Study Information

Title: Aggregate count statistics on COVID-19 patients using Clinical Query 2 (CQ2)

Protocol #: 2020D000290

Application Type: Determination of Human Subject Research

PI: Weber, Griffin

PI Dept: Medicine

PI Div: Med: IMBIO

Author: Weber, Griffin

General Information

1. Is the activity an investigation?

Yes

2. The activity involves the following:

-

3. Is the investigation systematic?

Yes

4. Is the systematic investigation designed to develop or contribute knowledge?

Yes

5. Is the knowledge generalizable?

Yes

6. Will you gather data about living individuals through intervention or interaction?

No

7. Will you gather data about living individuals that is private?

Yes

8. Will you gather data about individuals that is identifiable?

No

9. Does the activity involve the collection or analysis of newborn dry blood spots, either identifiable or de-identified?

No

10. Does the activity involve an individual as a recipient of any test articles (i.e drug, medical device) or as a control?

No

11. Does the activity involve an individual on whose specimen a medical device will be used?

No

12. Will tissue/specimens be used to test the effectiveness of a medical device (including in vitro diagnostic devices)?

No

13. Will the information obtained be submitted to the FDA?

No

14. Describe the purpose, specific aims, and/or objectives:

I manage the Clinical Query 2 (CQ2) website at BIDMC. It is powered by an open source tool called i2b2, which is used at 200+ institutions worldwide. Dozens of these i2b2 sites have launched a project to use their databases to pull together information about COVID-19 patients from around the world. Each site is asked to provide four tables of aggregate counts about their COVID-19 patients: (1) the number of new cases per day, (2) the sex and age breakdown of the patients, (3) the average values for a set of about 15 lab tests, and (4) the top diagnoses/comorbidities of the patients. These counts will be merged with the goal of publishing a paper that characterizes a large fraction of the global COVID-19 positive population.

15. Describe target population:

Patients who have had a positive COVID-19 test at BIDMC.

16. Describe procedures used to gather information (e.g., communication or interpersonal contact with individuals, manipulation of individuals, manipulation of individual's environment, or physical procedures). Indicate if these procedures would be conducted as part of standard of care, regardless of the research:

I will use Clinical Query 2 (CQ2) to generate tables of aggregate counts (e.g., number of COVID-19 patients at BIDMC who have diabetes). Age groups are in bins (e.g., 50-69 years old), with the top bin "80+ years old". Any count less than 10 will be reported as "less than 10". Each participating site will place their tables in a central repository for analysis. The number of participating sites are continually increases as institutions join this effort. Note that CQ2 already allows researchers to run aggregate count queries without separate IRB approval. However, I'm not sure if there is added sensitivity right now about sharing aggregate information about COVID-19 patients at BIDMC with other institutions. Each site is generating files with aggregate counts. Those files will be merged and made available to the public for research. I cannot control how the data will then be used. The data are not coded. The files I would be creating only contain aggregate counts describing the population of patients at BIDMC with COVID-19. The files do not contain data on individual patients. Any counts corresponding to fewer than 10 patients will be reported as "less than 10 patients".

17. Describe how you will use the information collected from this activity (e.g. data will be used to directly improve clinical care at the medical center):

We will merge the counts from all participating institutions to generate histograms describing the global COVID-19 patient population and publish these results.

18. Describe the type of data/samples which will be collected and/or analyzed. If data/samples were collected for other purposes, describe how the data/samples originally gathered from individuals (e.g., obtained as part of another IRB approved protocol at this institution/another institution or as part of routine clinical practice):

The data will come from Clinical Query 2, which obtains its data from the BIDMC clinical data repository. Those data were collected through the routine clinical practice. The data types include date of COVID-19

test, sex, age group, whether the patient was in the ICU, whether the patient died, lab values, and diagnoses. Though, these will all be obtained as aggregate counts with low-count masking. This study will not collect data on individual patients.

19. Were data/samples obtained as part of a research study?

No

20. Did participants provide consent?

21. Did the consent obtained permit this use of data/sample? (Please consider whether the consent form allowed for the sharing and future use of data/samples or whether the proposed activity is consistent with the participant's understanding of how their data/samples could be used.)

22. Address if the data/samples can be directly or indirectly associated/linked with individual identities. If data/samples are coded address any agreement, policy or legal restriction which would prevent the provider of the data/samples from releasing the key to the code:

No. Only aggregate counts will be generated. Small counts less than 10 will be masked.

23. Can others directly or indirectly associate/link the collected information with individual identities?

No. Only aggregate counts will be generated. Small counts less than 10 will be masked.

24. Do you plan to submit data to the NIH Genome-Wide Association Study (GWAS) data repository?(Check with NIH)

No

Internal Staff:

Name	Role	Mail	Consenting	MD	Credentialed
Weber, Griffin M	PI	Yes	Yes	No	

PI Statement

PI accepted agreement on: 03/27/2020