



COVID-19 in primary care EMR data

Creating a COVID-19
cohort using pan-
CPCSSN data

July 2021

COVID-19 Lab Data

Of the 113,804 COVID lab reports (n=83,507 patients) in the dataset there were:

- 2,480 'Positive'
- 101,265 'Negative'
- 510 'Inconclusive'
- 4,136 'Cancelled'
- 5,413 no result found

Note: this is for patients inclusive of Jan.1-
Dec. 31, 2020

Lab Positive Cohort

2,360 Positive COVID19 Labs

2,360 patients with (≥ 1) Positive COVID19 lab

108 patients with ≥ 2 Positive COVID19 labs

89.8% had < 21 days between positive lab tests

	n	%
Sex		
Female	1318	55.99
Male	1036	44.01
Age		
0-18	316	13.57
19-49	1124	48.26
50-64	495	21.25
65+	394	16.92

no recorded sex, n=6

no recorded birthyear, n=31

COVID-19 in Health Conditions Table

There were 820 COVID “diagnoses” in the health conditions table

The original text associated with these diagnoses:

- ~7% were blank.
- ~17% say COVID in some form (SARS-associated coronavirus, covid19, coronavirus etc).
- 30% explicitly indicate ‘COVID Positive’.
- 43% are unique strings that need further analysis.

Additional 374 (31%) have a positive COVID-19 lab record.

Verdict: High potential to include these patients in COVID19 positive cohort after some validation and further analysis of text strings.

	DiagnosisText_orig	Frequency Count	Percent of Total Frequency
1	SARS-associated coronavirus	126	11.371841155
2	Infectious Disease (COVID-19 Confirmed Case)	115	10.379061372
3		86	.
4	Coronavirus	46	4.1516245487
5	Confirmed COVID-19	43	3.880866426
6	COVID-19 ASSOC CORONAVIRUS	28	2.5270758123
7	COVID-19	19	1.714801444
8	COVID Positive	15	1.3537906137
9	COVID-19 Positive	14	1.2635379061
10	Covid +ve Dec/20	13	1.1732851986
11	COVID	11	0.9927797834
12	COVID19 POS (Nov 2020) (Concerns)	11	0.9927797834
13	+ve COVID (Concerns)	10	0.9025270758
14	COVID +	10	0.9025270758
15	COVID+	10	0.9025270758
16	COVID19 POS (Dec 2020) (Concerns)	10	0.9025270758
17	Covid Positive 2020.	9	0.8122743682
18	*covid-19	8	0.7220216606
19	+ve covid (Concerns)	7	0.6317689531
20	SARS-ASSOCIATED CORONAVIRUS	7	0.6317689531
21	COVID Dec 2020	5	0.4512635379
22	COVID19	5	0.4512635379
23	CoVID-19 pos/	5	0.4512635379
24	COVID-19 (Oct 2020)	4	0.3610108303
25	COVID-19 virus DETECTED by real-time PCR.	4	0.3610108303
26	Coronavirus SARS-associated	4	0.3610108303
27	Covid positive	4	0.3610108303
28	Covid19 POSITIVE May 2020	4	0.3610108303
29	My patient since MAY 2015. RSRS chart reviewed. /COVID pos NOV/20 while in hospital /***** MEDICAL *****/--CARDIO--- /0000: HYPERTENSION /2013 OCT: ATRIAL FIBRILLATION; started on Pradaxa BID /- was never on beta blocker /--ENDO--- /2001: OSTEOPOROSI	4	0.3610108303
30	- Osteoarthritis - bilateral shoulder replacements, bilateral hip replacements, right total knee replacement. /- Cataract surgery. /- Bell's palsy. /- COVID19 positive 2020, hospitalized/	3	0.2707581227
31	--- (Covid positive)	3	0.2707581227
32	2020 COVID infection	3	0.2707581227
33	COVID + 2020	3	0.2707581227
34	COVID 19 POS (May 2020) (Concerns)	3	0.2707581227
35	COVID 19: Positive/	3	0.2707581227
36	COVID Positive Test Result	2	0.1804762454

COVID-19 in Encounter Diagnosis Table

26,154 observations of COVID-19 (n=13,716 patients)

Original Text associated with these observations:

- ~65% say 'SARS-associated coronavirus, COVID, COVID19, coronavirus, COVID 19.'
- About 7% are blank
- 5-10% have a standardized ICD-9 name, plus another ICD-9 standard diagnosis, and so may or may not be related to a positive covid case
 - ex. 'Viral Condition in Conditions Classified Elsewhere, and of Un, COVID19, Depression'.

COVID-19 in Billing Table

118,155 COVID-19 billings (n=55,412 patients)

The original text associated with these billings:

- 88% were blank
- 5% list an ICD-9 code (O79.82) or 'Covid19'.

COVID-19 Case Definitions

COVID-19 Positive

Inclusion Criteria

- A COVID-19 lab result of 'Positive' and/or COVID in health condition table and/or COVID in encounter diagnosis table.

Exclusion Criteria

- None

COVID-19 Negative

Inclusion Criteria

- None

Exclusion Criteria

- A COVID-19 lab result of 'Positive'

Options for COVID-19 cohort

- Continue to refine COVID-19+ cohort:
 - More work to pull data out of PDFs (e.g., important for MB CPCSSN data)
- Develop other tables that will be useful for examining effects of long-haul COVID (examples):
 - Exam table (only contains physical measurements currently)
 - Referral table (develop ways to examine which specialists COVID + patients were referred to)

COVID-19 cohort (preliminary)

- n=3,174
- 56% female
- 35% 0-18 years; 24% 19-39 years; 28% 40-64 years; 13% 65+
- Most cases in our cohort from Alberta and Ontario, followed by British Columbia
 - Ability to extract data from PDFs (e.g, MB)
 - Provincial CPCSSN node (e.g. lots of COVID in Quebec, but smaller network)