.55%

DQD for Cohorts

This has opened up many questions and areas for research…

- What are the use cases?
- Is this covered by cohortDiagnostics or should it be added in some form?
- Should the thresholds for the full database differ from the thresholds at the cohort level?
- What is the correct interpretation of the DQD results at the cohort level?

Lots of Helpful Guides

Ohdsi.github.io/DataQualityDashboard

Data Quality Check Type Definitions

Clair Blacketer

2020-07-30

Source: `vignettes/CheckTypeDescriptions.rmd`

Introduction

Running the DQD on a Cohort

Clair Blacketer

ere are two options in the `executeDqChecks` function,
t the DQD to the schema where the cohort table is located
nes that the table being referenced is the standard OHDSI
and **subject_id**. For example, if I have a cohort number
ecuteDqChecks function would look like this:

Failure Thresholds and How to Change Them

Clair Blacketer

2020-07-30

Source: `vignettes/Thresholds.rmd`

DQD Failure Thresholds

As described in the introduction to the tool, the Data Quality Dashboard works by systematically applying 20 data quality check types to

Future Projects

Near term goals

- Implementation of a unique check id
- Configure the tool to only run checks with a non-zero denominator
- Integrate CDM ETL standards with DQD checks

Long term goals

- Create a fit-for-use roadmap
  - Build out the DQD to assess fitness for use leveraging the phenotype library and characterization power of CHARYBDIS