PopMedNet™

DEPARTMENT OF POPULATION MEDICINE

HARVARD MEDICAL SCHOOL | Harvard Pilgrim Health Care Institute

# Distributed Health Data Networks: A Rapid, Systematic Approach to Verifying Machine-Generated Queries Within Different Systems

**Kyle Erickson; Jessica M. Malenfant, MPH; Kimberly Barrett, MPH; Adam Paczuski; Chayim Herzig-Marx, PhD; Jeffrey S. Brown, PhD**

**AMIA 2017 Annual Symposium,** Washington D.C., November 4-8, 2017

## Background

Distributed health data networks allow participants to maintain ownership over their data and to run queries and analytic programs in the privacy of their own data environments. Some networks, such as the PCORnet Distributed Research Network, utilize machine-generated queries powered by PopMedNet (PMN).

## Problem

Machine-generated queries, like those used in PCORnet, can be verbose and difficult for a human to parse. This presents various challenges when testing queries to ensure they perform as expected within different relational database management systems (RDBMS), especially in the fast-paced context of a distributed research network.

## Solution

We developed an open-source Test Case Inserter (TCI) tool and a standard testing process, enabling efficient verification of machine-generated queries with a high degree of confidence without requiring the tester to have any SQL knowledge. This process is not specific to PMN or PCORnet and can be used with any querying application, data model, and supported RDBMS.

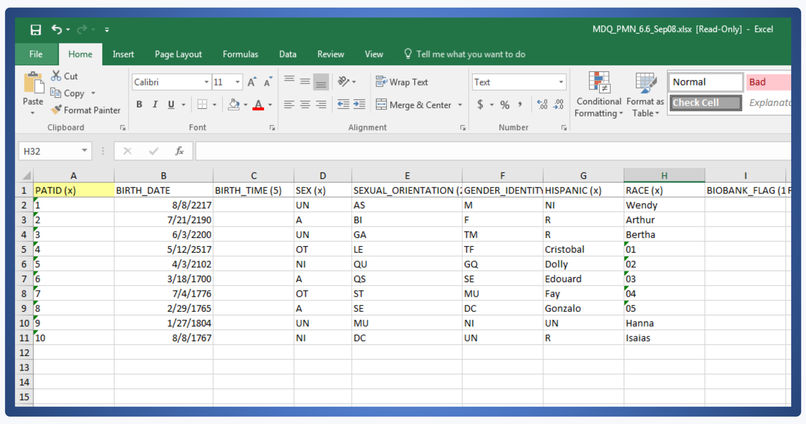
## Define Queries

**MDQ2: 3 Criteria Groups**

| Step | Action | Expected Behavior | Success (Y/N)? | Notes |
|---|---|---|---|---|
| 1 | Log into PMN 5.3.2 and create a new PCORnet MDQ request | Login is successful, pages load and display as expected | | |
| 2 | Name the MDQ "PX Codes Complex Relationships MDQ2: 3 Criteria Groups" and save the metadata | Metadata saves and the page looks the same as it does in PMN 5.3.1 | | |
| 3 | Add the Procedure Codes term, choose "HCPCS or CPT" as the code type, and enter the following codes as exact match: 19300, 19301; 19302; 19303; 19304; 19305; 19306; 19307; 38500 | The term should add to the criteria group the same way other terms are, and all info entered into the term should save | | |
| 4 | Add the Sex term and select Female | Term appears and info is entered the same way as in 5.3.1 | | |
| 5 | Add the Observation Period term and enter 1/1/2011 as the min date and enter 12/31/2015 as the max date | The term should add to the criteria group, AND with other terms, and all info entered into the term should save | | |
| 6 | Add the Age term and enter 18 as the min age, enter 89 as the max age, select the 'age as of user-entered date' calculation type, and enter the date: 12/31/2014 | The term should add to the criteria group, AND with other terms, and all info entered into the term should save | | |
| 7 | Add a second criteria group and name it "Group 2: DX inclusion" | The criteria group should AND with the previous criteria group | | |
| 8 | Add a Diagnosis code term to the second criteria group, choose "ICD-9" as the code type, and enter the following codes as starts with: 174 | The term should add to the criteria group the same way other terms are, and all info entered into the term should save | | |
| 9 | Add a third criteria group and name it "Group 3: DX exclusion" and select the exclusion criteria checkbox | The criteria group should AND NOT with the previous criteria group | | |
| 10 | Add a Diagnosis code term, choose "ICD-9" as the code type, and enter the following codes as exact match: V16.3 | The term should add to the criteria group the same way other terms are, and all info entered into the term should save | | |
| 11 | Add the Hispanic and Race stratification and remove all other stratifications | The Hispanic and Race stratification should be the only ones listed | | |
| 12 | Add the 12 DataMarts from the About section and click submit | The page reloads and the routing status for all 12 DataMarts is now "Submitted" | | |
| 13 | Wait for the requests to execute | Eventually, all DataMarts should execute successfully | | |
| 14 | Navigate to the request in the Web Portal | All DataMart routings should have a status of Complete | | |
| 15 | View and download the Individual results and Aggregate results for all 12 DataMarts | The individual results should match the expected patients and the aggregate results should match the individual results added up together<br>Expected results: Chris, Florence, Helene, Joyce, Michael | | |

- With more complex research questions and query enhancements, functionality is continuously tested
- A specific query or set of queries are defined & test plans are developed
- Menu-Driven Queries in PopMedNet allow for theoretically limitless query complexity

## Create Test Patients



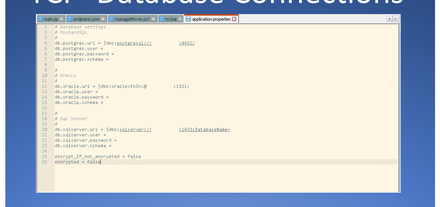

| | A | B | C | D | E | F | G | H | I |
|---|---|---|---|---|---|---|---|---|---|
| 1 | PATID (x) | BIRTH_DATE | BIRTH_TIME (5) | SEX (x) | SEXUAL_ORIENTATION (x | GENDER_IDENTITY | HISPANIC (x) | RACE (x) | BIOBANK_FLAG (1F |
| 2 | 1 | 8/8/2217 | | UN | AS | M | NI | Wendy | |
| 3 | 2 | 7/21/2190 | | A | BI | F | R | Arthur | |
| 4 | 3 | 6/3/2200 | | UN | GA | TM | R | Bertha | |
| 5 | 4 | 5/12/2517 | | OT | LE | TF | Cristobal | 01 | |
| 6 | 5 | 4/3/2102 | | NI | QU | GQ | Dolly | 02 | |
| 7 | 6 | 3/18/1700 | | A | QS | SE | Edouard | 03 | |
| 8 | 7 | 7/4/1776 | | OT | ST | MU | Fay | 04 | |
| 9 | 8 | 2/29/1765 | | A | SE | DC | Gonzalo | 05 | |
| 10 | 9 | 1/27/1804 | | UN | MU | NI | UN | Hanna | |
| 11 | 10 | 8/8/1767 | | NI | DC | UN | R | Isaias | |
| 12 | | | | | | | | | |
| 13 | | | | | | | | | |
| 14 | | | | | | | | | |
| 15 | | | | | | | | | |

- We create patients to meet our test cases in an Excel document adhering to the PCORnet Common Data Model
- We assign fake identifiers to each test patient for tracking
- Patients are created to either be included or excluded in query results
- The TCI allows us to create as many patients as appropriate

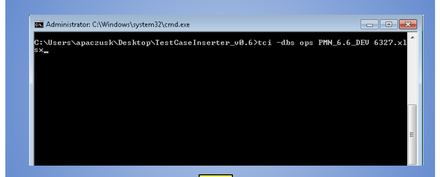
## Insert Data

- Patients are inserted into multiple instances of Oracle, Postgres, and SQL Server via the TCI
- TCI is a Java-based command line application
- TCI ensures the same patients are entered in all RDBMS instances

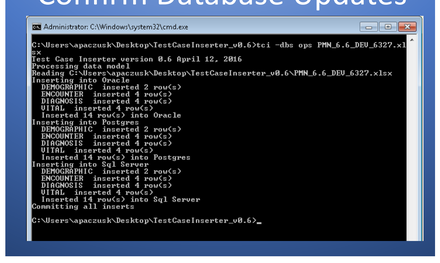
TCI –Database Connections



Run TCI

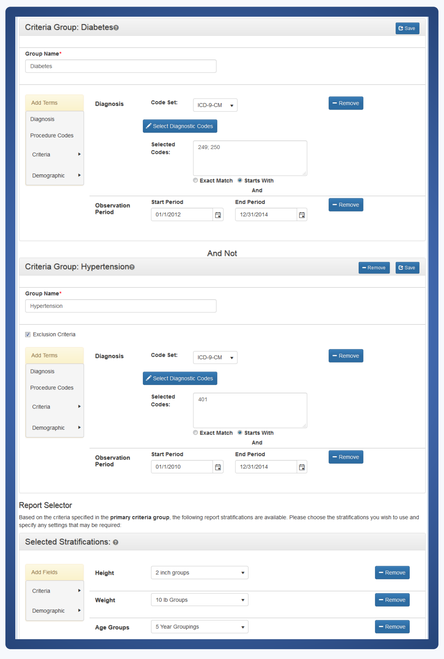



Confirm Database Updates




## Run Queries

- Queries being tested are created in PopMedNet and executed against updated databases
- Some validation queries are modified to stratify on fields used for fake patient identifiers




## Verify Query Results

- Results are reviewed against the expected patient results as defined in original test plans
- Queries are given a Pass or Fail verification

| Hispanic | Race | Patients |
|---|---|---|
| Barry | 06 | 1 |
| Y | Arlene | 1 |
| Y | Chris | 1 |
| Y | Cindy | 1 |
| Y | Don | 1 |
| Y | Florence | 1 |
| N | Helene | 1 |
| N | Isaac | 1 |
| N | Joyce | 1 |
| N | Kirk | 1 |
| N | Michael | 1 |
| Y | Nadine | 1 |
| Y | Sean | 1 |

- If a query fails, the source data are investigated
- Additional test queries and/or patients are created and SQL is inspected to reproduce and identify issues
- If PMN defects are found, they are logged, fixed, and retested using the same process and test cases in all supported systems

| Currently supported RDBMS Platforms | Currently supported CDM versions |
|---|---|
| Oracle 11, 12 | PCORnet CDM 3.0, 3.1 |
| Postgres 9.4, 9.5 | PCORnet CDM 3.0, 3.1 |
| SQL Server 2012, 2014 | PCORnet CDM 3.0, 3.1 |

## Outcomes

- This testing process has successfully identified unexpected query behavior in various conditions, most of which are complex queries involving joins with patient-record information across several tables, or utilizing different encounter time windows each associated with different patient information within a single query.
- Unexpected use of query input parameters can be a major contributor to errors in resulting query behavior. PopMedNet employs the use of request templates to reduce user error without sacrificing flexibility.

## Future Work

- This methodology will continue to be used to verify PCORnet Menu-Driven Queries in PopMedNet
- Investigate the use of PATID fields and other ways of identifying specific test patients
- Investigate test automation where applicable

Special Thanks: Hozefa Divan, PhD, MSPH for scientific guidance in creating test plans and Zachary Wyner, MPH for poster design.

Scan for more information on the TCI: