COVID-19 Testing Description Document

Peter Higgins

5/15/2021

COVID-19 Testing Dataset Introduction

Source: This data set is from Amrom E. Obstfeld, who de-identified data on COVID-19 testing during the year 2020 at the Children's Hospital of Pennsylvania (aka CHOP).

This data set contains data concerning testing for SARS-CoV2 via PCR as well as associated metadata. The data has been anonymized, time-shifted, and permuted.

Abstract   This data set contains data on SARS-CoV2 testing at a single hospital on

Background   The COVID-19 pandemic in 2020 required rapid development and deployment of a novel PCR test, and application to many outpatients with symptoms, and all inpatients upon admission to the emergency room or the hospital. These anonymized data (with fake subject_ids, names, ages, and genders) provide a snapshot of the first ~100 days of testing at a single pediatric hospital, with over 15,000 tests performed on patients in a variety of settings. The vast majority were a single type of test. These tests occurred in 88 named clinics (clinic_name) and in 5 demographic groups (demo_group). Roughly half of the tests were performed at drive-through sites (drive_thru_ind).

The test result (result) can be either negative, positive, or invalid. The ct_result provides the cycle number at which the positive threshold was reached during PCR. Lower integer values for ct_result occur with a higher viral load (higher number of viral particles) in the test sample (deep nasal swab).

The test order can have been placed as a one-off order, or within an order set. The payor_group for an ordered test can be one of 7 categories (includes insurance type and self-pay). The patients are categorized into 9 patient classes, including inpatient, emergency, outpatient, etc. The col_rec_tat variable records the time elapsed (in hours) between sample collection and receipt in the lab. The rec___ver_tat variable records the time elapsed (in hours) between sample receipt in the lab and result verification/posting.

Study Design   Prospective Cross-Sectional Cohort

Subjects & Variables   15,524 participants were tested, from day 4 to day 107 of the COVID-19 pandemic.

N = 15,524 subjects
17 variables

Potential Data Analysis   There is a dichotomous outcome variable (result) and a continuous outcome variable (ct_result).

You can also look at the specimen processing time to receipt, and from receipt to verification as an outcome. Did processing times improve over the first 100 days?

You can look at trends over time by pan_day.

There are a lot of categorical variables that could be associated with the outcomes, including patient_class, clinic_name, drive_thru_ind, demo_group, payor_group, gender, and orderset.

These categorical variables could also be interesting to explore with correlations and mosaic plots.