Assessment of the effectiveness of required weekly COVID-19 surveillance antigen testing at a university

Christopher W. Ryan, MD, MS, MSPH*

Clinical Associate Professor of Family Medicine, SUNY Upstate Medical University Binghamton Clinical Campus, PO Box 778, Johnson City, NY, 13790, US

Medical Director, Broome County Health Department, 225 Front Street, Binghamton, NY, 13905, US

2023-11-22 16:00:14

Word count: approximately 2991, excluding title page, abstract, required statements, and figure/table placement markers

Number of figures: 3

Number of tables: 3

1 Keywords

COVID-19, surveillance testing, pandemic control

2 Three-question summary box

What is the current understanding of this subject? Assessments of long-term operational effectiveness of COVID-19 “surveillance testing” have not been published.

What does this report add to the literature? During the 2020-2021 academic year at one university, people with COVID-19 detected via compulsory weekly surveillance antigen testing were equally likely to be symptomatic at time of detection, and for just as long, as similar-aged people detected via testing venues in the community.

What are the implications for public health practice? Surveillance testing programs during the pandemic consumed a large amount of time, money, and effort. In future respiratory pandemics, resources might be better devoted to other mitigation measures.

*Corresponding author. Address: SUNY Upstate Medical University Binghamton Clinical Campus, PO Box 778, Johnson City, NY, 13790, US. Phone: 607-772-3526. Email: cryan@binghamton.edu

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract

Objectives
To mitigate the COVID-19 pandemic, many institutions implemented a regimen of periodic required testing, irrespective of symptoms. The effectiveness of this “surveillance testing” requires assessment.

Methods
I fit a zero-inflated negative binomial model to COVID-19 testing and case investigation data between 1 November 2020 and 15 May 2021, from young adult subjects in one community. I compared the duration of symptoms at time of specimen collection in those diagnosed via (1) surveillance testing at a university, (2) the same university’s student health services, and (3) all other testing venues.

Results
The data comprised 2926 records: 393 from surveillance testing, 493 from student health service, and 2040 from other venues. About 65% of people with COVID-19 detected via surveillance testing were already symptomatic at time of specimen collection.

Predicted mean duration of pre-testing symptoms was 1.7 days (95% CI 1.59 to 1.84) for the community, 1.81 days (95% CI 1.62 to 1.99) for surveillance, and 2 days (95% CI 1.83 to 2.16) for student health service. The modelled “inflated” proportions of asymptomatic subjects from the surveillance stream and the other/community stream were comparable (odds ratio 0.95, p = 0.7709). Comparing surveillance testing with the student health service, the proportion of “excess” zero symptom durations was significantly higher in the former (Chi-square = 12.08, p = 0.0005).

Conclusions
Surveillance testing at a university detected 393 people with COVID-19, but no earlier in their trajectory than similar-aged people detected in the broader community. This casts some doubt on the public health value of such programs, which tend to be labor-intensive and expensive.
Introduction

Hoping to control the spread of COVID-19 attributable to asymptomatic and presymptomatic infections, many institutions implemented a regimen of periodic required testing, irrespective of symptoms, for people notionally under their jurisdiction—often called “surveillance testing.” Among them were many colleges and universities. After abruptly halting in-person, on-campus instruction in March 2020, surveillance testing was widely considered important for a safe return to in-person collegiate education in fall 2020 and was widely adopted.1–12

Early mathematical modeling studies attempted to predict the effects of surveillance testing,2,13–19 and several pilot or short-duration studies of implementation were also conducted.4,6,7 As these programs are costly and labor-intensive, it is now important to assess their effects from a longer-term perspective. The present study concerns a single university (“University”) where, serendipitously, surveillance test results could be confidently distinguished from those generated by the student health service and by other testing venues in the wider community. The research question is whether University’s COVID-19 surveillance testing program detected COVID-19 infections significantly earlier in their clinical evolution, compared to University’s own student health service, and to the broader community where the the like-aged population was not generally subject to surveillance testing.

Methods

The analysis involved all positive COVID-19 test results and the subsequent case investigation reports among subjects aged 17.5 to 27 years, inclusive, in a county of approximately 198,000 population, between 1 November 2020 and 15 May 2021. For most of that interval, wild-type SARS-CoV-2 was the predominant circulating strain; by the end of the study period, the alpha strain accounted for about half of sequenced specimens worldwide.20 The county encompasses University, which enrolls approximately 18,000 students, undergraduate and graduate. Generally healthy people of study age became eligible for COVID-19 vaccination toward the end of the study period.21 The Institutional Review Board of SUNY Upstate Medical University determined that this project comprised public health surveillance and did not meet the definition of human subjects research (determination 1933720-1).

During the study period, the local health department was obligated to investigate all occurrences of COVID-19 via structured telephone interview, which included the presence and duration of symptoms. All those with COVID-19 received official isolation orders and instructions. Exposed contacts disclosed by the patients were also interviewed and quarantined. The health department maintained continuous, close coordination with University.

Weekly, students at University were required to have a BINAXNow COVID-19 antigen test,22 conducted at the student union (“surveillance testing”). This began in late August, but with a
different testing product; that period is excluded from this analysis. Students with symptoms of COVID-19 were not to attend their next scheduled surveillance test but rather seek clinical evaluation promptly. For many, this would be via University’s student health service. Nevertheless, many students disclosed symptoms at the surveillance site, and if their test was positive, they were automatically considered by the local health department to have COVID-19. Subjects with a positive antigen test but who denied symptoms at the surveillance site were sent, immediately or the next day, to the student health center for a point-of-care Cue Health isothermal nucleic acid amplification test (NAAT). If positive on NAAT, they too were considered to have COVID-19 and investigated as above; if negative (discordant results), they were not considered infected and no further action was taken.

Employees of University were also subject to weekly required surveillance testing and are subsumed here under the term “students.” There is no reliable, automated way to distinguish laboratory results pertaining to students from those pertaining to employees.

Ill students presenting to the student health center were tested as clinically indicated, with one or more of an antigen test, a point-of-care NAAT, or a specimen sent to a reference laboratory for a NAAT (“shs testing”). Employees were not generally eligible for clinical care at the student health center, except for the follow-up NAAT for those asymptomatic but with a positive surveillance test.

Students could also avail themselves of other testing venues: one state-operated and two hospital-operated drive-through NAAT sites, and a mobile antigen testing service operated by the local health department. These venues are here aggregated under the category “community/other testing.”

The results of all COVID-19 tests were reported into the state’s Electronic Clinical Laboratory Reporting System (ECLRS) and thence to the local health department where the person resides. College students attending University were considered residents of the study county, although in practice some results were mis-routed to the health department of their family home; those were not accessible for this study. I used all COVID-19 test results present in the study county’s ECLRS database as of 14 February 2023.

Each result record contained a Clinical Laboratory Improvement Act (CLIA) number that uniquely identified the laboratory that analyzed the specimen. By convenient coincidence, University used different CLIA numbers for its surveillance testing and its student health service testing, a decision for made for internal, operational reasons. All non-University CLIA numbers were categorized here as “community/other.”

All positive COVID-19 laboratory reports were automatically transmitted into a commercial database called CommCare (Dimagi, Inc.), which was provided by the state for this one purpose. CommCare served as the data system for guiding the investigator’s interview and for recording the findings. I used all records of investigations present in CommCare as of 27 December 2022.

The lab result and investigation datasets were each separately filtered to eliminate those with a variety of anomalies, including:

- specimen collection date prior to February 2020 or as dates that had not yet occurred
• duplicate entries: results from the same specimen from the same person on the same day
• investigations marked as entered in error, duplicate, or outside the local health department’s jurisdiction

Using the fastLink package\textsuperscript{24} in R,\textsuperscript{25} I then matched the two data sources on the basis of last name, first name, day of the month of the birthdate (1-31), and age in fractional years on the date the specimen was collected. Records were considered successfully matched if the posterior probability of an accurate match was greater than 85%.

I then filtered the matched records to those meeting the following criteria:

• specimen collection date (SCD) was within the study period
• the laboratory stream was recorded unambiguously
• the presence or absence of symptoms was recorded unambiguously
• if symptoms were present, the symptom onset date (SOD) was within seven days either side of the SCD
• the SCD in ECLRS was within four days either side of the SCD in CommCare

This yielded the final, analytical dataset.

Using the countreg package\textsuperscript{26-28} in R, I fit a zero-inflated negative binomial model, with the number of pre-test symptomatic days as the response and the testing stream by which the infection was detected as the predictor of interest. The “other/community” stream served as the baseline category. When symptoms began after specimen collection or were entirely absent, the response variable was set to zero. All students with COVID-19 first detected via a positive surveillance antigen test were attributed to that testing stream, irrespective of any immediate follow-up NAAT testing at student health service. I controlled for overall community incidence with a kernel smoothed estimate of the proportion of tests across the entire county (all three testing streams) that were positive each week.

Briefly, a zero-inflated negative binomial model accommodates a mixture distribution where subjects with zero counts (here, zero pre-collection symptomatic days) can arise in two different ways: via a count process with overdispersion, or via a dichotomous process in which a count is either zero or non-zero. A zero-inflated model can accommodate a larger number of subjects with zero counts than can be well-fit by typical Poisson or negative binomial models—a so-called “excess” of zeros.

I assessed goodness of model fit with graphical methods in the countreg package. I compared coefficient estimates on the lab streams with Wald tests. Lastly, I conducted 5000 simulations from the model, at a typical age of 22 years and at two different levels of overall community incidence, generating point estimates and confidence intervals for stream-specific probability distributions of the number of pre-test symptomatic days.
Results

The county-wide percent of tests each week that were positive ranged from 1.14% to 12%, with a median of 5.52%.

The query of the state laboratory database yielded 2292320 records. Excluding those with anomalous dates left 1384786. Limiting those to subjects aged 17.5 to 27 years left 340051, of which 12887 were positive.

The query of the Commcare case investigation database yielded 97543 records. Excluding records marked as duplicate, erroneous, or out-of-jurisdiction left 76615. Of those, 8935 were between 17.5 and 27 years of age, inclusive. All non-erroneous records in Commcare by definition represent patients with positive test results.

Record linkage yielded 7870 matched records. After the exclusions itemized above and summarized in Table 1, 2926 matched records remained. These comprise the final, analytical dataset to which models were fit.

[Table 1 about here.]

The testing streams from which the 2926 records in the analytical dataset emerged are summarized in Table 2.

[Table 2 about here.]

The distribution of elapsed days between SOD (as determined from the case investigation interview) and SCD (as determined from the state laboratory result database) is shown in Figure 1.

[Figure 1 about here.]

Modeling results

The fitted model is shown in Table 3. The coefficient estimates and standard errors in Table 3 are on the scale of the linear predictor, so positive values imply an increase in the probability of an “excess” zero (top half of Table 3) or in the number of pre-collection symptomatic days discounting the excess zeros (bottom half of Table 3). Model fit seems adequate, as explored in more detail in the Supplemental Materials.

[Table 3 about here.]

Part 3a of Table 3 models the “excess” portion of subjects asymptomatic at time of specimen collection. These “excess zeros” are those poorly accounted for by the negative binomial count process modeled in subtable 3b. Subtable 3a suggests that the “excess” proportion of asymptomatic subjects presenting to the student health service (shs) is significantly lower than that in the other/community
setting (odds ratio 0.26, p = 0.0002.) This is perhaps not surprising. On the other hand, the “excess” proportions of asymptomatic subjects from the surveillance stream and the other/community stream were comparable (odds ratio 0.95, p = 0.7709). Comparing surveillance testing with the student health service, the proportion of “excess” zero symptom durations was significantly higher in the former (Chi-square = 12.08, p = 0.0005)—also not surprising.

Part 3b of Table 3 models the negative binomial count process for the number of symptomatic days prior to specimen collection (which can itself also yield zeros). Subtable 3b suggests that, discounting the “excess” asymptomatic infections (that were accounted for by the “excess zero” component in 3a), the mean duration of pre-collection symptoms was not significantly different in either the surveillance stream or the student health stream, compared to the baseline other/community stream (odds ratios 0.96, p = 0.4265 and 1.06, p = 0.3374, respectively.)

Combining the two components of the model, we can predict two informative things:

1. The probability that a person detected as having COVID-19 would have any particular number of pre-test symptomatic days. This is shown in Figure 2, for each lab stream, at each of two illustrative overall community incidence levels: “low” (the observed lower quartile of 3.9%) and “high” (the observed upper quartile of 8.75%). At either incidence level, the probability profiles of University’s surveillance testing stream and that from the wider community are nearly identical. On the other hand, people identified via the student health service stream were less likely to have been asymptomatic.

2. The mean duration of symptoms prior to specimen collection. This is shown in Figure 3 for each lab stream, again at “low” and “high” community incidence. Within each graph, there is much overlap between predicted mean pre-collection symptom durations from each of the three testing streams. The scale of the horizontal axis should also be noted: what differences there may be in the predicted means are on the order of a few tenths of a day, i.e. 6-12 hours.

[Figure 2 about here.]

[Figure 3 about here.]

Discussion

At a three-year remove from the onset of the pandemic, the effects of surveillance testing programs can now be assessed with some perspective. The serendipitous use of a separate CLIA number for surveillance testing at University enabled the present analysis to compare surveillance testing with (1) University’s student health service and (2) testing venues in the wider community used by college-aged people.

The interpretation of the present findings is nuanced. About 35% of students with COVID-19 detected via University’s surveillance regimen had no symptoms at the time of specimen collection.
This compares with about 20% of those detected via the student health service. This could be considered a success for disease control, in that it enabled early isolation of those asymptomatic but infected students, and identification and quarantine of their contacts.

Nevertheless, that leaves about 65% of surveillance-detected students already symptomatic at the time of detection, despite instructions that symptomatic students should not wait for their scheduled surveillance test. The concern is that those students could potentially transmit infection while waiting. On the other hand, a more encouraging interpretation is that those infections might have gone entirely undetected but for University’s surveillance regimen.

The fundamental question in the present analysis is whether infections with COVID-19 were detected earlier in their trajectory among a population subject to surveillance testing compared to a like-aged population that generally was not. This analysis provided no evidence that they were. Those identified via University’s surveillance testing were as likely to be symptomatic at the time of specimen collection as similar-aged subjects arising from various testing venues in the community, and model-predicted mean duration of pre-collection symptoms differed only trivially between the three different testing streams. These findings raise important questions about the utility of resource-intensive surveillance testing programs in mitigating respiratory pandemics.

Public health interventions take place in a larger behavioral, social, ecological, economic, occupational, and academic context. These contextual factors may influence the effects of such programs, perhaps by influencing care-seeking (or care-avoiding) behavior—in potentially paradoxical ways.

In any one-month period, approximately 75% of adults will develop some sort of symptom, yet only one-third of those will seek medical care. Symptoms are by definition subjective and not verifiable externally. They also have meaning for the person experiencing them, and that meaning can change with time and events. Minor symptoms initially self-interpreted as unworthy of mention or action may later be reinterpreted, especially after being presented with a test result.

People with COVID-19 could not attend class, work, or other group activities. Neither could the people who were identified as having been close around them. So besides the medical ramifications, the decision to seek prompt clinical testing, as opposed to awaiting a scheduled surveillance test, entailed opportunity costs, and not just for the person infected.

“Risk homeostasis,” still a controversial concept, posits that people might become less careful when they perceive their individual risk to be reduced by external mitigation measures. This could manifest as a delay in seeking care if one felt reassured by the results of a recent (negative) surveillance test or by the sure knowledge of another test coming up soon. Several early mixed-methods studies in university settings suggested that some subjects receiving a negative result on a recent test indeed felt reassured and thus willing to liberalize their social interactions.
Limitations

This analysis was based on data from weekly antigen-based surveillance testing at a single university and may not be generalizable. Replication in other settings, where other procedures were used, would be valuable.

Release from isolation depended mainly on the presence and duration of symptoms, which were measured by telephone interview after diagnosis. It is possible that some subjects fabricated symptoms, or provided spuriously early symptom onset dates, to expedite their release. This caveat would apply to both students (subject to surveillance testing) and non-students (not generally subject to surveillance testing), so any bias may not have been differential.

A small number of county residents unaffiliated with University were subject to surveillance testing, mainly the employees of skilled nursing facilities. Test results from these settings were included here in the “other/community” stream, since they could not easily be otherwise distinguished. Compared to University, these surveillance testing programs were relatively low-volume, with a still smaller fraction of participants expected to be of study age.

No subjects were interviewed specifically for the purposes of this study. The case investigation records did not record any subject’s thoughts, feelings, or intentions around getting or not getting tested through any particular stream. That would be fertile ground for further study.

Conclusions

People with COVID-19 detected via a compulsory weekly COVID-19 surveillance antigen testing program at a university were equally likely to be symptomatic at time of detection, and for just as long, as similar-aged people whose COVID-19 was detected via testing venues in the broader community. This casts some doubt on the public health value of such programs, which are generally expensive and labor-intensive. Given the mean duration of symptoms found here, a twice- or thrice-weekly testing frequency may have been necessary to achieve earlier detection, with a concomitant increase in the effort and resources required. Alternatively, in future respiratory pandemics, resources might be better devoted to other mitigation measures.

3 Funding sources

The author received no financial support for the research, authorship, and/or publication of this article.

4 Conflicts of interest

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.
References


Symptom onset date compared with specimen collection date

Red line at zero indicates specimen collection. Negative values mean symptoms preceded collection; conversely, positive values mean onset of symptoms followed specimen collection.

Figure 1: Distribution of times between onset of symptoms (from the case investigation interview) and the specimen collection date (from the state laboratory result database) in the three different testing streams, as a proportion of all results from the respective streams. Negative values mean onset of symptoms preceded specimen collection; conversely, positive values mean onset of symptoms followed specimen collection.
Figure 2: The model-estimated proportion of subjects with any given number of symptomatic days (from 0 to 7) prior to specimen collection, at each of two levels of community incidence, for each of three laboratory streams. The probability of symptom duration beyond 7 days was very low and is not shown. “shs” = student health service.
Figure 3: 5000 values of mean duration of symptoms simulated from the model, at each of two levels of community incidence, for each of three laboratory streams. Envelopes represent the density of the mean duration values obtained, while the dark bars show the means and their 95% two-sided confidence intervals. “shs” = student health service.
Table 1: Reasons matched records were excluded from further analysis and modeling. SOD = symptom onset date. SCD = specimen collection date. Reasons are not mutually exclusive, i.e. some records may have qualified for exclusion on more than one account. Total number excluded was 4944

<table>
<thead>
<tr>
<th>exclusion criterion</th>
<th>number excluded</th>
</tr>
</thead>
<tbody>
<tr>
<td>more then 4 days discordance on SCD</td>
<td>670</td>
</tr>
<tr>
<td>missing indicator of symptoms</td>
<td>175</td>
</tr>
<tr>
<td>ambiguous lab indicator</td>
<td>88</td>
</tr>
<tr>
<td>more than 7 days between SCD and SOD</td>
<td>662</td>
</tr>
<tr>
<td>missing duration between SCD and SOD</td>
<td>1986</td>
</tr>
<tr>
<td>prior to study period</td>
<td>628</td>
</tr>
<tr>
<td>after study period</td>
<td>3950</td>
</tr>
</tbody>
</table>
Table 2: Sources of positive COVID-19 test results in the final analytical dataset: other = non-university-affiliated settings; shs = university student health service; surveillance = university surveillance testing.

<table>
<thead>
<tr>
<th>testing stream</th>
<th>number</th>
<th>percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>other</td>
<td>2040</td>
<td>70%</td>
</tr>
<tr>
<td>shs</td>
<td>493</td>
<td>17%</td>
</tr>
<tr>
<td>surveillance</td>
<td>393</td>
<td>13%</td>
</tr>
</tbody>
</table>
Table 3: Zero-inflated negative binomial model. The coefficient estimates and standard errors are on the scale of the linear predictor. The “other” lab stream, i.e. the broader community outside University, serves as the baseline category.

| Coefficient               | Estimate | Std. Error | t value | Pr(>|t|) |
|---------------------------|----------|------------|---------|----------|
| intercept                 | -1.18611 | 0.5927     | -2.001  | 0.045    |
| shs lab                   | -1.34612 | 0.3617     | -3.721  | <0.001   |
| surveillance lab          | -0.05549 | 0.1905     | -0.291  | 0.771    |
| smoothed daily incidence  | 2.63214  | 2.1972     | 1.198   | 0.231    |
| age                       | -0.00407 | 0.0261     | -0.156  | 0.876    |

(a) The zero component.

| Coefficient               | Estimate | Std. Error | t value | Pr(>|t|) |
|---------------------------|----------|------------|---------|----------|
| intercept                 | 0.9865   | 0.18755    | 5.260   | <0.001   |
| shs lab                   | -0.0428  | 0.05383    | -0.795  | 0.427    |
| surveillance lab          | 0.0564   | 0.05881    | 0.959   | 0.337    |
| smoothed daily incidence  | -1.0210  | 0.70775    | -1.443  | 0.149    |
| age                       | -0.0031  | 0.00826    | -0.375  | 0.708    |

(b) The count component