Published in final edited form as:

Med Decis Making. 2021 November ; 41(8): 970–977. doi:10.1177/0272989X211021603.

Maximizing the Efficiency of Active Case-Finding for SARS-CoV-2 Using Bandit Algorithms

Gregg S. Gonsalves, PhD^{1,11,*}, J. Tyler Copple, MS^{1,11}, A. David Paltiel, PhD^{2,11}, Eli P. Fenichel, PhD³, Jude Bayham, PhD⁴, Mark Abraham, BA⁵, David Kline, PhD⁶, Sam Malloy, MPA⁷, Michael F. Rayo, PhD⁸, Net Zhang⁷, Daria Faulkner, BA⁹, Dane A. Morey, MS⁸, Frank Wu, BS^{1,11}, Thomas Thornhill, MPH^{1,11}, Suzan Iloglu, PhD^{1,11}, Joshua L. Warren, PhD^{10,11}

¹Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, CT.

²Department of Health Policy and Management, Yale School of Public Health, New Haven, CT.

³Yale School of the Environment, New Haven, CT.

⁴Department of Agricultural and Resource Economics, Colorado State University, Fort Collins, CO.

⁵DataHaven, New Haven, CT.

⁶Center for Biostatistics, Department of Biomedical Informatics, The Ohio State University, Columbus, OH.

⁷Battelle Center for Science, Engineering, and Public Policy, John Glenn College of Public Affairs, The Ohio State University, Columbus, OH.

⁸Integrated Systems Engineering, The Ohio State University, Columbus, OH.

⁹College of Public Health, The Ohio State University, Columbus, OH.

¹⁰Department of Biostatistics, Yale School of Public Health, New Haven, CT.

¹¹Public Health Modeling Unit, Yale School of Public Health, New Haven, CT.

Even as vaccination for SARS-CoV-2 expands in the United States, cases will linger among unvaccinated individuals for at least the next year allowing the spread of the coronavirus to continue in communities across the country. Detecting these infections, particularly asymptomatic ones, is critical to stemming further transmission of the virus in the months ahead. This will require active surveillance efforts in which these undetected cases are proactively sought out rather than waiting for individuals to present to testing sites for diagnosis. However, finding these pockets of asymptomatic cases (i.e., hotspots) is akin to searching for needles in a haystack as choosing where and when to test within communities is hampered by a lack of epidemiological information to guide decision-makers' allocation of these resources. Making sequential decisions with partial information is a classic problem

*Correspondence to: gregg.gonsalves@yale.edu.

Research data: All code and data associated with the web app are available here at: <https://github.com/NetZissou/Bandit>. The web app itself is available here at: <https://netzissou.shinyapps.io/BanditDemo/>.

in decision-science, the explore versus exploit dilemma. Using methods—bandit algorithms—similar to those used to search for other kinds of lost or hidden objects, from downed aircraft or underground oil deposits, we can address the explore vs. exploit trade-off facing active surveillance efforts and optimize the deployment of mobile testing resources to maximize the yield of new SARS-CoV-2 diagnoses. These bandit algorithms can be implemented easily as a guide to active case finding for SARS-CoV-2. A simple Thompson sampling algorithm and an extension of it to integrate spatial correlation in the data are now embedded in a fully functional prototype of a web app to allow policymakers to use either of these algorithms to target SARS-CoV-2 testing. In this instance, potential testing locations were identified by using mobility data from UberMedia to target high-frequency venues in Columbus, Ohio as part of a planned feasibility study of the algorithms in the field. However, it is easily adaptable to other jurisdictions, requiring only a set of candidate test locations with point-to-point distances between all locations, whether or not mobility data is integrated into decision-making in choosing places to test.

Introduction

Even as vaccinations against SARS-CoV-2 roll out in the United States (US), new infections continue to mount across the country (1). However, while new infections will decrease as more people are immunized, lingering cases of the coronavirus will still exist in communities across the US, frustrating attempts to fully suppress transmission and end the pandemic (2). Even though multiple sites for testing for SARS-COV-2 are available in many communities, tracking down and detecting many cases of the virus will require active surveillance, in which public health workers seek out infections rather than waiting for individuals to present for diagnosis (3).

While active surveillance efforts, particularly tied to contact tracing, were stymied by the scale of the epidemic in 2020 in the US, other countries, particularly, China, South Korea, Hong Kong, Singapore, Taiwan, Australia, Vietnam, and New Zealand achieved low community transmission levels through robust control measures that include extensive surveillance efforts (4,5). In fact, relying on presentation to healthcare facilities for diagnosis cannot contain the pandemic and community-based surveillance and contact tracing are vital to early detection of new cases and prevention of the resurgence of the disease (5,6).

Community-based surveillance efforts in the US are now widespread and have taken SARS-CoV-2 testing out of the healthcare setting. There are different types of off-site testing designs, retrofitting existing structures (e.g. a sporting venue), using tents for pop-up clinics and vans to take SARS-CoV-2 testing fully mobile (7). Our efforts are directed towards the last two design choices, in which SARS-CoV-2 testing does not depend on individuals seeking out a fixed site, but can be moved to seek out places in the community at higher risk for transmission than others and reach vulnerable populations that may not be able to reach other venues at a distance from their homes (e.g. the elderly, those without transport).

These kinds of mobile testing opportunities have been successful in increasing uptake of HIV testing in the US and abroad, particularly targeting high-risk populations that otherwise may not come forward for screening (8–10). This targeted testing among high-risk

communities is now also happening in the context of COVID-19 (11). Combining these types of testing interventions with geo-spatial and phylogenetic data, information on social and sexual networks has also been proposed as a way to home in on hotspots of HIV, increasing the yield of testing (12). However, even with these kinds of efforts, approximately 14% of Americans living with HIV remain unaware of their HIV serostatus, leaving the detection of undiagnosed infections as the “holy grail” of HIV control efforts indicating additional approaches are necessary to reach these individuals (13). A similar situation may persist with SARS-CoV2 in the US, with many undiagnosed infections still undetected across the country (14).

As happened with HIV infection, new platforms for SARS-CoV-2 testing are emerging quickly, from at-home tests, to rapid antigen assays to supplement standard laboratory-based PCR diagnostics, with saliva-based alternatives to invasive nasopharyngeal swabs for the comfort and convenience of patients (15). But these technologies cannot address the simple question that underlies active surveillance efforts: *where can we find lingering cases of SARS-COV-2?* Identifying most, if not all infections in the US through active case finding, contact tracing, isolation of infected individuals and quarantine of their contacts is the ideal route to containing SARS-CoV-2 (16–18). Universal testing, contact tracing, and isolation strategies for SARS-CoV-2 deployed in places like Wuhan, China and akin to the universal test-and-treat efforts for HIV in some countries are expensive, resource-intensive and, in the context of SARS-CoV-2 control in Wuhan, raised human rights concerns (19,20). If universal testing, contact tracing and isolation are infeasible for the US, how can we maximize the yield of new cases detected? While we might target pop-up or mobile testing to places where we believe the prevalence of undetected infection to be very high (e.g., apartment buildings, nursing homes, police stations) in many cities and towns choosing between these venues may be difficult both in terms of their epidemiological value (i.e., the underlying prevalence at a site) and the willingness of people passing through these locations to volunteer for testing.

The elements of the predicament for screening for SARS-CoV-2 in the US are clear: many people remain undiagnosed with SARS-CoV-2 and the prospects of universal testing of entire communities are slim. Thus, we want to maximize the number of cases detected with the limited resources we do have, while cognizant that we *also* have imperfect information about where these undetected infections are to be found. Policymakers have choices about how to address this problem. In deciding where to deploy their testing resources, policymakers must choose between: making the best possible choice based on their current understanding of the evidence or; investing in improving their understanding of the evidence in the hope that it will lead to even better choices in subsequent periods. That is, policymakers can go with what they know and target testing at places they assume are high risk (e.g., nursing homes), but the kinds and numbers of high-risk venues, which might yield the most cases, may be large and diverse and shift over time(21). In addition, SARS-CoV-2 is an over-dispersed pathogen tending to spread in clusters with heterogeneity and stochasticity in transmission and targeting locations based on simple assumptions about risk environments may turn up to be dead-ends (22). Beyond universal testing, the alternative is to test randomly across a community, with the hopes of finding ‘hotspots,’ but facing the prospect that the number of positive diagnoses may wane at promising locations as testing

uncovers most of the undiagnosed cases there or the epidemic moves onto new places in a community. Both of these choices open to policymakers present an iterative series of questions: where do we test today, how long do we stay in that location, when do we move on and where do we go next? How can policymakers best make these complicated decisions between the choices open to them for venues to test for SARS-CoV-2 on an ongoing basis and in an evidence-based fashion? Here we describe how a set of tools we have modified and adapted from the sequential decision-making field, namely bandit algorithms, may help solve the conundrum of SARS-CoV-2 testing in the context of limited resources.

The Explore vs. Exploit Dilemma and Bandit Algorithms

The “explore-versus-exploit dilemma” is a classic problem, where a limited resource must be deployed across alternative targets in a way that maximizes overall gain, when the critical attribute of each target is only partially known at the time of deployment but may become better understood as a result of the deployment decision. This is a problem that has been studied in-depth in the fields of operations research and decision science. How do you mine your best current prospects (“exploit”) while keeping your eye open to better opportunities (“explore”)? The tools that are used to address this dilemma are called *bandit algorithms* (23). They are widely used to guide sequential decisions under uncertainty in a range of settings, from commercial applications (e.g., oil exploration) to military efforts (e.g., searching for downed airplanes) (24).

We have previously studied the use of bandit algorithms for HIV testing—to identify undiagnosed HIV infection using mobile testing units—for several years (24,25). Using model-based simulation studies, we have shown that bandit algorithms outperform more traditional approaches for deploying HIV testing resources, including going this year where you found the most HIV cases last year (i.e. choosing current testing locations based on historical yield of new diagnoses in these places), or sampling a large number of candidate locations before settling down on the best place to test (24,25). The basic bandit algorithm—known as Thompson sampling—outperforms these other methods (26). Thompson sampling is an adaptive Bayesian approach. First, it makes an inventory of all possible target settings. Second, a policymaker offers an initial assessment of the prevalence of undetected infection in each target setting. This takes the form of a probability distribution and is, by design, a subjective exercise that permits the policymaker to be as definitive or tentative as their prior information directs them to be. These prior probability distributions are updated as new information arrives. Third, the decision to deploy testing resources on the first day is made via a random selection from the various prior distributions assigned to each candidate setting. Fourth, on each day a record is maintained in every active testing setting of the total tests performed and the number of positive cases detected. This information is used to update the prior distribution for that site. Then, the decision to deploy testing resources on the next day is repeated via a random selection using the updated priors (i.e., posteriors) and the process repeats.

At the outset, this strategy assigns greatest priority to settings based entirely on the policymaker’s initial assessment, which is based on the data available at the moment. But as the testing campaign proceeds, this strategy provides new information about the probability

of finding a case among the available locations by using the daily test results (positive and negative) at each site selected to refine the understanding of the prevalence at each location. Over time, this continuous process of “learning-while-doing” homes in on the places with highest potential yield of finding new cases faster than other strategies. In addition, if the situation changes—for instance, if one has saturated a given location and depleted the number of undetected cases in that location—the posterior distribution associated with that location will reflect those shifts as well, making it less likely that the site will be chosen in the future. The algorithm is flexible in its accommodation of the time-value of information. If, for example, there is reason to believe that information acquired in previous rounds should have diminishing influence over time (e.g. the site has not been visited in a month) or some exogenous factor has changed the underlying environment (e.g. emergence of new viral strains or variants), one can apply a “discount factor” that assigns less and less weight to prior observations as time goes by. Alternatively, if there is reason to privilege initial assumptions and to make decisions increasingly resistant to new observations, one can apply a different discount factor that assigns decreasing weight to newer data (27). In practical terms, this strategy is meant on an ongoing basis to guide and draw those performing testing in the field to the sites where more people are willing to test *and* with a premium on locations with a higher prevalence of undetected infection. That is, the goal is to maximize yield of positive cases, not to estimate local prevalence of disease. The details of the Thompson sampling strategy are described in Table 1.

We have also developed a variation on Thompson sampling to account for spatial correlation in the prevalence of undetected infections between adjacent geographical areas using a hierarchical Bayesian spatial modeling framework employing an intrinsic conditional autoregressive (ICAR) prior distribution for the spatial random effects and exchangeable, normally distributed random effects to account for non-spatial heterogeneity (25,28). The details of the spatial algorithm are shown in Table 2.

Adapting the Bandit Algorithm to SARS-CoV-2

Bandit algorithms, including Thompson sampling and those that model spatial correlation, can be implemented easily as a guide to active case finding and screening for SARS-CoV-2. In order to provide priors to initialize the algorithms for use with SARS-CoV-2, we defined a set of highly trafficked candidate locations for daily testing. For Columbus, Ohio we used a raw data from the UberMedia COVID-19 recovery data set (<https://covid19.ubermedia.com/covid-19-recovery-insights/>), which contains pairs of individual smart devices within 5 meters of each other within a 5 minute window. These data were cleaned for obvious geolocation errors and to remove contacts on roadways using the Census Tiger Lines. These data were then spatially joined with Loveland Landgrid (<https://landgrid.com/>) parcel data. We produced indices of contacts and unique contacts per parcel, we also rarified the window of the contact definition. These various indices generally produced consistent rankings. We chose highly trafficked locations as these venues create more opportunities for casual encounters for testing but also because these bandit algorithms are sensitive to testing volume and a low number of volunteers could hamper the effectiveness of this approach. In fact, in our studies of bandit algorithms in the context of HIV infection, at less than 10 tests per day, bandit algorithms performed poorly (24). Before deploying these algorithms in field

testing, establishing a floor for daily testing volume at potential locations through simulation using best estimates of local epidemiology of SARS-CoV-2 will be an important operational consideration and performing more tests at fewer sites may be a trade-off to weigh for those using these algorithms in practice. We supplemented this list of highly trafficked locations with residential settings (e.g., apartment buildings), where close contacts are numerous and frequent and where adherence to social distancing and other infection control measures may be more difficult. These are likely to be potential hotspots for disease transmission and could be important places to search for new cases, particularly in census tracts with no public locations for testing. Because we are assessing the yield of testing at point locations rather than areal zones, we use the Euclidean distance between locations rather than map adjacencies when defining spatial proximity. However, distances between locations can be defined in other ways given topographical considerations and local contexts. We set prior distributions for the prevalences in these locations with a Beta(0.50,0.50) distribution (i.e., Jeffreys prior), indicating our lack of knowledge of SARS-CoV-2 prevalence in any of these areas in Columbus. With more data on the epidemiology of SARS-CoV-2, these prior distributions could be crafted to reflect more knowledge of local epidemics. Potential testing sites can also be determined in other ways beyond the use of cell phone and epidemiological data. For instance, targeting industries that are low-work-from-home and demand high-physical proximity for workplace testing could be considered as potential targeting sites where social gatherings regularly still take place even in the context of the pandemic (e.g., houses of worship)(29,30).

We have now set up a fully functional prototype of a web app to allow policymakers in Columbus to use Thompson sampling to target SARS-CoV-2 testing in the city (Figure 1).

It is straightforward to adapt the algorithm to other jurisdictions, requiring only a set of candidate test locations with point-to-point distances between all locations. Finally, the only inputs required for these bandit algorithms between testing forays are the number of tests performed on a given day in each location and the number of positive tests obtained on that day in that location. This makes the algorithm easy to use by local public health departments. The algorithm is simple enough that it can make virtually instantaneous use of new data to inform the deployment of testing units for the next day's effort. This would make this approach best suited for rapid diagnostic tests, particularly rapid lateral-flow antigen-based assays, which though with lower sensitivity than standard PCR are well-positioned to detect individuals with high titers of SARS-CoV-2 and most likely to transmit in a given settings [7]. In fact, while a pilot in Columbus, Ohio is still in the planning stage, we intend to use the BinaxNOW™ COVID-19 Ag Card provided through our collaboration with the Ohio Department of Health(31). However, even with standard PCR based assays and the delays in reporting of SARS-CoV-2 results in practice, these tools can be useful. In this case, the updating of prior distributions for testing locations will be lagged, so the allocation of new testing assignments will only benefit from additional information about potential yield of new diagnoses among locations as previous days testing results become available. Furthermore, bandit algorithms complement test pooling strategies, and can help to concentrate positives within pools to minimize the number of second round tests needed. Finally, although these algorithms are simple to set up and run, the practical considerations of deploying testing teams to multiple locations, often shifting teams at

least initially from day to day, requires the support and engagement of local public health officials, the resources to mount and maintain a mobile testing program over time and a seamless integration of the algorithms into the normal workflow of testing sites. In the context of SARS-CoV-2, ensuring safety of those getting tested as well as staff at a given testing location is also paramount (7).

Bandit algorithms could provide a useful, simple tool to find needles in a haystack—the tens of thousands of undiagnosed infections from coast to coast—when one cannot test everyone, everywhere. As discussed above, bandit algorithms are widely used to successfully address the explore versus exploit dilemma in several other fields. Deployment of bandit algorithms for SARS-CoV-2 may provide useful answers, enable more cost-effective testing, and offer a lifeline to policymakers trying to figure out where to test next for SARS-CoV-2.

Funding:

Financial support for this study was provided entirely by grants from the National Institute on Drug Abuse DP2 (DA049282 to GSG), R37 (DA15612 to GSG. and ADP), the National Institute of Allergy and Infectious Diseases R01 (AI137093 to JLW) and grants from National Science Foundation Northeast Big Data Innovation Hub Subaward 4(GG01486-02) PTE Federal Award (No. OAC-19165850) (EPF), Yale-AWS Enterprise Agreement (EPF), and from the Tobin Center for Economic Policy at Yale University (EPF). The funding agreement ensured the authors' independence in designing the study, interpreting the data, writing, and publishing the report.

References:

1. Sharma B, Mashal M. Covid-19: U.S. Vaccinations Increase, but Virus Continues to Spread. The New York Times [Internet]. 2021 3 19 [cited 2021 Apr 12]; Available from: <https://www.nytimes.com/live/2021/03/19/world/covid-vaccine-coronavirus-cases>
2. Phillips N The coronavirus is here to stay—here's what that means. Nature. 2021;590(7846):382–4. [PubMed: 33594289]
3. Murray J, Cohen AL. Infectious Disease Surveillance. In: Quah SR, editor. International Encyclopedia of Public Health (Second Edition) [Internet]. Oxford: Academic Press; 2017 [cited 2021 Apr 12]. p. 222–9. Available from: <https://www.sciencedirect.com/science/article/pii/B9780128036785005178>
4. Clark E, Chiao EY, Amirian ES. Why Contact Tracing Efforts Have Failed to Curb Coronavirus Disease 2019 (COVID-19) Transmission in Much of the United States. Clinical Infectious Diseases [Internet]. 2020 8 6 [cited 2021 Apr 12];(ciaa1155). Available from: 10.1093/cid/ciaa1155
5. Lokuge K, Banks E, Davis S, Roberts L, Street T, O'Donovan D, et al. Exit strategies: optimising feasible surveillance for detection, elimination, and ongoing prevention of COVID-19 community transmission. BMC Medicine. 2021 12 17;19(1):50. [PubMed: 33596902]
6. Ferguson N, Laydon D, Nedjati Gilani G, Imai N, Ainslie K, Baguelin M. & Dighe A (2020). Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID19 mortality and healthcare demand. 2020; Available from: <https://spiral.imperial.ac.uk/handle/10044/77482>.
7. Off-Site COVID-19 Testing Toolkit [Internet]. Network for Regional Healthcare Improvement. [cited 2021 Apr 12]. Available from: <https://www.nrhi.org/offsite-testing-toolkit/>
8. Bassett IV, Govindasamy D, Erlwanger AS, Hyle EP, Kranzer K, van Schaik N, et al. Mobile HIV Screening in Cape Town, South Africa: Clinical Impact, Cost and Cost-Effectiveness. PLoS One [Internet]. 2014 1 22 [cited 2015 Oct 24];9(1). Available from: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3898963/>
9. Rubin R Have Tent, Will Do Pop-up HIV Screening. JAMA. 2018 11 27;320(20):2063–5. [PubMed: 30422167]
10. Knight V, Gale M, Guy R, Parkhill N, Holden J, Leeman C, et al. A novel time-limited pop-up HIV testing service for gay men in Sydney, Australia, attracts high-risk men. Sex Health. 2014 8 28;11(4):345–50. [PubMed: 25163573]

11. Rapid COVID-19 Test Shows Promise in Community Test Setting [Internet]. Rapid COVID-19 Test Shows Promise in Community Test Setting | UC San Francisco. [cited 2021 Apr 13]. Available from: <https://www.ucsf.edu/news/2020/10/418761/rapid-covid-19-test-shows-promise-community-test-setting>
12. Burns DN, DeGruttola V, Pilcher CD, Kretzschmar M, Gordon CM, Flanagan EH, et al. Toward an endgame: finding and engaging people unaware of their HIV-1 infection in treatment and prevention. *AIDS research and human retroviruses*. 2014;30(3):217–24. [PubMed: 24410300]
13. Eyawo O, Hogg RS, Montaner JS. The Holy Grail: The search for undiagnosed cases is paramount in improving the cascade of care among people living with HIV. *Canadian journal of public health=Revue canadienne de sante publique*. 2013;104(5):e418. [PubMed: 24183185]
14. Stout RL, Rigatti SJ. Seroprevalence of SARS-CoV-2 Antibodies in the US Adult Asymptomatic Population as of September 30, 2020. *JAMA Netw Open*. 2021;4(3):e211552. [PubMed: 33724387]
15. Kevadiya BD, Machhi J, Herskovitz J, Oleynikov MD, Blomberg WR, Bajwa N, et al. Diagnostics for SARS-CoV-2 infections. *Nature Materials*. 2021;21:1–13.
16. Li Z, Chen Q, Feng L, Rodewald L, Xia Y, Yu H, et al. Active case finding with case management: the key to tackling the COVID-19 pandemic. *The Lancet*. 2020;396(10243):63–70.
17. Wilasang C, Sararat C, Jitsuk NC, Yolai N, Thammawijaya P, Auewarakul P, et al. Reduction in effective reproduction number of COVID-19 is higher in countries employing active case detection with prompt isolation. *Journal of travel medicine*. 2020;27(5):taaa095. [PubMed: 32519743]
18. Kretzschmar ME, Rozhnova G, van Boven M. Isolation and contact tracing can tip the scale to containment of COVID-19 in populations with social distancing. *Frontiers in Physics*. 2021;8:677.
19. Havlir D, Lockman S, Ayles H, Larmarange J, Chamie G, Gaolathe T, et al. What do the Universal Test and Treat trials tell us about the path to HIV epidemic control? *J Int AIDS Soc*. 2020;23(2):e25455. [PubMed: 32091179]
20. Wee S-L, Wang V. Here's How Wuhan Tested 6.5 Million for Coronavirus in Days. *The New York Times* [Internet]. 2020 5 26 [cited 2021 Apr 13]; Available from: <https://www.nytimes.com/2020/05/26/world/asia/coronavirus-wuhan-tests.html>
21. Benzell SG, Collis A, Nicolaides C. Rationing social contact during the COVID-19 pandemic: Transmission risk and social benefits of US locations. *PNAS*. 2020;117(26):14642–4. [PubMed: 32522870]
22. Althouse BM, Wenger EA, Miller JC, Scarpino SV, Allard A, Hébert-Dufresne L, et al. Stochasticity and heterogeneity in the transmission dynamics of SARS-CoV-2. *arXiv preprint arXiv:200513689*. 2020;
23. Berry DA, Fristedt B. *Bandit problems: sequential allocation of experiments (Monographs on statistics and applied probability)* [Internet]. Springer; 1985 [cited 2015 Nov 21]. Available from: <http://link.springer.com/content/pdf/10.1007/978-94-015-3711-7.pdf>
24. Gonsalves GS, Crawford FW, Cleary PD, Kaplan EH, Paltiel AD. An Adaptive Approach to Locating Mobile HIV Testing Services. *Medical decision making: an international journal of the Society for Medical Decision Making*. 2018;38(2):262. [PubMed: 28699382]
25. Gonsalves GS, Copple JT, Johnson T, Paltiel AD, Warren JL. Bayesian adaptive algorithms for locating HIV mobile testing services. *BMC medicine*. 2018;16(1):155. [PubMed: 30173667]
26. Thompson WR. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*. 1933;28:5–94.
27. Davidson-Pilon C. *Bayesian Methods for Hackers: Probabilistic Programming and Bayesian Inference* [Internet]. Addison-Wesley Professional; 2015 [cited 2015 Nov 21]. Available from: <https://books.google.com/books?hl=en&lr=&id=rMKiCgAAQBAJ&oi=fnd&pg=PT14&dq=Probabilistic+Programming+and+Bayesian+Methods+for+Hackers:+Using+Python+and+PyMC.&ots=DLkm-jGzjs&sig=xV7KouHpeL8hakwOCtP9d2pg2p8>
28. Besag J, York J, Mollié A. Bayesian image restoration, with two applications in spatial statistics. *Ann Inst Stat Math*. 1991;39(1):1–20.

29. Mongey S, Pilossoph L, Weinberg A. Which Workers Bear the Burden of Social Distancing Policies? (4 26, 2020). University of Chicago, Becker Friedman Institute for Economics Working Paper No. 2020-51, Available from: .
30. Liu T, Gong D, Xiao J, Hu J, He G, Rong Z, et al. Cluster infections play important roles in the rapid evolution of COVID-19 transmission: A systematic review. *International Journal of Infectious Diseases*. 2020;101;99:374–80. [PubMed: 32768702]
31. Pilarowski G, Marquez C, Rubio L, Peng J, Martinez J, Black D, et al. Field Performance and Public Health Response Using the BinaxNOW™ Rapid Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) Antigen Detection Assay During Community-Based Testing. *Clinical Infectious Diseases* [Internet]. 2020 12 26 [cited 2021 Apr 17];(ciaa1890). Available from: 10.1093/cid/ciaa1890 [PubMed: 33367619]

Figure 1/.

Home page of the web app for targeting SARS-CoV-2 testing with mobile units (<https://netzissou.shinyapps.io/BanditDemo/>).

Table 1

Thompson sampling strategy for identifying testing sites for SARS-CoV-2

Algorithm 1 Thompson sampling Strategy

For each potential testing site $i = 1 \dots, n$ set $X_i(0)=0$, $Y_i(0)=0$.

for each $t = 1, 2 \dots t_{max}$ **do**

 For each testing site $i = 1 \dots, n$, sample $p_i(t-1)$ from the $\text{Beta}(\alpha_i + X_i(t-1), \beta_i + Y_i(t-1))$ distribution.

 Select testing site $j = \text{argmax}_i p_i(t-1)$.

 Perform m Bernoulli trials in testing site j and observe x_j successes and $(m-x_j)$ failures.

 Let $X_j(t) = X_j(t-1) + x_j$ and $Y_j(t) = Y_j(t-1) + (m-x_j)$.

 For all testing sites $i \neq j$, let $X_i(t) = X_i(t-1)$ and $Y_i(t) = Y_i(t-1)$.

End

Table 2

Hierarchical Bayesian spatial strategy for identifying testing sites for SARS-CoV-2

Algorithm 2 BYM StrategyFor each potential testing site $i = 1, \dots, n$ set $X_i(0)=0$, $Y_i(0)=0$.**do while the number of unique potential testing sites is < 10 :**For each testing site $i = 1, 2, \dots, n$, sample $p_i(t-1)$ from the $\text{Beta}(\alpha_i + X_i(t-1), \beta_i + Y_i(t-1))$ distribution.Select testing site $j = \text{argmax}_i p_i(t-1)$.Perform m Bernoulli trials in testing site j and observe x_j successes and $(m-x_j)$ failures.Let $X_j(t) = X_j(t-1) + x_j$ and $Y_j(t) = Y_j(t-1) + (m-x_j)$.For all testing sites $i \neq j$, let $X_i(t) = X_i(t-1)$ and $Y_i(t) = Y_i(t-1)$.**do while the number of unique visited zones is ≥ 10 :**

Fit the hierarchical Bayesian spatial logistic regression model to the complete set of data:

$$X_i(t-1) | p_i \sim \text{Binomial}(Y_i(t-1) + X_i(t-1), p_i)$$

$$\text{logit}(p_i) = \beta_0 + \phi_i + \theta_i$$

where $X_i(t-1)$ is the total number of identified SARS-CoV2 cases in testing site i up to time $t-1$,

$Y_i(t-1) + X_i(t-1)$ is the total number of administered tests in testing site i up to time $t-1$, p_i represents the true but unobserved prevalence in testing site i , β_0 is an intercept parameter, ϕ_i is a spatial random effect which follows the ICAR distribution, and θ_i is an exchangeable random effect (independent and normally distributed with variance parameter, σ_θ^2). Note that only previously sampled zones contribute data to the fitting of this model.

The ICAR random effects are defined conditionally as

$$\phi_i | \phi_{-i}, \sigma_\phi^2 \sim \text{Normal} \left(\frac{\sum_{j=1}^n w_{ij} \phi_j}{\sum_{j=1}^n w_{ij}}, \frac{\sigma_\phi^2}{\sum_{j=1}^n w_{ij}} \right)$$

where $\phi_{-i} = (\phi_1, \dots, \phi_{i-1}, \phi_{i+1}, \dots, \phi_n)^\top$ and w_{ij} describes the spatial proximity between spatial zones i and j (e.g., touching borders, inverse distance weights) with $w_{ii} = 0$ for all i .

After fitting the model, we obtain posterior samples from $f(p_i | \mathbf{X}(t-1), \mathbf{Y}(t-1))$ for each testing site (even those that haven't been visited yet) where $\mathbf{X}(t-1)$ and $\mathbf{Y}(t-1)$ are the complete set of data from all zones ($\mathbf{X}(t-1) = (X_1(t-1), \dots, X_n(t-1))^\top$; $\mathbf{Y}(t-1)$ defined similarly). We then randomly select a posterior sample from each testing site and define it as $p_i(t-1)$.

Select testing site $j = \text{argmax}_i p_i(t-1)$.Perform m Bernoulli trials in testing site j and observe x_j successes and $(m-x_j)$ failures.Let $X_j(t) = X_j(t-1) + x_j$ and $Y_j(t) = Y_j(t-1) + (m-x_j)$.For all zones $i \neq j$, let $X_i(t) = X_i(t-1)$ and $Y_i(t) = Y_i(t-1)$.**End**

Prior Specifications:

- $\beta_0 \sim \text{Normal}(0, 2.85)$; Results in $\approx$ Uniform(0,1) prior probabilities for each testing site *a priori* assuming no excess variability in the data.
- σ_ϕ^2 (variance parameter for the ICAR random effect) $\sim \text{Inverse - Gamma}(3.00, 2.00)$
- σ_θ^2 (variance parameter for the exchangeable random effect) $\sim \text{Inverse - Gamma}(3.00, 2.00)$