

Phase 1 Proposal: GOALS

<https://github.com/GriffinWeber/covid19i2b2>

- Get minimal amount of useful information about COVID-19 as fast as possible
- Avoid IRBs if possible
- Avoid IT if possible (no firewalls, new VMs, software installs, etc.)
- Avoid "data harmonization" (ontology changes)

Phase 1 Proposal: STEPS

<https://github.com/GriffinWeber/covid19i2b2>

- Sites create four CSV files containing only aggregate counts (daily counts, demographics, labs, and diagnoses)
- Sites create these files locally in whatever way is easiest for them (via the i2b2 UI, database SQL script, etc.)
- Sites send their files to X, so they can be combined into four master CSV files
- Rapidly publish the results

DailyCounts-SiteID.csv

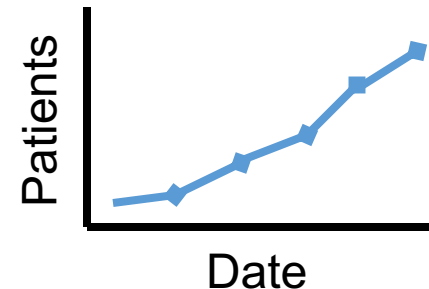
Fields: siteid, date, new_positive_cases, patients_in_icu, new_deaths

Notes:

1. One row per date (2020-03-15, 2020-03-16, etc.)
2. Site is a unique identifier for your institution (e.g., "BIDMC")
3. new_positive_cases, patients_in_icu, new_deaths are # of distinct patients
4. If a patient has multiple positive test results, use the earliest date
5. Set patients_in_icu = -2 if you do not have ICU data
6. Set new_deaths = -2 if you do not have death data
7. Obfuscate small counts with "-1" as required by your institution

Examples: (DailyCounts-BIDMC.csv)

```
BIDMC, 2020-03-20, -1, 0, -2
BIDMC, 2020-03-21, 6, -1, -2
BIDMC, 2020-03-22, 14, 5, -2
BIDMC, 2020-03-23, 26, 10, -2
```



Demographics-SiteID.csv

Fields: siteid, sex, total_patients, age_0to2, age_3to5, age_6to11, age_12to17, age_18to25, age_26to49, age_50to69, age_70to79, age_80plus

Notes:

1. One row per sex (values = "Male", "Female", "Other", "All")
2. Count anyone who is not male or female as "Other"
3. Include all patients in the "All" row
4. The total_patients value in the sex=All row should be everyone
5. Obfuscate small counts with "-1" as required by your institution

Examples: (Demographics-BIDMC.csv)

BIDMC, Male, 26, 0, 0, 0, 0, 0, 0, 6, 8, 12

BIDMC, Female, 24, 0, 0, 0, 0, 0, 0, -1, 10, 11

BIDMC, Other, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0

BIDMC, All, 50, 0, 0, 0, 0, 0, 0, 9, 18, 23

Labs-SiteID.csv

Important lab tests from Gabe's list

TEST	LOINC
white blood cell count (Leukocytes)	6690-2
neutrophil count	751-8
lymphocyte count	731-0
albumin	1751-7
lactate dehydrogenase (LDH)	2532-0
alanine aminotransferase (ALT)	1742-6
aspartate aminotransferase (AST)	1920-8
total bilirubin	1975-2
creatinine	2160-0
cardiac troponin	49563-0
D-dimer	7799-0
prothrombin time (PT)	5902-2
procalcitonin	33959-8
C-reactive protein (CRP)	1988-5

Labs-SiteID.csv

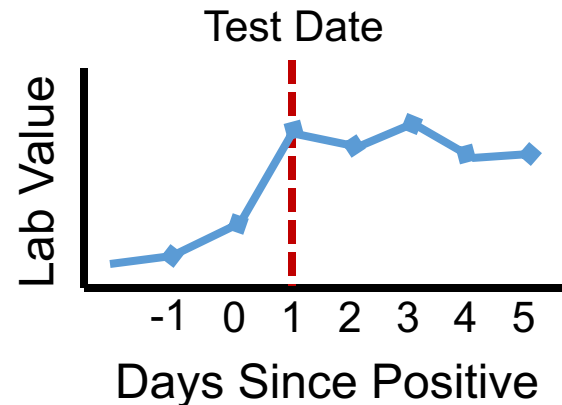
Fields: siteid, loinc, days_since_positive, num_patients, mean_value, stdev_value

Notes:

1. One row per loinc and days_since_positive
2. days_since_positive = 1 on the date the patient has a positive COVID test result
3. Start the table at days_since_positive = -6 (seven days before the positive test)
4. Go for as many days as you have data: days_since_positive = 5, 6, 7, ...
5. Map your local loinc codes to the loinc code in Gabe's list
6. Only use Gabe's loinc codes in this list, not your local codes
7. Obfuscate small counts with "-1" as required by your institution

Examples: (Labs-BIDMC.csv)

```
BIDMC, 1920-8, -2, -1, 27.0, -1
BIDMC, 1920-8, -1, 4, 41.2, 23.567
BIDMC, 1920-8, 0, 4, 47.1, 17.876
BIDMC, 1920-8, 1, 38, 59.2, 40.152
BIDMC, 1920-8, 2, 35, 52.2, 35.432
BIDMC, 1920-8, 3, 12, 40.4, 25.678
BIDMC, 1975-2, 1, 37, 39.5, 20.100
```



Diagnoses-SiteID.csv

Fields: siteid, icd_code, icd_version, num_patients

Notes:

1. One row per ICD diagnosis code
2. All diagnoses the patients have starting seven days before the positive test
3. icd_version = "9" or "10"
4. Obfuscate small counts with "-1" as required by your institution

Examples: (Diagnoses-BIDMC.csv)

BIDMC, B97.29, 10, 25

BIDMC, J12.89, 10, 19

BIDMC, R03.0, 10, 15

BIDMC, U07.1, 10, 13

BIDMC, 123, 9, -1