Introduction

The safe re-opening of the University at Buffalo (UB) campuses following closure in March 2020 due to the SARS-COV-2 virus represents a challenge for which sound scientific methods and logistic preparations are required. Current planning capabilities are limited due to lack of knowledge of the baseline rates of infection among students, faculty, and staff, particularly in the asymptomatic population in which transmission still occurs. Baseline diagnostic testing and the associated estimation process of the SARS-COV-2 virus prevalence rates is necessary for the modelling of subsequent virus transmission and the development of monitoring goals and plans, subject to constraints, in order to minimize the likelihood of campus closure during the Fall 2020 semester. This document outlines the recommendations for the sampling and estimation of the campus baseline prevalence rates, or mathematically equivalent, the number of cases on campus, among students, faculty, and staff. It should be noted that this plan does not represent a proposed campus-level surveillance program as that would require extensive diagnostic testing performed at repeated intervals throughout the semester, and associated isolation of identified cases. Such an initiative should be carefully planned, including the use of established epidemiologic and biostatistical models so to maximize efficiencies given resource limitations. The testing procedures recommended in this document should be done in conjunction with, not in place of, the promotion of face coverings, hygiene, social distancing, and symptom-monitoring.

Subpopulations and sampling plan

We acknowledge that virus transmission and a given individual’s probability of being SARS-COV-2 virus positive is dependent on a number of factors. Although the majority of the student population may not be within the highest risk categories for either acute infection or severe disease-related morbidity, the cohabitation living arrangements, campus activities, and commonly attended social events represent an environment for accelerated virus transmission. In addition, a portion of the student body will arrive at UB at the semester start from outside areas currently experiencing higher infection rates. Other students, many which would fall into the non-traditional category, would experience a reduced risk of infection as campus activities would be limited, perhaps only to in-person classes where distancing measures have already been implemented. Faculty and staff would likely experience yet a different infection risk profile.

A stratified random sampling plan is propose based on the following three subpopulations of individuals which have differing degrees and types of on-campus in-person interactions:

1) Students that reside in campus dormitories during the Fall 2020 semester.
2) Students that live off-campus within Erie or Niagara County during the Fall 2020 semester.
3) University faculty and staff that live in Erie or Niagara County with some degree of identifiable on-campus in-person interactions during the Fall 2020 semester.

The following individuals should be excluded from this sample:

- Faculty and staff who will not have identifiable consistent on-campus in-person interactions during the Fall 2020 semester in its entirety, i.e., fully remote faculty or staff. These individuals would not directly contribute to disease spread on campus. In addition, those not coming to campus either have medical reasons or may be living remotely and would have difficulty in coming onto campus as they have chosen, for whatever reason, to stay remote.
- Students who live off campus and outside of Erie and Niagara County. Given the burden testing would represent for these students, a high level of noncompliance would be expected.
Appendix A

- Individuals whom are already known to be currently infected with the SARS-COV-2 virus. It should be noted though that data from these individuals will also be useful in campus decision making.
- Athletes from the student population due to testing protocols currently in place. Data from student athletes should be used in conjunction with the data to be gathered as part of the testing procedures described in this document.

Within each of the three stratified subpopulations, a total of approximately 279 individuals (rationale provided below) should be randomly sampled for a total of approximately 837 across the three strata. Random sampling should be implemented using a unique identifier, e.g., person number. Individuals selected for random testing should be contacted and tested immediately. It is expected that there will not be complete testing compliance by all selected individuals. Efforts should be made to replace those individuals with other randomly selected individuals in order to achieve the targeted sample size. Data on the selected individuals with missing diagnostic test values due to noncompliance should be examined along with the completed testing data in order to statistically assess potential biases that may result from noncompliance.

Pooled testing methods should be utilized as they represent a resource-conservative approach. Pool size should be based on testing lab recommendations and logistical considerations. In the event of a positive pooled sample, individuals within the pool should be retested so to identify which pool members are infected. These infected individuals should then be subsequently isolated and contact tracing methods should be applied to potentially identify other individuals at risk. After a positive pooled result but before the results of the subsequent individual testing, isolation of all pool members may be considered in order to mitigate virus transmission.

Further design modifications, e.g., blocking on specific individual-level characteristics, may be made to the above plan depending on the long-term use of the data. A survey of individual selected for testing should be implemented in order to obtain potentially important demographic data not normally collected by the University as well as measures of behavior known to be associated with virus transmission. Such collected data should be used to guide future adaptive sampling plans and modeling.

Statistical analysis methods

The analysis objectives once testing data is obtained are to 1) estimate the SARS-COV-2 virus prevalence within each of the stratum, and 2) estimate the overall SARS-COV-2 virus prevalence. In addition to point estimates, corresponding 95% confidence intervals should be computed to convey the uncertainty with the estimation process. Additional subgroup analyses defined by collected data or variables already within University records may be performed, but it should be noted such parameter estimation may lack adequate precision depending on the sample size of said subgroups. A complication in the estimation process is due to the likely imperfect diagnostic test utilized. In such scenarios, the use of the relative frequency results in a biased estimation process for the true prevalence. It is recommended that the adjusted procedure as described in Rogan and Gladen (1978) be utilized for estimation of true prevalence. Note that the sensitivity and specificity of the diagnostic test must be known for proper implementation, and if unknown, reasonable values must be utilized. Exact 95% confidence intervals should be computed using any of the methods discussed in Reiczigel, Földi, Özsvári (2010) although Blaker’s method is recommended for general use by the authors.

Sample size calculation

In that accurate assessment of the true prevalence is of utmost importance in the proposed plan, sample size is based on the precision of the associated estimates which may be conveyed using the expected confidence interval half width (CIHW). It is recommended that the CIHW is set to 2 percentage points as it provides a compromised between feasibility constraints and having adequate precision. The required sample size for a given CIHW is a function of the true unknown prevalence value and the sensitivity and specificity of the diagnostic test to be utilized for sample testing. Based on available data it is expected that the true prevalence is less than 1%. The required sample size increases for values of the true prevalence as they approach 50% from either side. Thus, the required sample size at 1% will provide a conservative
calculation. For values of sensitivity and specificity, 84% and 99%, respectively, were assumed. Using the method in Humphry, Cameron, and Gunn (2004), we see that a sample size of 251 per stratum is required. Allowing for missingness of 10% of testing values, a total of approximately 279 individuals per stratum should be randomly selected to participate in testing.

Acknowledgement

These recommendations are based on the contributions of the following UB faculty:

- Michael E. Cain, M.D., Vice President for Health Sciences and Dean, Jacobs School of Medicine and Biomedical Sciences
- Michael J. LaMonte, Ph.D., MPH, Research Associate Professor, Department of Epidemiology and Environmental Health
- John A. Sellick, Jr., DO, MS, Professor, Department of Medicine
- John E. Tomaszewski, M.D., MASC, SUNY Distinguished Professor and Peter A. Nickerson, PhD, Chair of Pathology and Anatomical Sciences
- Jean Wactawski-Wende, Ph.D., SUNY Distinguished Professor and Dean, School of Public Health and Health Professions
- Gregory E. Wilding, Ph.D., Professor and Chair, Department of Biostatistics

References

