STATE OF MAINE HEPATITIS C TRACKING

STUDENT:

JOSHUA BROBST

STAKEHOLDER:
CHLOE
MANCHESTER

WHAT IS HEPATITIS C (HCV)?

- Hepatitis C Virus (HCV) is a liver disease that passes through contact (even microscopic amounts) of infected blood.
- When left untreated, it can lead to liver scarring or cancer.
- It is a "notifiable condition" in Maine, meaning that all positive lab results are required by law to be reported to the Department of Health and Human Services (DHHS).

PRIMARY QUESTIONS

HVC Clearance Cascade:

This is the highest priority analysis for the project. The CDC is interested in finding out how many people are at each stage of the HCV 'clearance cascade' -

1. Antibody Test
2. RNA Test
3. Genotype Test (Optional)
4. Cured/Cleared Infection
5. Reinfection (Hopefully none).

Monitoring how often patients make it from one end to the other is important in identifying where resources are lacking and what treatments work.

Testing patterns:

Looking at the Hepatitis C labs, it will be analyzed what patterns are able to be noticed in the testing behavior. Example testing questions are:

- How many antibody tests are patients getting before they get a confirmatory RNA test?
- What factors are associated with repeat RNA tests but not achieving cure?

DATA ACQUISITION

- Started in July of this year.
- All data comes directly from the state of Maine CDC.
- Datasets are case-patient records meaning data points in the surveillance system represent individuals with HCV.
- Some case-patient investigations are more complete than others because of how different Hepatitis C conditions are prioritized.
- Due to the disease's nature as a "notifiable condition", there is high confidence that the dataset is representative of the whole population in the state of Maine.

INITIAL DATA STRUCTURE

- Two datasets -
  - Cases: 34,686 rows x 14 cols [485,604 cells]
  - Labs: 832,106 rows x 8 cols [665,6848 cells]

Cases	Definition
Disease	Disease status, either acute or chronic
HCV_Genotype	Genotype test result (genotype)
HCV_Genotype_Detected	Genotype test result (Y/N)
HCV_RNA	RNA test result
HCV_RNA_Date	RNA test collection date
Investigation_Case_Status	Probable/Confirmed Status
Year	Year of Investigation
Patient_State	State, should be Maine
Specimen_Collection_Date__HCV_Ge	Genotype test collection date
total_anti_HCV	Anti-HCV test result
total_anti_HCV_Date	Anti-HCV test collection date
County	Patient County
Patient ID (encoded)	Encoded Patient Tracker

- When combined, without cleaning -
  - 855,098 rows x 21 cols [17,957,058 cells]

Labs	Definition
Coded_Result	Lab Result
Date_Specimen_Collected	Specimen collection date
Numeric_Results	Lab Result
Resulted_Test_Name	Name of test performed
Test_Result_Code	Lab Result
Text_Result	Lab Result
Reporting_Facility	Facility that submitted the lab
Patient ID (encoded)	Encoded Patient Tracker

DELIVERABLE 1:

REPRODUCIBLE DATA CLEANING

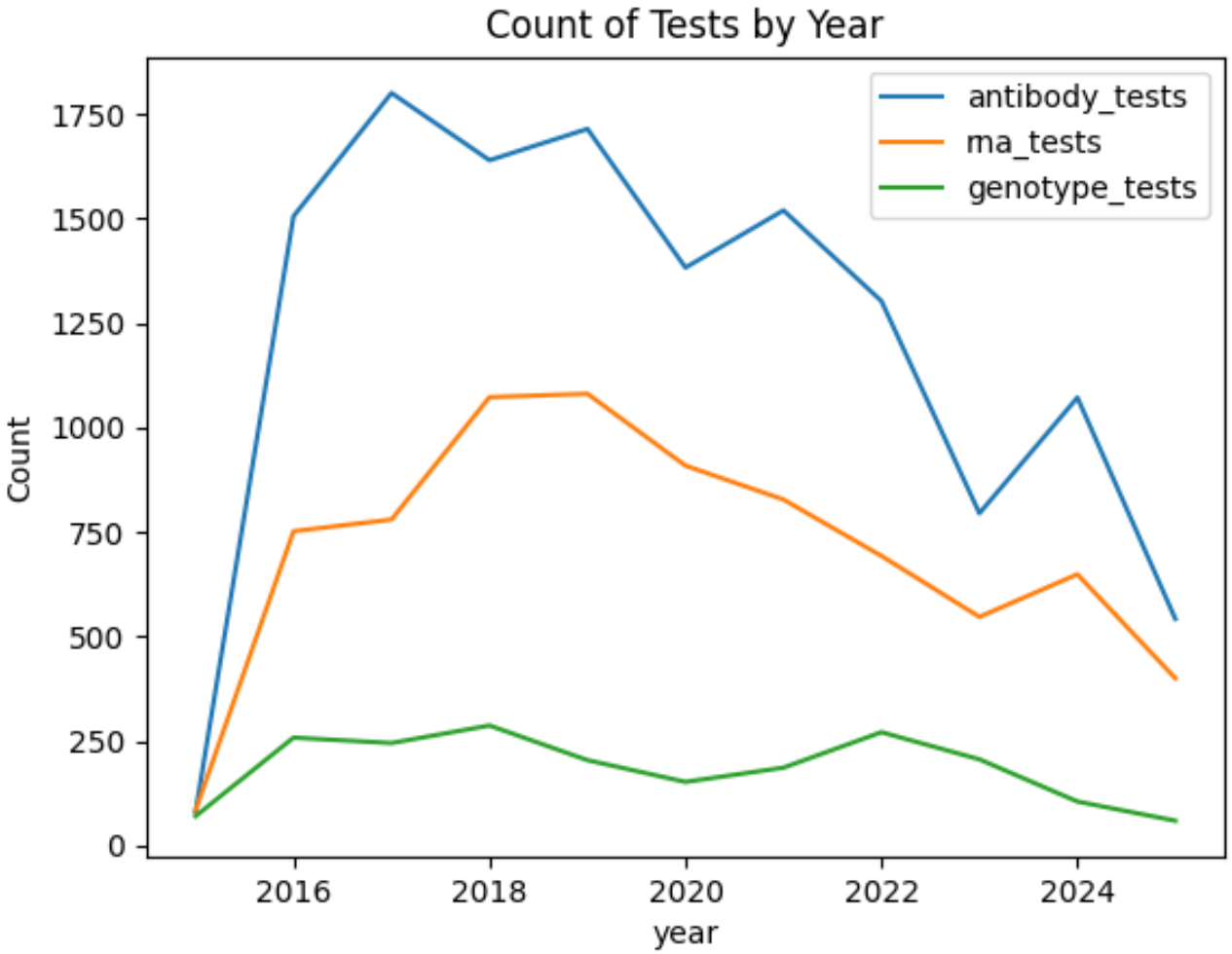
- Due to the data's nature, the CDC did not have a reproducible view that combines the prior Excel files.
- There are also repeat columns that should be combined, and some tests that are improperly captured (such as tests for Hepatitis A/B/D).
- In total, this meant that tracking the Clearance cascade was a challenge, because tests could not be aggregated or sorted chronologically.

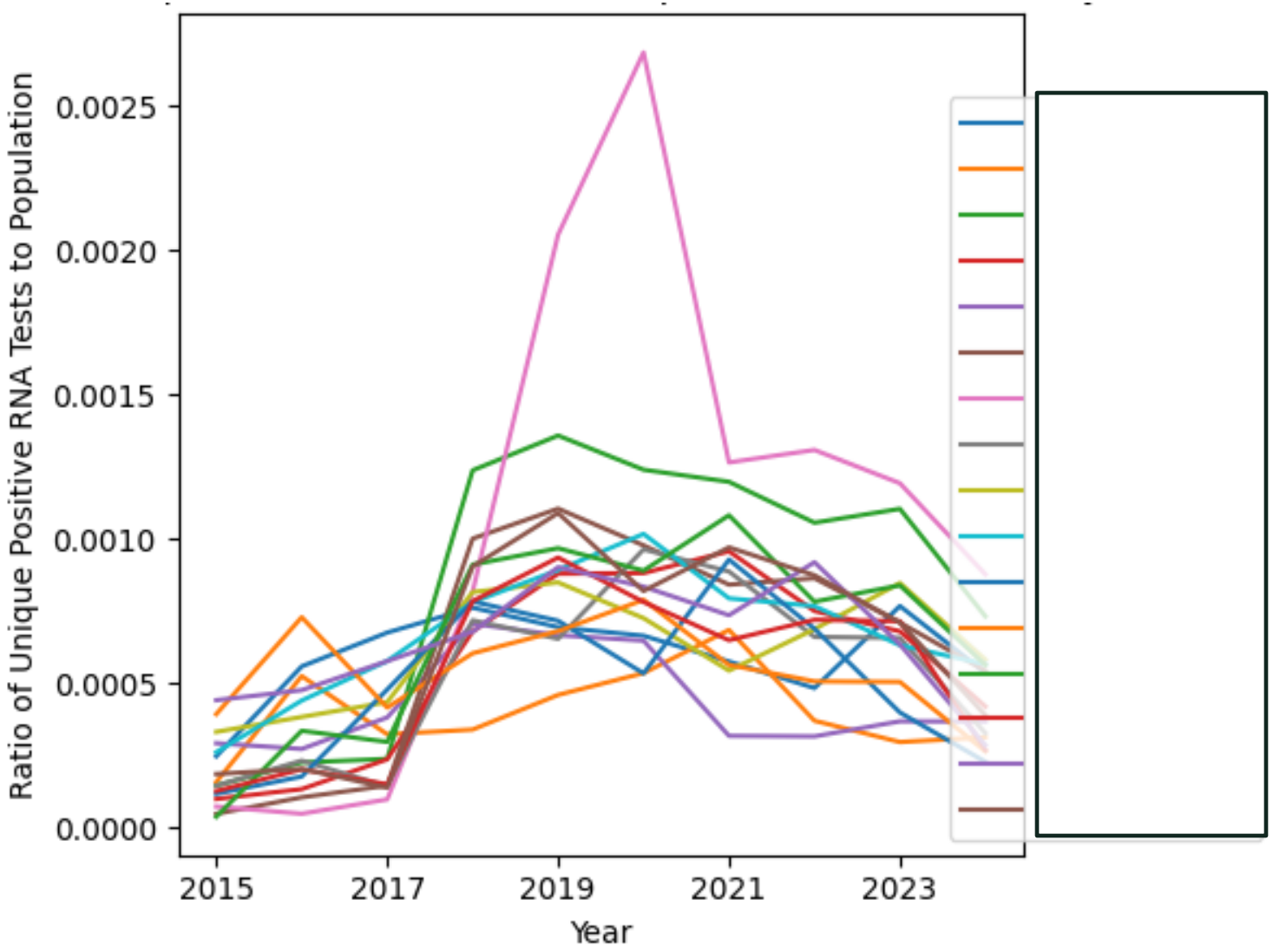
LONGFORM

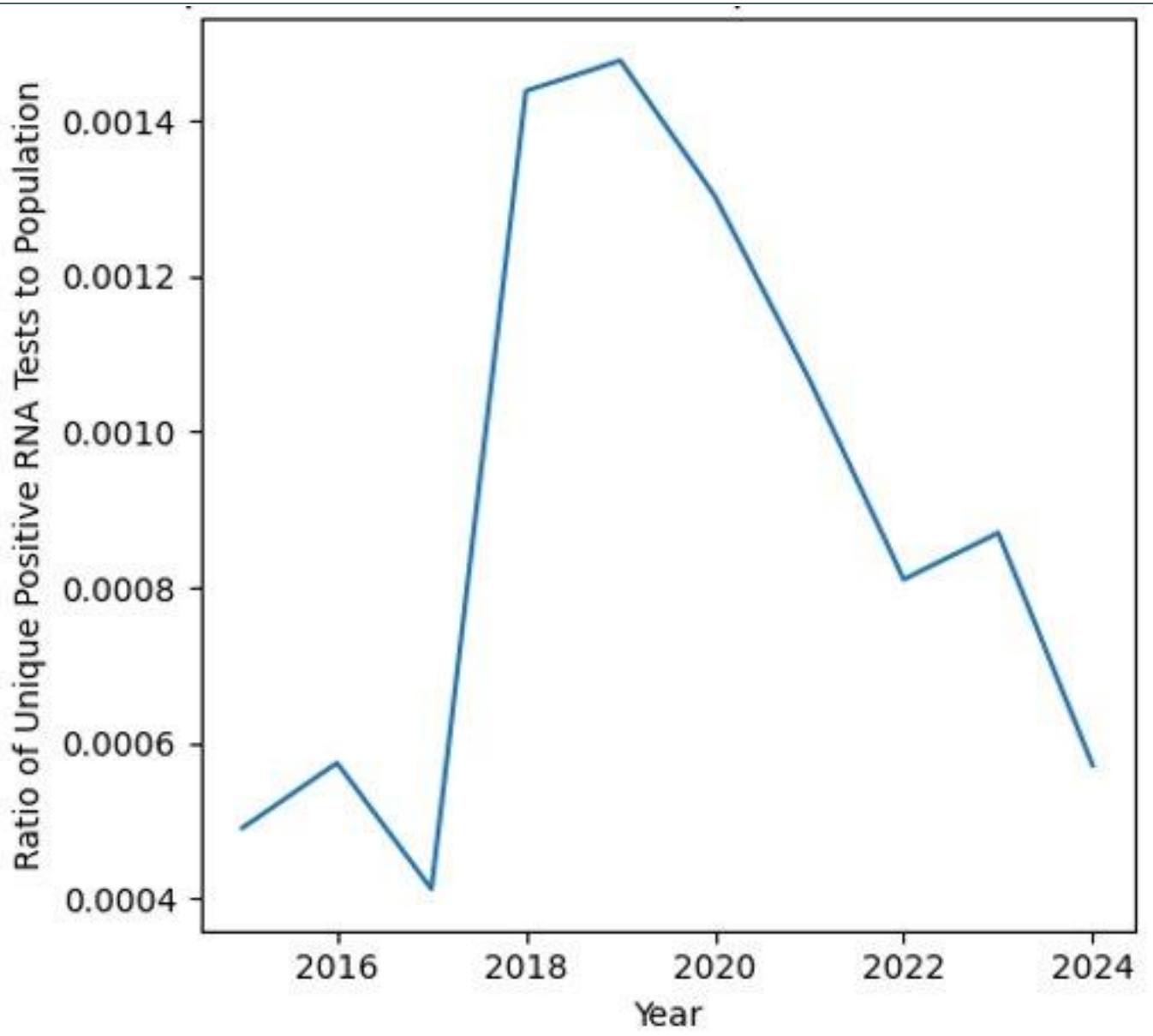
patient_id	classification	years	test_facility	test_date	test_type	test_result
PAT0000001			Affiliated Laboratories	2017-11-14	antibody	negative
PAT0000004	chronic	[2018]		2018-04-10	antibody	positive
PAT0000007			NorDx - Scarborough	2017-03-20	antibody	negative
PAT0000009			NDX-CORE LAB	2019-04-18	rna	negative
PAT0000011			NDX-CORE LAB	2024-03-11	antibody	positive
PAT0000011			NDX-CORE LAB	2024-03-12	rna	negative
PAT0000012	chronic	[2018]		2018-09-06	antibody	positive
PAT0000012	chronic	[2018]	Affiliated Laboratories	2023-02-08	antibody	positive
PAT0000012	chronic	[2018]	ARUP LABORATORIES	2023-03-07	genotype	1a or 1b
PAT0000012	chronic	[2018]	Affiliated Laboratories	2023-03-07	rna	no_result

Lab Number	1	1	1	1	2	2	2	2
patient_id	test_date	test_facility	test_result	test_type	test_date	test_facility	test_result	test_type
PAT0000001	2017-11-14	Affiliated Laboratories	negative	antibody				
PAT0000004	2018-04-10		positive	antibody				
PAT0000007	2017-03-20	NorDx - Scarborough	negative	antibody				
PAT0000009	2019-04-18	NDX-CORE LAB	negative	rna				
PAT0000011	2024-03-11	NDX-CORE LAB	positive	antibody	2024-03-12	NDX-CORE LAB	negative	rna
PAT0000012	2018-09-06		positive	antibody	2023-02-08	Affiliated Laboratories	positive	antibody
PAT0000016	2022-09-25	NDX-CORE LAB	negative	rna				
PAT0000017	2021-01-11		positive	antibody	2021-01-18	Affiliated Laboratories	positive	antibody
PAT0000018	2023-05-31	MAINEGENERAL MEDICAL CENTER	positive	antibody	2023-05-31	MAINEGENERAL MEDICAL CENTER	negative	rna

WIDEFORM


Count of Tests by Year
antibody_tests
ma_tests
genotype_tests
1750
1500
1250
1000
750
500
250
0
Count
2016
2018
2020
2022
2024
year


Ratio of Unique Positive RNA Tests to Population
0.0025
0.0020
0.0015
0.0010
0.0005
0.0000
2015
2017
2019
2021
2023
Year


Ratio of Unique Positive RNA Tests to Population
0.0014
0.0012
0.0010
0.0008
0.0006
0.0004
2016
2018
2020
2022
2024
Year

DELIVERABLE 11:

ANALYSIS OF THE CLEARANCE CASCADE

test_facility	count	share	neg_share	posneg_share
NDX-CORE LAB	19615	52.058	70.964	66.431
Affiliated Laboratories Inc	5406	14.348	0.65	0.642
LABCORP	4355	11.558	16.451	15.889
QUEST NEW ENGLAND	2041	5.417	0.053	0.161
MAYO CLINIC DEPT. OF LAB MED AND PATHOLOGY	1175	3.118	0.544	4.289
QUEST CHANTILLY	1026	2.723	4.037	2.493
MAINEGENERAL MEDICAL CENTER	871	2.312	3.304	2.11
ARUP LABORATORIES	706	1.874	0.006	2.369
ST MARY'S REGIONAL MEDICAL CENTER	457	1.213		0.916
Quest Diagnostic Nichols Institute	320	0.849	0.089	0.402

NEXT STEPS

1. Pull in death statistics.

2. Preform ad hoc analysis on testing patterns.

3. Write a peer reviewed paper on our findings related to the clearance cascade.

KEY TAKEAWAYS

- **C's**
- **C**onfidentiality
- **C**leaning
- **C**ompletion
- **C**learance

EMAIL:
BROBST.J@NORTHEASTERN.EDU

GITHUB:
BROBST-J

LINKEDIN