CSE8803 Project: OHDSI Surgeon Scorecard

Michael Shamberger

***Abstract*—CSE8803 Project: OHDSI Surgeon Scorecard creates a new OHDSI tool based on the OMOP data format.**

This project attempts to reproduce the work done by Pro Publica data scientists that made the individual surgeon complication rate for eight low risk inpatient procedures in the United States public for the first time in 2015.

***Index Terms*—Big data, Health analytics, Data mining, OHDSI, Surgeon Scorecard, Pro Publica, Open source**

I. Introduction and Motivation

DATA scientists at Pro Publica performed a study on the US Medicare claims data to determine the complication rate of surgeons for eight low risk inpatient surgeries. They created an application that allows the general public to query the complication rates of local surgeons. This was the first time that surgeons in the United States had their complication rates published to the public. This was made possible by a change in how Medicare codes the physician names. In 2014 the encrypted identifiers were changed to allow access to the real names of the surgeons.

There is a rigorous methodology available for evaluating surgical outcomes. It is from the American College of Surgeons, and is called NSQIP. It is the "leading nationally validated, risk-adjusted, outcomes-based program to measure and improve the quality of surgical care in the private sector." [1] The problem with NSQIP is that algorithm is not published and the results must remain confidential to those who use it.

The researchers at Pro Publica published their methodology but did not release their source code. Having an open source algorithm is necessary to allow researchers to modify and iterate on the algorithms as there is no public consensus in the medical profession exactly how a surgeon's complicaton rate should be measured [3]. The goal of this work is to reproduce the Surgeon Scorecard and publish the application code as a scalable open source OHDSI application.

II. Methods

A. Data

The data used for analysis is the synthetically made Medicare 2008-2010 Data Entrepreneurs SynPUF converted to an OMOP format. The orinal Pro Publica study was performed against the Medicare 100% Standard Analytic Files for years 2009-2013.

The Data Entrepreneurs SynPUF contains the medical claims information for 2.32 million synthetic patients.

The medicare SynPuf data preserves the detailed data structure of key variables at both the beneficiary and claim levels [2]. There is no guarantee that the data will produce meaningful results at the individual surgeon level.

B. Identifying Patient Cohort

The Cohort is made up of the 1.88 million patients that had a hospital visit during the period 2008-2010. No screening for existing comorbidies was applied to the Cohort.

Patients	N
All Patients	2,326,857
Patients with visit	1,885,277

TABLE I
COHORT INFORMATION

C. Identifying Procedures for Analysis

Similar to the Pro Publica study, eight elective inpatient procedures were selected. These surgeries include hip replacement, knee replacement, 3 types of spinal fusion, gall bladder removal, prostate removal, and prostate resections.

ICD9 CODE	Procedure	N
51.23	Laparoscopic cholecystectomy	11,263
60.5	Radical prostatectomy	2,585
60.29	Transurethral prostatectomy (TURP)	2,746
81.02	Cervical fusion of the anterior column, anterior technique	2,634
81.07	Lumbar and lumbosacral fusion of the posterior column, posterior technique	11,122
81.08	Lumbar and lumbosacral fusion of the anterior column, posterior technique	10,622
81.51	Total hip replacement	10,903
81.54	Total knee replacement	29,013
Total		80,888

TABLE II
MEDICARE SYNPUF STUDY: PROCEDURES INCLUDED IN ANALYSIS AND COUNT OF INDEX ADMISSIONS IN PATIENT COHORT

D. Identifying Index Admissions

An index procedure is defined as one of the individual surgeries for which the outcome was analyzed.

- Each procedure in Table I was identified by a set of qualifying concept primary admission icd9 codes.
- The procedure is checked to make sure it occurs during an inpatient visit.

The following were part of the pro-publica study but were not implemented:

- Excludes patients if the surgery occurred during an ER visit
- Excludes patients transferred from a nursing home or correctional facility.
- A complication risk adjustment for each surgeon based on patient comorbities and age is not applied.

E. Complications

This show the total complications found from the patient cohort where the complication icd9 code is the primary admission code.

Complication Type	N	Example
Infection	2,834	998.59-Postoperative infection
Clot	118	415.11-Iatrogenic pulmonary embolism
Reaction	242	996.69 - Infection and inflammatory reaction due to internal joint prosthesis
Mechanical	513	996.47 - Mechanical complication of prosthetic joint implant
Sepsis	15,399	03.89 - Septicemia
Bone	406	996.44 - Peri-prosthetic fracture around prosthetic joint
Hemotoma	574	998.12 - Hematoma complicating a procedure
Wound	154	998.2 - Accidental puncture or laceration during a procedure
Hemmorrhage	567	998.11 - Hemorrhage complicating a procedure
Pain	230	338.18 - Acute postoperative pain
Digestive	0	997.49 - Digestive system complications
C.diff	1,521	00.845 - Intestinal infection due to Clostridium difficile
Misc. Comp.	35	787.22 - Dysphagia, oropharyngeal phase
Seroma	113	998.13 - Seroma complicating a procedure
Fever	49	780.62 - Postprocedural fever
Urinary	192	997.5 - Surgical complications of the urinary tract
Total	22,947	

TABLE III
EXAMPLES OF 20 TYPES OF COMPLICATION ASSOCIATED WITH SURGERY

For each patient that was found to have an index procedure, the data was searched for two negative outcomes.

- Death - Patient died during hospital stay or within 30 days of being discharged.
- Complication - Patient was discharged alive but was admitted to a hospital within 30 days[4] of discharge with a principal diagnosis indicating a negative surgical outcome.

The index readmissions were detected by searching a set of icd9 primary diagnosis codes for codes that were indicating a complication based on the index procedure and withing the time window.

For the full list of surgical codes that were determined to indicate surgical complications for the procedures of interest, see Appendix B online here.

F. Surgeon Complication Rate

The complication rate for each surgeon was calculated by taking the total amount of index procedures and dividing those by the sum of instances of death and readmission for that procedure. While the Pro Publica study applied a risk adjustment to each complication rate this study only produces the raw surgeon complication rate.

G. Further Information

More detailed information about the medical justifications for selecting these procedures and the ICD9 codes that indicate them in the Medicare data can be found in the Pro Publica whitepaper Surgeon Level Risk Methodology.

ICD9 CODE	Procedure	N
51.23	Laparoscopic cholecystectomy	136
60.5	Radical prostatectomy	14
60.29	Transurethral prostatectomy (TURP)	12
81.02	Cervical fusion of the anterior column, anterior technique	21
81.07	Lumbar and lumbosacral fusion of the posterior column, posterior technique	97
81.08	Lumbar and lumbosacral fusion of the anterior column, posterior technique	91
81.51	Total hip replacement	139
81.54	Total knee replacement	383
Total		893

TABLE IV
MEDICARE SYNPUF STUDY: READMISSIONS DUE TO SURGICAL COMPLICATION FOR EACH OF THE PROCEDURES INCLUDED IN ANALYSIS, BROKEN DOWN BY PRINCIPAL DIAGNOSIS ON READMISSION.

Procedure	N
51.23	7,389
60.5	2,103
60.29	2,248
81.02	2,189
81.07	6,964
81.08	6,719
81.51	6,648
81.54	14,041

TABLE V
MEDICARE SYNPUF STUDY: NUMBER OF SURGEONS BY PROCEDURE

III. IMPLEMENTATION

A. Analytic Infrastructure

A dedicated server with 64 GB of ram and 8 cpu cores was used for the scalability testing of the PySpark application.

B. Technologies

The following technologies were used to implement the Cohort and Readmission components as well as the Surgeon Scorecard application:

- Python - The OHDSI communities existing tools show that they are most comfortable with the scripting languages and sql logic. PySpark with Spark Sql was used.
- Python Pandas - The Surgeon Scorecard sometimes sends processed data to a pandas dataframe. This is only done when it is known that the data can fit into the memory of a single server. There are more API's implemented for processing Dataframes in pandas then are currently available in PySpark allowing for rapid development.
- Spark - The ability to cache results and run across clusters of machines makes Spark a superior choice to relational databases once the data size is beyond what a single server can process.

C. Software Components

- Cohort- A PySpark application class Cohort was developed. Based on a properties files, OMAP format data is read in from CSV files and a cohort of patients is selected. The data can be read uncompressed or compressed in various formats. After the cohort is selected, the remaining data is filtered to remove data that is not related to this cohort. The option exists to write this cohort and filtered data back out to OMAP format CSV files. This smaller dataset can be loaded into a database in order to be able to use the full set of OHDSI tools.

- Readmissions- The PySpark application class Readmission takes a set of OMOP data and detects hospital inpatient readmissions following a procedure of interest. It can be used in combination with the Cohort class to find readmissons on a filtered cohort of users. The codes of procedures of interest and their possible complication codes are defined in a properties files diagnosis.properties and readmission.properties. There is no limit to the amount of procedures that can be investigated during a single run.

D. Library Gaps

The application can work with other codes besides icd9 such as icd10 and SNOMED as long as consistent codes are used in both the diagnosic.properties and readmission.properties as well as the data being analyzed. There is no functionality for conversion of codes.

There are no existing open source Python libraries available for conversion of medical codes such as between icd9 and icd10 or making relationships between these codes and comorbidies. An R library exists for icd conversion and relation to comorbidies but was not converted to python and incorporated into the application.

E. Data Issues

There are some data conversion issues with the existing OHDSI CMS ETL that impacted the study. The Pro Publica study used primary admission diagnosis for identifying a qualifying procedure. Examination of the logic in the ETL showed that while the primary admission diagnosis was mismapped for outpatient to a first position diagnosis, it was completely dropped when doing an inpatient record migration. The mapping to the CONDITION_TYPE_CONCEPT_ID that holds the information on whether the diagnosis was primary or not was incomplete in both inpatient and outpatient record conversions.

Concept Type ID	Description	N
38000200	Inpat Condition 1st Postion	8,317,475
38000251	Inpat Procedure 1st Position	3,592,580
Total		11,910,055

TABLE VI
INITIAL INPATIENT COUNTS OF CONCEPT_TYPE_ID FROM CMS-ETL

To fix this, the python ETL was modified and github pull request was sent to the OHDSI CMS-ETL project. The actual position in the Medicare source data was not considered important. Only the difference between primary or not primary. The ETL was re-run and the scorecard was recalculated based on the correctly mapped data.

F. Scalability Issues In Spark

Spark has some 2G limits in it's 2.x codebase. These are due to 32 bit types used in various places in the code. The first of these happens when reading a data block that has been stored to disk. The code uses an instance of MappedByteBuffer which cannot exceed a size of 2G. The second is when using kryo serialized data. The serialized data is stored in a byte[] the size of which cannot exceed 2G. The third situation occurs when RPC writes data to be sent to a channel. The code uses a ByteBuf which means that it cannot transfer data over 2G in memory. The fourth and final issue occurs when an RPC message is received where again a ByteBuf is used which cannot exceed 2G. The result to the end user is the same for all of the above issues. Their big data application crashes with a "Size exceeds Integer.MAX_VALUE at sun.nio.ch.FileChannelImpl.map".

Concept Type ID	Description	N
38000199	Inpat Condition Primary Position	959,691
38000200-209	Inpat Condition Not Primary	7,932,634
DNE	Inpat Procedure Primary Position	291,287
38000251-260	Inpat Procedure Not Primary	3,383,288
Total		12,566,900

TABLE VII
FINAL COUNTS OF CONCEPT_TYPE_ID FROM CMS-ETL

The recommended workaround for this limitation is to set additional partitions beyond the default 200 on the dataframes. The properties file for the application allows the partition size to be set for all dataframes created. Tests were done with up to 4000 partitions but instability was still seen as the study parameters were modified.

There is currently a patch pending to fix this issue but it has not been incorporated to a released version of spark. The Spark foundation team has so far refused to incorporate the patch as it contains 143 files and 4400 lines of code. There are some discussion that the issue is not serious because users can workaround it by increasing the number of partitions in their Dataframes or modify their application code. I disagree as users with huge datasets have had the issue occur from simple joins and counts. See Table V for the breakdown of the 1.6 billion records in the SynPuf dataset.

In order for the scorecard to run reliably, the patch was compiled into the main branch and resulting binary was deployed to the implementation server. The patch solved the issue and this was reported back to the pull request in github.

IV. RESULTS

A. Surgeon Scorecard

A set of reusable usable software components were developed that can be used with the OHDSI OMOP data format. These components were used to reproduce the Surgeon Scorecard raw readmission rates.

The library successfully calculates the complication rate for each of the 8 inpatient surgical procedures of interest. The scorecard results for the 1000 patient dataset can be seen here. The full data set scorecard results are available here.

Similar to the Pro Publica Surgeon Level Risk Appendix, index procedure counts using icd9 codes of interest were generated and can be seen at this location. The counts of index complication codes were calculated and available here.

OMOP Table	N
procedure_cost	637,854,416
observation_period	2,098,516
provider	905,493
measurement	74,128,430
observation	37,531,052
location	3,186
payer_plan_period	7,789,513
device_exposure	4,465,486
care_site	320,546
death	107,645
procedure_occurrence	264,056,962
drug_cost	111,085,970
person	2,326,857
visit_occurrence	111,637,583
condition_occurrence	276,693,282
drug_exposure	126,048,051
Total	1,657,052,988

TABLE VIII
MEDICARE SYNPUF STUDY: RECORD COUNTS

B. Comparison To Pro Publica Results

The Pro Publica study used the Medicare 100% Standard Analytic Files for years 2009-2013 and included those patients that were enrolled in the Medicare Fee-For-Service program. The Medicare Fee-For-Service program included around 52.5 million patients in 2013.

The SynPuf dataset is about 1/20 of the datasize and half the duration so it is expected that fewer index procedures and index complications would be identified.

Description	N Medicare	N SynPuf
Total Procedures	2,376,851	80,888
Total Complications	64,367	893

TABLE IX
COMPARISON OF FULL MEDICARE DATA RESULTS VS SYNPUF DATASET

C. Cohort Selection

Another benefit of this work is to allow the OHDSI tools to be used even when the data is initially to massive to be loaded into a relational database without requiring special hardware and skills. Researchers are able to build a Cohort on a subset of the data based on their specific criteria and generate csv format files that can be loaded into a relational database to make use of the rich set of OHDSI tools already in place.

D. Big Data Tool Improvements

The application encountered data size limitations in Spark and published a method to bypass them by patching the main branch of code.

The CMS-ETL produced the first publically available massive health care dataset in the OMOP format. Defects in the ETL related to conversion of the primary admission diagnosis were uncovered and a fix provided to the OHDSI.

V. FINAL COMMENTS

The synthetic data was adequate to build the application, but access to the real deidentified Medicare data is still limited to those with corporate resources or scientific grants. Even with access to the real data such as Pro Publica had there are still limitations on making the results of studies public. Statewide data sets exist but have to be purchased individually where each can cost in the tens of thousands of dollars.

More comprehensive open source medical libraries need to be built into the OHDSI tools. The current OHDSI tools are built as applications but would benefit from having a set of reusable components implemented in a scripting language such as python. Python API's for medical code conversion (For example: icd9/10 to SNOMED) and relationships to comorbidies are also needed. The benefits of having a common schema to combine datasets are limited if it is not possible to have common medical codes within it.

The synthetic data provided by CMS is a good start and the open source Surgeon Scorecard application could not have been developed without it. There is a need for a refresh of the data since the medicare coding has moved to icd10 codes in 2015 from icd9. An application will need to be able to handle both data with icd9 and icd10 codes as well as convert the icd9 codes to icd10 in order to perform multi-year studies.

VI. SUPPLEMENTAL MATERIAL

Source code - https://github.com/opme/SurgeonScorecard/

REFERENCES

[1] Acs national surgical quality improvement program. https://www.facs.org/quality-programs/acs-nsqip. Accessed: 2016-11-23.

[2] De 1.0 data users document appendix a. https://www.cms.gov/Research-Statistics-Data-and-Systems/Downloadable-Public-Use-Files/SynPUFs/Downloads/SynPUF_DUG.pdf. Accessed: 2016-11-23.

[3] M. W. Friedberg, K. Y. Bilimoria, P. J. Pronovost, S. D. M., C. L. Damberg, and A. M. Zaslavsky. Response to propublica's rebuttal of our critique of the surgeon scorecard. *http://www.rand.org/pubs/perspectives/PE170z1.html Santa Monica, CA: RAND Corporation, 2015.*

[4] H. M. Krumholz, Z. Lin, E. E. Drye, M. M. Desai, L. F. Han, M. T. Rapp, and S. L. T. Normand. An administrative claims measure suitable for profiling hospital performance based on 30-day all-cause readmission rates among patients with acute myocardial infarction. *Cardiovascular Quality and Outcomes 4(2), 243-252.*