Statistical Science
2022, Vol. 37, No. 2, 270–277
https://doi.org/10.1214/22-STS859
© Institute of Mathematical Statistics, 2022

Being a Public Health Statistician During a Global Pandemic

Bhramar Mukherjee

Abstract. In this perspective, I first share some key lessons learned from the experience of modeling the transmission dynamics of SARS-CoV-2 in India since the beginning of the COVID-19 pandemic in 2020. Second, I discuss some interesting open problems related to COVID-19 where statisticians have a lot to contribute to in the coming years. Finally, I emphasize the need for having integrated and resilient public health data systems: good data coupled with good models are at the heart of effective policymaking.

Key words and phrases: COVID-19 pandemic, epidemiological models, forecasting, public health data systems, research infrastructure, transmission dynamics, vaccine effectiveness.

INTRODUCTION

For the last two years, our research team has been modeling the SARS-CoV-2 pandemic in India [10, 81, 83, 88, 104]. The work that grew out of an altruistic need to contribute to the pandemic fight by my country of origin, gradually became a source of many interesting statistical questions and ideas. The learning curve has been steep since we did not have prior experience in building forecasting models for infectious diseases before the pandemic. After the rapid-paced "nowcasting" (a term used for real-time forecasting) work that was primarily focused on informing the public and the policymakers in India subsided, we were able to focus more on other more nuanced statistical issues. Our team explored the electronic health records data from Michigan Medicine to answer different types of questions related to COVID-19 testing, outcomes and risk factors [37, 87, 89, 103]. In this perspective, my goal is to share some key lessons that I learned through this modeling experience and discuss some of the pressing problems where I think statisticians have a lot to contribute to in the months and years to come.

Key Lessons

There were several key lessons that I learned which go far beyond our knowledge of mathematical models and emphatically depart from the standard academic approach to research. Our goal from the very beginning was to make real-time public health impact. The resources and stamina one needs to attain this goal are very different in nature.

(1) *Having an interdisciplinary team*: To conduct real time forecasting you need a team who will be largely available to link data sources, maintain software and write code in real time. Methods development was secondary in the initial phase as we were using existing models for virus transmission [100]. Tuning, running, and updating models daily are not part of the typical research projects of a graduate student or a faculty member. The rewards for such "chores" are very little in traditional academia. Most of this work will not be translated into papers, but it takes considerable time and capability to do it well. This implies you need full-time staff programmers for day-to-day operations. Fortunately, Michigan Biostatistics had a cadre of excellent programmers who were willing to volunteer and devote their time to this project.

(2) *Public Communications*: Our initial projections for the daily new cases in India were released through a Medium article [19] where we focused on reaching a broader audience in a non-technical way. The article had a link to an accompanying technical report that had the detailed descriptions of the models we used. We could not choose a peer-reviewed outlet as time was of essence. A typical statistics paper, even if the review process is quick, takes about 6–9 months to be published. That was not the purpose of this work. The medium article had a major impact. It was circulated widely, covered in the media worldwide and was quoted as a critical piece of evidence while considering India's national lockdown that started on March 25, 2020. At that time there were no formal projections available for India.

Since then, our research has been constantly in the media, and we have adopted various modes of communication. We never knew that the pandemic would continue this long but had the foresight to create an app

Bhramar Mukherjee is Professor of Biostatistics, Epidemiology and Global Public Health at the University of Michigan, Ann Arbor, Michigan 48109-2029, USA (e-mail: bhramar@umich.edu).

(covind19.org) [20] that updates daily projections and epidemiologic metrics for India, its states and union territories. This app became a "go to" destination for policymakers, journalists, epidemiologists, and the public. We also wrote multiple newspaper articles and opinion pieces [21, 33, 71–73] with a wide reach.

Television was a new medium for me, it was daunting at first. Through popular prime time news programs in India, we were able to relay our message and alert the public when it was needed the most. During the devastating second wave in 2021, we predicted it very early as well as pointed out the dire consequences of holding unmasked political rallies and religious festivals with little or no public health interventions in place. I also used my social media outlets with current modeling updates to maintain a compelling and regular data narrative. One reason why I have continued to engage with the media though statisticians often prefer to work behind the scenes is I wanted statistics to be at the table, to be recognized as a principal scientific discipline contributing to the pandemic. Otherwise, we would only see economists, physicists, mathematicians, and clinicians talking about epidemiology and modeling in India. We statisticians are the ones who have intimate and primary knowledge of modeling, and we should claim our position.

It really helped to have formal media training and support from the University communications team. We need to have more resources for statisticians to engage with media, to have training in public communications and to learn effective ways of partnering with journalists. As the late Dr. Janet Norwood, a pioneering satistician who worked tirelessly on data and policy said "do not just publish your research, publicize it" [29]. In the process, I became trusted friends with many journalists that I currently correspond and work with. Data journalists played a huge role in identifying gaps and patterns in the data from India [27, 34, 92].

(3) *Keeping up with the pace of Science*: The compartmental models traditionally used for disease transmission have many parameters that require external information. For example, generation interval (the time it takes for an infected person to infect the next person, often approximated by test-to-test times) keep changing with emerging variants [19, 100]. The generation interval has decreased over the course of the pandemic [40], from around 5.5 days for the Alpha variant and 4.7 days for the Delta variant [41] to approximately 3 days for Omicron [3]. This meant you had to really follow the evolving literature on the emerging variants and change these parameters as we progressed over time. Twitter and medrxiv were extremely helpful to keep up with the latest developments. The incredible work that Dr. Eric Topol has done in using his twitter handle (@EricTopol) for synthesizing evidence by picking out key papers and preprints was invaluable [28]. Dr. Topol served as an unofficial primary reviewer of a large number of preprints as we were swimming in information, often conflicting with each. I spent a significant amount of time reading the papers and followed literature in various disciplines like in evolutionary virology journals that were closely related to our modeling work.

(4) *A fair reward system*: While as a senior faculty I was able to devote a significant portion of my time to this work which is hard to capture in a CV, my junior colleagues often asked me the question that if they spent time doing such pro-bono and largely unfunded COVID research, whether and how it will be considered towards their promotion. I played safe and told them to balance between their mainstream research and their COVID work. Similarly, I had to protect my own students and keep them focused on their methodological research. They were clearly more interested in COVID research. This balancing act was a constant struggle. Finally, I decided to divide COVID research into three buckets, one which is rapid nowcasting (just for public service), applied analysis (published in public health or medical journals) and methodological papers that are longer term (published in statistical journals). I facilitated the structure of each project in such a way that the workload was distributed across staff, undergraduate trainees, Masters students from Statistics/Biostatistics, Masters students from other disciplines and finally the doctoral students. I tried to match projects with interests and with incentives that are most valued in each training track. While I was pragmatic in my recommendations to junior faculty, it also raised the question in me: "are we playing it too safe in academia?", it has become more of a game than a pursuit. Perhaps the pandemic can help us all to reflect on what we value in the academe.

(5) *Humility of Knowledge*: As we moved forward with time, things changed. It was very important to quantify the huge uncertainty in the models and state emphatically that long range predictions are impossible with these models [38, 53, 74]. They can only be used for short term projections. Over time, things became more complex even for short term projections (2–4 weeks of time horizon), we needed a system of models: for human behavior, for virus evolution, for policy measures, for vaccine dissemination/effectiveness—just having a disease transmission model was not enough. One had to be flexible in terms of shifting pivots and changing scientific beliefs. For example, our notion of protective herd immunity changed as new variants with immune evasion properties emerged. The concept of herd immunity or endemicity of COVID was something around which the whole community drifted over time [6]. It was important to set the humble tone from the very beginning that we do not know many things and are writing the playbook as we go along.

IMPORTANT AREAS OF STATISTICAL RESEARCH

As an ardent reader of the COVID-literature, in this section I will try to classify the principal genres of statistical problems that are most prominent in the literature right now. We see most of the papers in this volume of *Statistical Science* addressing one or more of these issues.

(a) *Forecasting of infection and deaths at an aggregate level*: There are various kinds of models for disease transmission [11, 35, 90], we have heard more about the compartmental models but there are regression models [78, 81, 82] and hybrid models [44]. Models rely on availability of data. Some models are built on just coarse time series of cases, recoveries and death counts as that is all that may be available for many countries at a national level. Others decompose the compartments in terms of factors that may determine transmission like mobility, age-sex composition, which can be fit only if disaggregated data are available at a more granular level. Many of these models are governed by an underlying system of differential equations. To incorporate uncertainty around the generated forecasts, Bayesian methods are often applied. The models need external input as the latent time between getting exposed and becoming infectious or symptomatic are key parameters and they keep changing with new emerging variants. The immune evasion, virulence and transmissibility properties of a new variant are important to follow [5, 55, 64, 93] to build such models.

The other major issue for SARS-CoV-2 transmission models is estimating the large percentage of covert/unreported/asymptomatic infections that never get detected. Many models make ad hoc assumptions about the ascertainment rate, for example, for the index reported case there were k number of undetected infections at the start of the process. However, seroprevalence surveys (in the unvaccinated) are useful tools to identify this parameter. Similarly, many countries have a highly incomplete death reporting system [46, 62, 63] and models that assume a fixed infection fatality rate may also perform quite poorly. A thorough sensitivity analysis and propagation of uncertainty on these assumed values are necessary, making the Bayesian framework a natural avenue. Principled use of data integration tools that use external information are also essential.

(b) *Risk prediction at an individual level*: Who is at the highest risk of hospitalization, mortality and long COVID with and without vaccines/boosters? These are probably some of the key questions facing us. There has been ML and standard statistical models [17, 43, 69, 99] developed for this task but we need to understand the data and the science better to filter spurious associations/predictors. Most of these models are based on electronic health records or EHRs in the US where the issue of selection bias and poor phenotyping cannot be ignored. Who is in my study and what is the target population of inference has to be carefully thought [7]. Classic concepts like transportability and generalizability are critical to design good prediction algorithms [16, 42, 52, 101]. Temporal data are essential to ensure pre-diagnosis variables are used for prediction. Often these models condition on patients testing positive for COVID-19, but that may introduce bias depending on who has access to testing/care. Choosing the right control group is key to obtaining a valid prediction model. We have seen several designs at play here: test negative designs, untested controls, self-control case series, case-crossover designs [30, 60, 61, 80]. Each design has its pros and cons and assumptions that must be justified. Similarly, who is at the risk of long COVID and what the future entails for long COVID patients is important to characterize and understand. It is apparent that we are going to deal with a sicker population in the coming years and we need healthcare resources to combat the post pandemic lingering of post-acute sequelae of COVID [13, 14, 76] and other healthcare needs. Study of long COVID or broadly COVID survivorship is one of the areas where statisticians have a great deal to contribute in the coming years.

(c) *Vaccine Effectiveness*: Many designs have been adopted for this purpose using real world observational data [54]. Test negative designs that are commonly used suffer from testing bias and for their generalizability we need to know who is getting tested and why. As the disease progresses over time, it is hard to find people who are tested negative or find matched untested controls. It is important to choose historic controls or contemporaneous controls with a plausible reference time window to index test date and apply appropriate analytic techniques [85]. Many confusions were planted in the public mind around the effectiveness and safety of vaccines [2, 32, 56, 57]. Misinformation spun around these confusions have propagated and delayed progress with vaccinations in many countries. For example, just counting how many patients are in the hospital that are vaccinated and unvaccinated is meaningless unless you calculate the rate of hospitalization in the two groups by applying the right denominator [68]. With some countries having 80% people vaccinated, even a small rate of breakthrough hospitalizations can lead to a large absolute number of hospitalizations post vaccination. Statisticians have come forward to discuss these issues as well as pointed out the need to control for important confounders [67] while estimating vaccine effectiveness, including but not limited to past COVID infection and the dominant variant actively in circulation at the time of infection [84]. Having a common protocol to track vaccination data worldwide was very necessary [25] but did not happen. In many countries the testing, vaccination, sequencing and clinical outcome databases do not crosstalk and as a result we have no estimates of re-infection rates, breakthrough infection rates, hospitalization and death rates broken down by vaccination status. Nations with integrated data systems and large cohorts, like Public Health England, Clalit Health Systems in Israel, the Danish cohort study [8, 22, 36, 58] have all led to seminal observations regarding vaccine effectiveness, waning of vaccine-induced immunity and the need for boosters.

(d) *Estimation of excess mortality attributable to COVID-19*: This is a very important area with limited subnational death data available for many low- and middle-income countries and the cause of death data having massive missingness. The key questions are to quantify (a) excess mortality during the pandemic (b) what proportion is due to direct impact of COVID (c) what proportion is due to delayed impact of COVID and (d) what proportion of deaths is due to other causes related to the pandemic for example, delay in care, lack of screening, access to healthcare, mental health issues related to isolation, transportation barriers and alike. There was also potential reduction in deaths due to road accidents and from other infectious diseases due to changes in human behavior and public health interventions. Integrating diverse surveys to quantify excess death estimates is a key component to understand the true extent of COVID related mortality [45]. During the Omicron wave, incidental COVID or patients who are admitted "with COVID" or "for COVID" was important to distinguish while estimating COVID specific death rates [75]. To quantify increase in all-cause-mortality within 6 months of COVID diagnosis or within one year of COVID diagnosis are important to understand the overall death toll of COVID.

(e) *Policy evaluation*: We have seen worldwide natural experiments with an ensemble of non-pharmaceutical public health interventions (NPIs) like travel bans, facial covering, social distancing, school closing, night and weekend curfews being rolled out. Very few causally defensible estimates of the effect of these interventions are available in the literature [65, 79]. These measures have crushing social and economic consequences to human life. For building a pandemic resilient future we need geography-specific catalogs of what interventions worked and a framework for introduction of these measures at the right time [65]. Use of synthetic controls, matching, generation of counterfactuals, carefully developing analytic tools from causal inference literature to characterize intervention effects from observational data are needed to tease out meaningful policy relevant inference [1, 66, 86, 95].

(f) *Issues with Diagnostic Tests*: When we started out in 2020, RT-PCR tests were limited in supply. Thus, testing was selective and based on priority. Only people with symptoms, exposure, with certain occupation and co-morbidities could get tested. This resulted in huge selection biases in reported case-counts. Any fair model needed to consider the auxiliary propensity model of who got tested. Moreover, tests were not perfect, and had high degree of false negatives [102]. This can influence transmission models where people with a false negative test can have a false sense of security and mingle with more people [9, 70]. There were also cheaper rapid antigen tests at that time which were less accurate but could be used repeatedly to improve detection rates. Then came the home tests [98] to detect contagiousness. Optimal testing for surveillance, diagnostic detection of more symptomatic cases, optimal quarantine policies are interesting statistical questions with tests and resources being limited [26]. In 2022, with availability of self-tests, most tests remain unreported and it is important to do random testing and wastewater surveillance to track community prevalence and track new variants.

(g) *Evaluation of Therapeutics*: Finally, we recognize the importance of clinical trials and use of real-world observational data to evaluate therapeutic treatments for COVID. We have seen some classic narratives unfold and debunked around use of hydroxycholoroquine [94, 96], Ivermectin [31, 59], Remdesivir [4], Convalescent plasma [50] which will serve as textbook examples in statistics courses. Both randomized clinical trials as well as principled harnessing of real-world data are key to evaluate the role of emerging anti-virals [39].

(h) *Identifying new variants and genetics of COVID*: There is a very interesting area of evolutionary virology which studies the phylogeny of the virus genome to understand mutations as well as characterize their properties. The GISAID (https://www.gisaid.org/) scientists have been constantly looking at the sequencing data from all over the world to track new variants and classify variants of concern. A great visualization and bioinformatic tool is available at https://nextstrain.org/ Predicting mutations and their lethality require sophisticated modeling [15]. Wastewater surveillance is another source of data that have been collected for ecological level surveillance and tracking [24, 51].

Then comes the study of genetics of patients with COVID, what genetic factors are predictive for severe disease or long COVID and if there are other lifestyle or risk factors that modify this genetic risk [91]. The interaction of host and virus genome is also of paramount importance. This is an area of association and interaction studies where statisticians have a wealth of knowledge to contribute.

CONCLUSION

My experience of working in the data-poor environment in India has made me realize the importance of robust and resilient public health data systems that are linkable to other sources, crosstalking in real time. We have decent models, but we need good data. Without linking the testing, vaccination, sequencing, and clinical outcome data, it is hard to provide estimates of reinfection, breakthrough infection, vaccine effectiveness. Data paucity, data opacity and data denial have harmed transparent and data-driven policymaking in India and many other countries. Dynamic statistical agencies and integrated data systems have played a pivotal role in this pandemic. Countries with national health system, population-based cohorts or comprehensive healthcare/insurance systems like England, Denmark and Israel have produced seminal papers often saving millions of lives [22, 36, 58]. Official data and reporting resources like the UK Office for National Statistics [77], UK Health Security Agency [97], and OpenSAFELY [23] have alerted us about waning effect of boosters, transmissibility, immune evasion, and virulence properties of emerging variants.

Just like integrated data we need more integrated view of the pandemic and its challenges. Throughout this pandemic the discipline of Economics and Public health have been pitted against each other. We need a unified framework to assess the cost benefit analysis of mitigation measures. There is a significant literature on the statistical value of life [18, 47] which can be used to translate deaths averted by mitigation measures into a unit of economic savings and then the two can be compared in similar units. Such creative interdisciplinary collaborations are necessary. There are many causal questions on COVID's impact on mental health, learning outcomes from online education and school closing, childhood immunization rates in developing countries, nutrition programs, impact of long COVID on human health, changes in society and behavior that could use the best statistical minds and the best set of methods. Interrupted time series, difference in difference, regression discontinuity designs and various matching methods are being used to answer these questions but most papers are focused on developed countries [12, 48, 49]. We need more statistical work with a global scope. Infectious disease was not a thrust area in many Statistics/Biostatistics departments, but we need more training opportunities in this area. Despite the devastating effect of the pandemic, it has shifted public attention to data and methods. The pandemic provided extremely motivating examples and context to recruit the next generation of students and faculty to our field. Academia with all its imperfections, gave me the degrees of freedom to dive into a problem without seeking permission from anyone and build a team overnight to work on COVID in India. Very few professions have such resource and freedom. My hope is that the last two years will push our profession to value impact and come up with an incentive framework that allows and rewards high risk work influencing policy and improving human life with statistics.

ACKNOWLEDGMENTS

The author would like to thank Maxwell Salvatore for his assistance with compiling the references.

FUNDING

The research was supported by NSF DMS 1712933, NIH R01 HG008773, and NIH P30 CA046592.

REFERENCES

[1] ABADIE, A. (2021). Using synthetic controls: Feasibility, data requirements, and methodological aspects. *J. Econ. Lit.* **59** 391–425. https://doi.org/10.1257/jel.20191450

[2] ABBASI, J. (2022). Widespread misinformation about infertility continues to create COVID-19 vaccine hesitancy. *JAMA*. https://doi.org/10.1001/jama.2022.2404

[3] ABBOTT, S., SHERRATT, K., GERSTUNG, M. and FUNK, S. (2022). Estimation of the test to test distribution as a proxy for generation interval distribution for the Omicron variant in England. *EpiForecasts* **18**. Available at https://epiforecasts.io/omicron-sgtf-forecast/generation-time (accessed Mar. 07, 2022).

[4] ANSEMS, K. et al. (2021). Remdesivir for the treatment of COVID-19. *Cochrane Database Syst. Rev.* **2021**. https://doi.org/10.1002/14651858.CD014962

[5] ARAF, Y. et al. (2022). Omicron variant of SARS-CoV-2: Genomics, transmissibility, and responses to current COVID-19 vaccines. *J. Med. Virol.* **27588**. https://doi.org/10.1002/jmv.27588

[6] ASCHWANDEN, C. (2021). Five reasons why COVID herd immunity is probably impossible. *Nature* **591** 520–522. https://doi.org/10.1038/d41586-021-00728-2

[7] BEESLEY, L. J. and MUKHERJEE, B. (2022). Statistical inference for association studies using electronic health records: Handling both selection bias and outcome misclassification. *Biometrics* **78** 214–226. https://doi.org/10.1111/biom.13400

[8] BERG, S. K. et al. (2022). Long COVID symptoms in SARS-CoV-2-positive adolescents and matched controls (LongCOVIDKidsDK): A national, cross-sectional study. *Lancet Child Adolesc. Health* **6**. 240–248. https://doi.org/10.1016/S2352-4642(22)00004-9

[9] BHADURI, R. et al. (2022). Extending the susceptible-exposed-infected-removed (SEIR) model to handle the false negative rate and symptom-based administration of COVID-19 diagnostic tests: SEIR-fansy. *Stat. Med.* sim.9357. https://doi.org/10.1002/sim.9357

[10] BHATTACHARYYA, R., BHADURI, R., KUNDU, R., SALVATORE, M. and MUKHERJEE, B. (2020). Reconciling epidemiological models with misclassified case-counts for SARS-CoV-2 with seroprevalence surveys: A case study in Delhi, India. *Infectious Diseases (except HIV/AIDS)*. Preprint. https://doi.org/10.1101/2020.07.31.20166249

[11] BRAUER, F., CASTILLO-CHAVEZ, C. and FENG, Z. (2019). Simple compartmental models for disease transmission. In *Mathematical Models in Epidemiology* (F. Brauer, C. Castillo-Chavez and Z. Feng, eds.) 21–61. Springer, New York, NY. MR3969982 https://doi.org/10.1007/978-1-4939-9828-9

[12] BRODEUR, A., CLARK, A. E., FLECHE, S. and POWDTHAVEE, N. (2021). COVID-19, lockdowns and well-being: Evidence from Google trends. *J. Public Econ.* **193** 104346. https://doi.org/10.1016/j.jpubeco.2020.104346

[13] CENTERS FOR DISEASE CONTROL AND PREVENTION. Post-COVID Conditions: Information for Healthcare Providers. *Centers for Disease Control and Prevention*. Available at https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-care/post-covid-conditions.html (accessed Mar. 03, 2022).

[14] CHEN, C., HAUPERT, S. R., ZIMMERMANN, L., SHI, X., FRITSCHE, L. G. and MUKHERJEE, B. (2021). Global Prevalence of Post-Acute Sequelae of COVID-19 (PASC) or Long COVID: A Meta-Analysis and Systematic Review. *Epidemiology*. Preprint. https://doi.org/10.1101/2021.11.15.21266377

[15] CHEN, J., GAO, K., WANG, R. and WEI, G.-W. (2021). Prediction and mitigation of mutation threats to COVID-19 vaccines and antibody therapies. *Chem. Sci.* **12** 6929–6948. https://doi.org/10.1039/D1SC01203G

[16] COLE, S. R. and STUART, E. A. (2010). Generalizing evidence from randomized clinical trials to target populations: The ACTG 320 trial. *Am. J. Epidemiol.* **172** 107–115. https://doi.org/10.1093/aje/kwq084

[17] COVIDSURG COLLABORATIVE et al. (2021). Machine learning risk prediction of mortality for patients undergoing surgery with perioperative SARS-CoV-2: The COVIDSurg mortality score. *Br. J. Surg.* **108** 1274–1292. https://doi.org/10.1093/bjs/znab183

[18] COLMER, J. (2020). What is the meaning of (statistical) life? Benefit–cost analysis in the time of COVID-19. *Oxf. Rev. Econ. Policy* **36** S56–S63. https://doi.org/10.1093/oxrep/graa022

[19] COV-IND-19 STUDY GROUP. Predictions and role of interventions for COVID-19 outbreak in India. *Medium*. Available at https://bhramarm.medium.com/predictions-and-role-of-interventions-for-covid-19-outbreak-in-india-52903e2544e6 (accessed Mar. 03, 2022).

[20] COV-IND-19 STUDY GROUP (2020). COVID-19 Outbreak in India. Available at https://umich-biostatistics.shinyapps.io/covid19/.

[21] COV-IND-19 STUDY GROUP (2020). What can we learn from the Delhi Seroprevalence Study and a single number that was reported? *Medium*. Available at https://bhramarm.medium.com/what-can-we-learn-from-the-delhi-seroprevalence-study-and-a-single-number-that-was-reported-45e4cf185357 (accessed Mar. 03, 2022).

[22] DAGAN, N. et al. (2021). BNT162b2 mRNA Covid-19 vaccine in a nationwide mass vaccination setting. *N. Engl. J. Med.* **384** 1412–1423. https://doi.org/10.1056/NEJMoa2101765

[23] DATALAB AT THE UNIVERSITY OF OXFORD, EHR GROUP AT LONDON SCHOOL OF HYGIENE AND TROPICAL MEDICINE, TPP, AND EMIS. OpenSAFELY. Available at https://www.opensafely.org/ (accessed Mar. 07, 2022).

[24] DAUGHTON, C. G. (2020). Wastewater surveillance for population-wide Covid-19: The present and future. *Sci. Total Environ.* **736** 139631. https://doi.org/10.1016/j.scitotenv.2020.139631

[25] DEAN, N. (2021). COVID vaccination studies: Plan now to pool data, or be bogged down in confusion. *Nature* **591** 179–179. https://doi.org/10.1038/d41586-021-00563-5

[26] DU, J., BEESLEY, L. J., LEE, S., ZHOU, X., DEMPSEY, W. and MUKHERJEE, B. (2022). Optimal diagnostic test allocation strategy during the COVID-19 pandemic and beyond. *Stat. Med.* **41** 310–327. MR4380641 https://doi.org/10.1002/sim.9238

[27] DUTT, B. (2022). *To Hell and Back*: *Humans of COVID*. Juggernaut Publication.

[28] ERIC TOPOL (@ERICTOPOL) / TWITTER. *Twitter*. Available at https://twitter.com/EricTopol (accessed Mar. 03, 2022).

[29] FIENBERG, S. E. (1994). A conversation with Janet L. Norwood. *Statist. Sci.* **9** 574–590. MR1328613

[30] FUKUSHIMA, W. and HIROTA, Y. (2017). Basic principles of test-negative design in evaluating influenza vaccine effectiveness. *Vaccine* **35** 4796–4800. https://doi.org/10.1016/j.vaccine.2017.07.003

[31] GAREGNANI, L. I., MADRID, E. and MEZA, N. (2021). Misleading clinical evidence and systematic reviews on ivermectin for COVID-19. *BMJ Evid.-Based Med.* bmjebm-2021-111678. https://doi.org/10.1136/bmjebm-2021-111678

[32] GARETT, R. and YOUNG, S. D. (2021). Online misinformation and vaccine hesitancy. *Transl. Behav. Med.* **11** 2194–2199. https://doi.org/10.1093/tbm/ibab128

[33] GETTLEMAN, J., YASIR, S., KUMAR, H., RAJ, S. and LOKE, A. (2021). As Covid-19 devastates India, deaths go undercounted. *The New York Times*. Accessed: Jun. 01, 2021. [Online]. Available at https://www.nytimes.com/2021/04/24/world/asia/india-coronavirus-deaths.html.

[34] GHOSH, A. (2021). *Billions Under Lockdown*: *The Inside Story of India's Fight Against Covid*-19. Blosssbury India, New Delhi.

[35] GODIO, A., PACE, F. and VERGNANO, A. (2020). SEIR modeling of the Italian epidemic of SARS-CoV-2 using computational swarm intelligence. *Int. J. Environ. Res. Public. Health* **17** 10. https://doi.org/10.3390/ijerph17103535

[36] GRAM, M. A. et al. Vaccine effectiveness against SARS-CoV-2 infection, hospitalization, and death when combining a first dose ChAdOx1 vaccine with a subsequent mRNA vaccine in Denmark: A nationwide population-based cohort study. *PLOS Med.* **18** e1003874. https://doi.org/10.1371/journal.pmed.1003874

[37] GU, T. et al. (2020). COVID-19 outcomes, risk factors and associations by race: A comprehensive analysis using electronic health records data in Michigan Medicine. *Infectious Diseases (except HIV/AIDS)*. Preprint. https://doi.org/10.1101/2020.06.16.20133140

[38] GUGOLE, F., COFFENG, L. E., EDELING, W., SANDERSE, B., DE VLAS, S. J. and CROMMELIN, D. (2021). Uncertainty quantification and sensitivity analysis of COVID-19 exit strategies in an individual-based transmission model. *PLOS Comput. Biol.* **17** e1009355. https://doi.org/10.1371/journal.pcbi.1009355

[39] HAMMOND, J. et al. (2022). Oral nirmatrelvir for high-risk, nonhospitalized adults with Covid-19. *N. Engl. J. Med.* NEJMoa2118542. https://doi.org/10.1056/NEJMoa2118542

[40] HART, W. S. et al. (2022). Inference of the SARS-CoV-2 generation time using UK household data. *eLife* **11** e70767. https://doi.org/10.7554/eLife.70767

[41] HART, W. S. et al. (2022). Generation time of the alpha and delta SARS-CoV-2 variants: An epidemiological analysis. *Lancet Infect. Dis.* https://doi.org/10.1016/S1473-3099(22)00001-9

[42] HERNÁN, M. A. and VANDERWEELE, T. J. (2011). Compound treatments and transportability of causal inference. *Epidemiology* **22** 368–377. https://doi.org/10.1097/EDE.0b013e3182109296

[43] HIPPISLEY-COX, J. et al. (2021). Risk prediction of Covid-19 related death and hospital admission in adults after Covid-19 vaccination: National prospective cohort study. *BMJ* n2244. https://doi.org/10.1136/bmj.n2244

[44] JEWELL, N. P., LEWNARD, J. A. and JEWELL, B. L. (2020). Predictive mathematical models of the COVID-19 pandemic: Underlying principles and value of projections. *JAMA* **323** 1893. https://doi.org/10.1001/jama.2020.6585

[45] JHA, P. et al. (2022). COVID mortality in India: National survey data and health facility deaths. *Science* **375** 667–671. https://doi.org/10.1126/science.abm5154

[46] KARLINSKY, A. and KOBAK, D. (2021). Tracking excess mortality across countries during the COVID-19 pandemic with the world mortality dataset. *ELife* **10** e69336. https://doi.org/10.7554/eLife.69336
[47] KNIESNER, T. J. and VISCUSI, W. K. (2019). The value of a statistical life. *SSRN Electron. J.* https://doi.org/10.2139/ssrn.3379967
[48] KOLTAI, J., RAIFMAN, J., BOR, J., MCKEE, M. and STUCKLER, D. (2021). COVID-19 vaccination and mental health: A difference-in-difference analysis of the understanding America study. *Am. J. Prev. Med.* S0749379721006012. https://doi.org/10.1016/j.amepre.2021.11.006
[49] KOPPEL, S., CAPELLAN, J. A. and SHARP, J. (2022). Disentangling the impact of Covid-19: An interrupted time series analysis of crime in New York city. *Am. J. Crim. Justice*. https://doi.org/10.1007/s12103-021-09666-1
[50] KORLEY, F. K. et al. (2021). Early convalescent plasma for high-risk outpatients with Covid-19. *N. Engl. J. Med.* **385** 1951–1960. https://doi.org/10.1056/NEJMoa2103784
[51] KUMAR, M. et al. (2020). First proof of the capability of wastewater surveillance for COVID-19 in India through detection of genetic material of SARS-CoV-2. *Sci. Total Environ.* **746** 141326. https://doi.org/10.1016/j.scitotenv.2020.141326
[52] LESKO, C. R., BUCHANAN, A. L., WESTREICH, D., EDWARDS, J. K., HUDGENS, M. G. and COLE, S. R. (2017). Generalizing study results: A potential outcomes perspective. *Epidemiology* **28** 553–561. https://doi.org/10.1097/EDE.0000000000000664
[53] LEUNG, K. and WU, J. T. (2021). Quantifying the uncertainty of CovidSim. *Nat. Comput. Sci.* **1** 2. https://doi.org/10.1038/s43588-021-00031-0
[54] LIN, D.-Y. et al. (2022). Effectiveness of Covid-19 vaccines over a 9-month period in North Carolina. *N. Engl. J. Med.* https://doi.org/10.1056/NEJMoa2117128
[55] LIU, Y. and ROCKLÖV, J. (2021). The reproductive number of the delta variant of SARS-CoV-2 is far higher compared to the ancestral SARS-CoV-2 virus. *J. Travel Med.* **28** taab124. https://doi.org/10.1093/jtm/taab124
[56] LOCKYER, B. et al. (2021). Understanding COVID-19 misinformation and vaccine hesitancy in context: Findings from a qualitative study involving citizens in Bradford, UK. *Health Expect.* **24** 1158–1167. https://doi.org/10.1111/hex.13240
[57] LOOMBA, S., DE FIGUEIREDO, A., PIATEK, S. J., DE GRAAF, K. and LARSON, H. J. (2021). Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA. *Nat. Hum. Behav.* **5** 337–348. https://doi.org/10.1038/s41562-021-01056-1
[58] LOPEZ BERNAL, J. et al. (2021). Effectiveness of Covid-19 vaccines against the B.1.617.2 (Delta) variant. *N. Engl. J. Med.* **385** 585–594. https://doi.org/10.1056/NEJMoa2108891
[59] LÓPEZ-MEDINA, E. et al. (2021). Effect of ivermectin on time to resolution of symptoms among adults with mild COVID-19: A randomized clinical trial. *JAMA* **325** 1426. https://doi.org/10.1001/jama.2021.3071
[60] LU, C. Y. (2009). Observational studies: A review of study designs, challenges and strategies to reduce confounding. *Int. J. Clin. Pract.* **63** 691–697. https://doi.org/10.1111/j.1742-1241.2009.02056.x
[61] MACLURE, M. (1991). The case-crossover design: A method for studying transient effects on the risk of acute events. *Am. J. Epidemiol.* **133** 144–153. https://doi.org/10.1093/oxfordjournals.aje.a115853
[62] MAHAPATRA, P. and CHALAPATI RAO, P. V. (2001). Cause of death reporting systems in India: A performance analysis. *Natl. Med. J. India* **14** 154–162.
[63] MATHERS, C. D., FAT, D. M., INOUE, M., RAO, C. and LOPEZ, A. D. (2005). Counting the dead and what they died from: An assessment of the global status of cause of death data. *Bull. World Health Organ.* **83** 171–177. https://doi.org//S0042-96862005000300009
[64] MCCALLUM, M. et al. (2022). Structural basis of SARS-CoV-2 Omicron immune evasion and receptor engagement. *Science* **375** 864–868. https://doi.org/10.1126/science.abn8652
[65] MENDEZ-BRITO, A., EL BCHERAOUI, C. and POZO-MARTIN, F. (2021). Systematic review of empirical studies comparing the effectiveness of non-pharmaceutical interventions against COVID-19. *J. Infect.* **83** 281–293. https://doi.org/10.1016/j.jinf.2021.06.018
[66] MITZE, T., KOSFELD, R., RODE, J. and WÄLDE, K. (2020). Face masks considerably reduce COVID-19 cases in Germany. *Proc. Natl. Acad. Sci. USA* **117** 32293–32301. https://doi.org/10.1073/pnas.2015954117
[67] MORRIS, J. (2021). Israeli data: How can efficacy vs. severe disease be strong when 60% of hospitalized are vaccinated? *Covid Data Science*. Available at https://www.covid-datascience.com/post/israeli-data-how-can-efficacy-vs-severe-disease-be-strong-when-60-of-hospitalized-are-vaccinated (accessed Mar. 04, 2022).
[68] MORRIS, J. S. Covid-19 Data Science. *Covid Data Science*. Available at https://www.covid-datascience.com/ (accessed Mar. 03, 2022).
[69] MOULAEI, K., SHANBEHZADEH, M., MOHAMMADI-TAGHIABAD, Z. and KAZEMI-ARPANAHI, H. (2022). Comparing machine learning algorithms for predicting COVID-19 mortality. *BMC Med. Inform. Decis. Mak.* **22** 2. https://doi.org/10.1186/s12911-021-01742-0
[70] MOULIOU, D. S. and GOURGOULIANIS, K. I. (2021). False-positive and false-negative COVID-19 cases: Respiratory prevention and management strategies, vaccination, and further perspectives. *Expert Rev. Respir. Med.* **15** 993–1002. https://doi.org/10.1080/17476348.2021.1917389
[71] MUKHERJEE, B. (2022). It's not just India's fight, the world needs to pitch in. *The Times of India*. Accessed: Mar. 03, 2022. [Online]. Available at https://timesofindia.indiatimes.com/blogs/voices/its-not-just-indias-fight-the-world-needs-to-pitch-in/e.
[72] MUKHERJEE, B. (2022). Why we still don't know how many, and who exactly died of Covid in India. *The Times of India*. Accessed: Mar. 04, 2022. [Online]. Available at https://timesofindia.indiatimes.com/blogs/voices/why-we-still-dont-know-how-many-and-who-exactly-died-of-covid-in-india/.
[73] MURDUR, G. S. (2022). India is in a Covid-19 sandstorm, says Bhramar Mukherjee. *The Telegraph India*. Accessed: Mar. 03, 2022. [Online]. Available at https://www.telegraphindia.com/india/india-is-in-a-covid-sandstorm-expert/cid/1846564.
[74] MURRAY, C. J. L. and PIOT, P. (2021). The potential future of the COVID-19 pandemic: Will SARS-CoV-2 become a recurrent seasonal infection? *JAMA* **325** 1249. https://doi.org/10.1001/jama.2021.2828
[75] MURRAY, S. G., CROCI, R. and WACHTER, R. M. (2022). Perspective | is a patient hospitalized 'with' COVID or 'for' COVID? It can be hard to tell. *Washington Post*. Accessed: Mar. 04, 2022. [Online]. Available at https://www.washingtonpost.com/outlook/2022/01/07/hospitalization-covid-statistics-incidental/.
[76] NALBANDIAN, A. et al. (2021). Post-acute COVID-19 syndrome. *Nat. Med.* **27** 601–615. https://doi.org/10.1038/s41591-021-01283-z

[77] OFFICE FOR NATIONAL STATISTICS. Coronavirus (COVID-19). Available at https://www.ons.gov.uk/peoplepopulationandcommunity/healthandsocialcare/conditionsanddiseases (accessed Mar. 07, 2022).

[78] PAN, A. et al. (2020). Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China. *JAMA* **19** 1915.

[79] PEI, S., KANDULA, S. and SHAMAN, J. (2020). Differential effects of intervention timing on COVID-19 spread in the United States. *Sci. Adv.* **6** eabd6370. https://doi.org/10.1126/sciadv.abd6370

[80] PETERSEN, I., DOUGLAS, I. and WHITAKER, H. (2016). Self controlled case series methods: An alternative to standard epidemiological study designs. *BMJ* i4515. https://doi.org/10.1136/bmj.i4515

[81] PURKAYASTHA, S. et al. (2021). A comparison of five epidemiological models for transmission of SARS-CoV-2 in India. *BMC Infect. Dis.* **21** 533. https://doi.org/10.1186/s12879-021-06077-9

[82] QUICK, C., DEY, R. and LIN, X. (2021). Regression Models for Understanding COVID-19 Epidemic Dynamics With Incomplete Data. *J. Amer. Statist. Assoc.* **116** 1561–1577. MR4353694 https://doi.org/10.1080/01621459.2021.2001339

[83] RAY, D. et al. (2020). Predictions, role of interventions and effects of a historic national lockdown in India's response to the the COVID-19 pandemic: Data science call to arms. *Harv. Data Sci. Rev.* https://doi.org/10.1162/99608f92.60e08ed5

[84] ROBERTS, E. K., GU, T., MUKHERJEE, B. and FRITSCHE, L. G. (2022). Estimating COVID-19 Vaccination Effectiveness Using Electronic Health Records of an Academic Medical Center in Michigan. *Infectious Diseases (except HIV/AIDS)*. Preprint. https://doi.org/10.1101/2022.01.29.22269971

[85] ROTHMAN, K. J., LASH, T. L., VANDERWEELE, T. J. and HANEUSE, S. (2021). *Modern Epidemiology*, 4th ed. Wolters Kluwer, Philadelphia, PA.

[86] RUBIN, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *J. Amer. Statist. Assoc.* **100** 322–331. MR2166071 https://doi.org/10.1198/016214504000001880

[87] SALERNO, S. et al. (2021). Patterns of repeated diagnostic testing for COVID-19 in relation to patient characteristics and outcomes. *J. Intern. Med.* **289** 726–737. https://doi.org/10.1111/joim.13213

[88] SALVATORE, M. et al. (2020). Comprehensive public health evaluation of lockdown as a non-pharmaceutical intervention on COVID-19 spread in India: National trends masking state-level variations. *BMJ Open* **10** e041778. https://doi.org/10.1136/bmjopen-2020-041778

[89] SALVATORE, M. et al. (2021). A phenome-wide association study (PheWAS) of COVID-19 outcomes by race using the electronic health records data in Michigan medicine. *J. Clin. Med.* **10** 1351. https://doi.org/10.3390/jcm10071351

[90] SCHLICKEISER, R. and KRÖGER, M. (2021). Analytical modeling of the temporal evolution of epidemics outbreaks accounting for vaccinations. *Physics* **3** Art. no. 2. https://doi.org/10.3390/physics3020028

[91] SEYEDALINAGHI, S. et al. (2021). Genetic susceptibility of COVID-19: A systematic review of current evidence. *Eur. J. Med. Res.* **26** 46. https://doi.org/10.1186/s40001-021-00516-8

[92] SHRINIVASAN, R. (2021). Whole numbers and half truths: What data can and cannot tell us about modern India, Chennai: Westland Publications Private Limited.

[93] SHUAI, H. et al. (2022). Attenuated replication and pathogenicity of SARS-CoV-2 B.1.1.529 Omicron. *Nature*. https://doi.org/10.1038/s41586-022-04442-5

[94] SKIPPER, C. P. et al. (2020). Hydroxychloroquine in nonhospitalized adults with early COVID-19: A randomized trial. *Ann. Intern. Med.* **173** 623–631. https://doi.org/10.7326/M20-4207

[95] STUART, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statist. Sci.* **25** 1–21. MR2741812 https://doi.org/10.1214/09-STS313

[96] THE RECOVERY COLLABORATIVE GROUP (2020). Effect of Hydroxychloroquine in Hospitalized Patients with Covid-19. *N. Engl. J. Med.* **383** 2030–2040. https://doi.org/10.1056/NEJMoa2022926

[97] UK HEALTH SECURITY AGENCY. National flu and COVID-19 surveillance reports: 2021 to 2022 season. *GOV.UK*. Available at https://www.gov.uk/government/statistics/national-flu-and-covid-19-surveillance-reports-2021-to-2022-season (accessed Mar. 07, 2022).

[98] US FEDERAL DRUG AND FOOD ADMINISTRATION (2020). Coronavirus (COVID-19) Update: FDA Authorizes First COVID-19 Test for Self-Testing at Home. *FDA* **18**. Available at https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-authorizes-first-covid-19-test-self-testing-home (accessed Mar. 03, 2022).

[99] WAN, H., CUI, J.-A. and YANG, G.-J. (2020). Risk estimation and prediction of the transmission of coronavirus disease-2019 (COVID-19) in the mainland of China excluding Hubei province. *Infect. Dis. Poverty* **9** 116. https://doi.org/10.1186/s40249-020-00683-6

[100] WANG, L. et al. An epidemiological forecast model and software assessing interventions on the COVID-19 epidemic in China. *J. Data Sci.* **18** 409–432. https://doi.org/10.6339/JDS.202007_18(3).0003

[101] WESTREICH, D., EDWARDS, J. K., LESKO, C. R., STUART, E. and COLE, S. R. (2017). Transportability of trial results using inverse odds of sampling weights. *Am. J. Epidemiol.* **186** 1010–1014. https://doi.org/10.1093/aje/kwx164

[102] WOLOSHIN, S., PATEL, N. and KESSELHEIM, A. S. (2020). False negative tests for SARS-CoV-2 infection—challenges and implications. *N. Engl. J. Med.* **383** e38. https://doi.org/10.1056/NEJMp2015897

[103] YU, Y., GU, T., VALLEY, T. S., MUKHERJEE, B. and FRITSCHE, L. G. (2021). Changes in COVID-19-related outcomes, potential risk factors and disparities over time. *Epidemiol. Infect.* **149** e192. https://doi.org/10.1017/S0950268821001898

[104] ZIMMERMANN, L. V., SALVATORE, M., BABU, G. R. and MUKHERJEE, B. (2021). Estimating COVID-19—related mortality in India: An epidemiological challenge with insufficient data. *Am. J. Public Health* **111** S59–S62. https://doi.org/10.2105/AJPH.2021.306419