Data Impact Challenge Answer Submission

QUESTION: *What is the rate of repeated laboratory tests within a ninety (90) day period*?

TEAM MEMBERS

- **Ashraf Abusharekh, A2I Solutions**
- **Syed Sibte Raza Abidi, NICHE Research Group, Faculty of Computer Science, Dalhousie University**
  Samuel Alan Stewart, NICHE Research Group, Faculty of Medicine, Dalhousie University
- **Bryan Crocker, Calvino Cheng, Don Dorion, Pathology Lab, QEII HSC, Halifax, NS**

DESCRIPTION OF THE DATA AND THE ANALYSIS

- Data Custodian Organization(s) and data sources: **Pathology Lab, QEII HSC, Halifax, NS**
- List of Datasets Used: **Pathology Lab Clinical Database. Overall the dataset has 4,299,548 lab tests (17,094,825 procedures) done on 204,662 patients and ordered by 1,981 physicians.**
- Exclusions: **Hospital-based Inpatient and outpatient laboratory test orders**
- Nature and Size of Cohort (e.g. geographic area covered, number of patients included): **The Capital District Health Authority (CDHA), serving a population of over 400,000 people. The CDHA is now defunct, and it has been amalgamated with NS Health Authority (NSHA) in 2015.**
- Data timeframe: **2011-August-01 to 2014-June-20**

DATA ANALYSIS METHODOLOGY

In response to the above data impact challenge, this health data analytics project aims to analyze the lab test requisition pattern of family physicians within the CDHA jurisdiction. The lab tests were conducted by the QEII HSC Pathology Lab. De-identified flat files (in csv format) were exported from the Pathology Lab and stored into a secure HBase nosql datastore, and data analytics was done using the R statistical software (ver. 3.1.0) along with RHadoop.

The unit of analysis is a single ***lab test***, with the 15 most popular lab tests being studied. Note that a single test may comprise 1 or more specialized procedures (for example, the CBC test comprises 11 procedures, where each procedure has an individual result). Our analysis recognizes that one or more lab tests can be *ordered* by a physician for a patient in a single test requisition order. Each requisition order is time-stamped and the order date is used for the repeat test analysis. The lab test results are reported in terms of numeric values, and for some tests there is predefined normal range which determines whether the test result is normal or not.

Our methodology is as follows: (a) Each patient has a unique **PatientID** (say patient_A, patient_B) and each of the 15 lab tests studied has a unique **TestID** (say test_{CBC}, test_{AST}). To identify the specific lab tests ordered for each patient throughout the study period, we created a **Patient-TestID** by concatenating the PatientID and the TestID. A patient can have a range of different lab tests, therefore for each patient there will be multiple Patient-TestIDs such as patient_A-test_{CBC}, patient_A-test_{ALT} and so on; (b) for each Patient-TestID (say patient_A-test_{CBC}), we created a **timeline** showing the dates when the test was ordered, eg. patient_A-$\text{test}_{CBC}(t_1, t_2, \dots, t_n)$; (b) A **time lapse**, l_i, for each Patient-TestID was calculated as the difference between two consecutive tests using a patient-test timeline, i.e, $l_i = t_i - t_{i-1}$; (c) Any time lapse value that was less than 90 days was counted as a *repeat test.* Note that, the time lapse

period can be easily varied to study different repeat rates; (d) Steps a-c were applied for all the patient-testIDs—i.e. covering all tests ordered for all patients in the dataset; (e) The repeat test count for each of the 15 tests was aggregated (across all patients) to get the overall test repeat rate.

For analyzing the repeat tests, we also considered the results of the previous test order to get insights into the test requisition pattern and to establish the rationale (or not) for the repeated test order.

DESCRIBING THE FINDINGS AND DISCUSSION

Numerator: 803,227 (Repeated tests) **Denominator: 4,299,548** (Total tests ordered)
The table below presents the analytics results, highlighting the overall repeat test rate for all tests and repeat rates stratified by both test type and test result. **The overall lab test repeat rate is 18.68%**. But there are significant variations across specific tests, with cholesterol being repeated only 6.7% of the time and PT testing repeated 76.2% of the time.

When analyzing the test repetition pattern based on test results, it is noted that a positive test result leads to more repeat testing than after a negative result. We noted that 44.27% of tests after a positive result were repetitions, compared to 28.6% of tests were repeated after a negative result. This pattern holds across all lab tests types, albeit with varying magnitude of repetition. We posit that correlating the pathology of the patient diagnosis with the tests order may provide further rationale for this behaviour.

There are two limitations to this analysis: (i) the exclusion of hospital-based test orders prevents a an understanding test repetition pattern in hospitals; and (ii) the absence of the patients' diagnosis leaves a gap in interpreting the rationale for the repeated test.

	Repeat Tests	Total Tests	Repeat Rate	*After Positive Test Result*			*After Negative Test Result*		
				Repeat	Total	Rate	Repeat	Total	Rate
ALL Tests	**803227**	**4299548**	**18.68%**	**260332**	**588003**	**44.27%**	**337758**	**1182799**	**28.56%**
PT	108422	142308	**76.19%**						
CBC Auto Diff	146478	532801	**27.49%**	113204	231979	**48.80%**	9877	39487	**25.01%**
Glucose Random	17038	66192	**25.74%**	137	285	**48.07%**	474	951	**49.84%**
Creatinine	101531	446511	**22.74%**	33890	54568	**62.11%**	67641	218721	**30.93%**
Alkaline Phosphatase	44094	195550	**22.55%**	11094	20600	**53.85%**	33000	85526	**38.58%**
Urea	65252	290559	**22.46%**	20825	33950	**61.34%**	44427	133572	**33.26%**
Electrolyte Panel	86513	402525	**21.49%**	33727	72931	**46.25%**	52786	167414	**31.53%**
AST	51082	252190	**20.26%**	7995	15431	**51.81%**	43087	120650	**35.71%**
ALT	64350	353162	**18.22%**	15332	38044	**40.30%**	49018	162836	**30.10%**
GGT	18018	120304	**14.98%**	3715	9255	**40.14%**	14302	46787	**30.57%**
Glucose AC	23839	304959	**7.82%**	15537	93035	**16.70%**	8302	65795	**12.62%**
Cholesterol	19438	291114	**6.68%**						
Triglycerides	19403	290505	**6.68%**						
HDL Cholesterol	18049	288642	**6.25%**						
Thyroid Hormone	19720	322226	**6.12%**	4876	17925	**27.20%**	14844	141060	**10.52%**