STATISTICAL SCIENCE

Volume 37, Number 2 May 2022

Statistical Science **[ISSN 0883-4237 (print); ISSN 2168-8745 (online)]**, Volume 37, Number 2, May 2022. Published quarterly by the Institute of Mathematical Statistics, 9760 Smith Road, Waite Hill, Ohio 44094, USA. Periodicals postage paid at Cleveland, Ohio and at additional mailing offices. **POSTMASTER:** Send address changes to *Statistical Science*, Institute of Mathematical Statistics, Dues and Subscriptions Office, PO Box 729, Middletown, Maryland 21769, USA.

Statistical Science Volume 37, Number 2 (159–288) May 2022

Volume 37 Number 2 May 2022

Special Issue on Data Science in a Time of Crisis: Lessons from the Pandemic

Statistical Science
2022, Vol. 37, No. 2, 159
https://doi.org/10.1214/22-STS372PRE

Preamble

Sonia Petrone, Editor

Statistical Science
2022, Vol. 37, No. 2, 160–161
https://doi.org/10.1214/22-STS372IN

Data Science in a Time of Crisis: Lessons from the Pandemic

Chiara Sabatti and John M. Chambers

REFERENCES

[1] CAPPELLO, L., KIM, J., LIU, S. and PALACIOS, J. A. (2022). Statistical challenges in tracking the evolution of SARS-CoV-2. *Statist. Sci.* **37** 162–182.

[2] CHAMBERS, J. M. (2022). Data, science, and global disasters. *Statist. Sci.* **37** 284–288.

[3] COMESS, S., WANG, H., HOLMES, S. and DONNAT, C. (2022). Statistical modeling for practical pooled testing during the COVID-19 pandemic. *Statist. Sci.* **37** 229–250.

[4] JAHJA, M., CHIN, A. and TIBSHIRANI, R. J. (2022). Real-time estimation of COVID-19 infections: Deconvolution and sensor fusion. *Statist. Sci.* **37** 207–228.

[5] LIN, X. (2022). Lessons learned from the COVID-19 pandemic: A statistician's reflection. *Statist. Sci.* **37** 278–283.

[6] MUKHERJEE, B. (2022). Being a public health statistician during a global pandemic. *Statist. Sci.* **37** 270–277.

[7] NICHOLSON, G., BLANGIARDO, M., BRIERS, M., DIGGLE, P. J., FJELDE, T. E., GE, H., GOUDIE, R. J. B., JERSAKOVA, R., KING, R. E. et al. (2022). Interoperability of statistical models in pandemic preparedness: Principles and reality. *Statist. Sci.* **37** 183–206.

[8] WANG, Z., BOWRING, M. G., ROSEN, A., GARIBALDI, B., ZEGER, S. and NISHIMURA, A. (2022). Learning and predicting from dynamic models for COVID-19 patient monitoring. *Statist. Sci.* **37** 251–265.

[9] YU, B. and SINGH, C. (2022). Seven principles for rapid-response data science: Lessons learned from Covid-19 forecasting. *Statist. Sci.* **37** 266–269.

Chiara Sabatti is Professor, Biomedical Data Science and of Statistics, Stanford University, Stanford, California 94305-2070, USA (e-mail: sabatti@stanford.edu). John M. Chambers is Adjunct Professor in Statistics and Senior Advisor, Data Science, Stanford University, Stanford, California 94305-2070, USA (e-mail: jmc4@stanford.edu).

Statistical Science
2022, Vol. 37, No. 2, 162–182
https://doi.org/10.1214/22-STS853

Statistical Challenges in Tracking the Evolution of SARS-CoV-2

Lorenzo Cappello, Jaehee Kim, Sifan Liu and Julia A. Palacios

Abstract. Genomic surveillance of SARS-CoV-2 has been instrumental in tracking the spread and evolution of the virus during the pandemic. The availability of SARS-CoV-2 molecular sequences isolated from infected individuals, coupled with phylodynamic methods, have provided insights into the origin of the virus, its evolutionary rate, the timing of introductions, the patterns of transmission, and the rise of novel variants that have spread through populations. Despite enormous global efforts of governments, laboratories, and researchers to collect and sequence molecular data, many challenges remain in analyzing and interpreting the data collected. Here, we describe the models and methods currently used to monitor the spread of SARS-CoV-2, discuss long-standing and new statistical challenges, and propose a method for tracking the rise of novel variants during the epidemic.

Key words and phrases: Phylodynamics, genetic epidemiology, coalescent, Bayesian nonparametrics, birth-death processes, SIR models.

REFERENCES

ALTEKAR, G., DWARKADAS, S., HUELSENBECK, J. P. and RONQUIST, F. (2004). Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics* **20** 407–415.

AL KHATIB, H. A., BENSLIMANE, F. M., ELBASHIR, I. E., COYLE, P. V., AL MASLAMANI, M. A., AL-KHAL, A., AL THANI, A. A. and YASSINE, H. M. (2020). Within-host diversity of Sars-CoV-2 in Covid-19 patients with variable disease severities. *Front. Cell. Infect. Microbiol.* **10** 534. https://doi.org/10.3389/fcimb.2020.575613

ANDERSEN, K. G., RAMBAUT, A., LIPKIN, W. I., HOLMES, E. C. and GARRY, R. F. (2020). The proximal origin of Sars-CoV-2. *Nat. Med.* **26** 450–452. https://doi.org/10.1038/s41591-020-0820-9

ANDRÉOLETTI, J., ZWAANS, A., WARNOCK, R. C. M., AGUIRRE-FERNÁNDEZ, G., BARIDO-SOTTANI, J., GUPTA, A., STADLER, T. and MANCEAU, M. (2020). A skyline birth-death process for inferring the population size from a reconstructed tree with occurrences. *BioRxiv* 2020.10.27.356758. https://doi.org/10.1101/2020.10.27.356758

AYRES, D. L., DARLING, A., ZWICKL, D. J., BEERLI, P., HOLDER, M. T., LEWIS, P. O., HUELSENBECK, J. P., RONQUIST, F., SWOFFORD, D. L. et al. (2012). BEAGLE: An application programming interface and high-performance computing library for statistical phylogenetics. *Syst. Biol.* **61** 170–173.

BARIDO-SOTTANI, J., VAUGHAN, T. G. and STADLER, T. (2020). A multitype birth–death model for Bayesian inference of lineage-specific birth and death rates. *Syst. Biol.* **69** 973–986. https://doi.org/10.1093/sysbio/syaa016

BEAUMONT, M. A., ZHANG, W. and BALDING, D. J. (2002). Approximate Bayesian computation in population genetics. *Genetics* **162** 2025–2035. https://doi.org/10.1093/genetics/162.4.2025

BEHR, M., ANSARI, M. A., MUNK, A. and HOLMES, C. (2020). Testing for dependence on tree structures. *Proc. Natl. Acad. Sci. USA* **117** 9787–9792. MR4236178 https://doi.org/10.1073/pnas.1912957117

BERESTYCKI, N. (2009). *Recent Progress in Coalescent Theory. Ensaios Matemáticos* [*Mathematical Surveys*] **16**. Sociedade Brasileira de Matemática, Rio de Janeiro. MR2574323

BIERKENS, J. (2016). Non-reversible Metropolis-Hastings. *Stat. Comput.* **26** 1213–1228. MR3538633 https://doi.org/10.1007/s11222-015-9598-x

BILLERA, L. J., HOLMES, S. P. and VOGTMANN, K. (2001). Geometry of the space of phylogenetic trees. *Adv. in Appl. Math.* **27** 733–767. MR1867931 https://doi.org/10.1006/aama.2001.0759

BLEI, D. M., KUCUKELBIR, A. and MCAULIFFE, J. D. (2017). Variational inference: A review for statisticians. *J. Amer. Statist. Assoc.* **112** 859–877. MR3671776 https://doi.org/10.1080/01621459.2017.1285773

BLUM, M. G. B. and FRANÇOIS, O. (2010). Non-linear regression models for approximate Bayesian computation. *Stat. Comput.* **20** 63–73. MR2578077 https://doi.org/10.1007/s11222-009-9116-0

BONI, M. F., LEMEY, P., JIANG, X., LAM, T. T.-Y., PERRY, B. W., CASTOE, T. A., RAMBAUT, A. and ROBERTSON, D. L. (2020). Evolutionary origins of the Sars-CoV-2 sarbecovirus lineage re-

Lorenzo Cappello is Assistant Professor, Departments of Economics and Business, Universitat Pompeu Fabra, 08005, Spain (e-mail: lorenzo.cappello@upf.edu). Jaehee Kim is Assistant Professor, Department of Computational Biology, Cornell University, Ithaca, New York 14853, USA (e-mail: jaehee.kim@cornell.edu). Sifan Liu is a Ph.D. student, Department of Statistics, Stanford University, Stanford, California 94305, USA (e-mail: sfliu@stanford.edu). Julia A. Palacios is Assistant Professor, Departments of Statistics and Biomedical Data Sciences, Stanford University, Stanford, California 94305, USA (e-mail: juliapr@stanford.edu).

sponsible for the Covid-19 pandemic. *Nat. Microbiol.* **5** 1408–1417.

BOSKOVA, V., BONHOEFFER, S. and STADLER, T. (2014). Inference of epidemiological dynamics based on simulated phylogenies using birth-death and coalescent models. *PLoS Comput. Biol.* **10** 1–18. https://doi.org/10.1371/journal.pcbi.1003913

BOUCHARD-CÔTÉ, A., SANKARARAMAN, S. and JORDAN, M. I. (2012). Phylogenetic inference via sequential Monte Carlo. *Syst. Biol.* **61** 579–593.

BOUCHARD-CÔTÉ, A., VOLLMER, S. J. and DOUCET, A. (2018a). The bouncy particle sampler: A nonreversible rejection-free Markov chain Monte Carlo method. *J. Amer. Statist. Assoc.* **113** 855–867. MR3832232 https://doi.org/10.1080/01621459.2017.1294075

BOUCHARD-CÔTÉ, A., VOLLMER, S. J. and DOUCET, A. (2018b). The bouncy particle sampler: A nonreversible rejection-free Markov chain Monte Carlo method. *J. Amer. Statist. Assoc.* **113** 855–867. MR3832232 https://doi.org/10.1080/01621459.2017.1294075

BOUCKAERT, R., VAUGHAN, T. G., BARIDO-SOTTANI, J., DUCHÊNE, S., FOURMENT, M., GAVRYUSHKINA, A., HELED, J., JONES, G., KÜHNERT, D. et al. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15** e1006650.

BRITTON, T. (2010). Stochastic epidemic models: A survey. *Math. Biosci.* **225** 24–35. MR2642269 https://doi.org/10.1016/j.mbs.2010.01.006

CAPPELLO, L. and PALACIOS, J. A. (2021). Adaptive preferential sampling in phylodynamics with an application to Sars-CoV-2. *J. Comput. Graph. Statist.* 1–12. https://doi.org/10.1080/10618600.2021.1987256

CAPPELLO, L., VEBER, A. and PALACIOS, J. A. (2020). The Tajima heterochronous n-coalescent: Inference from heterochronously sampled molecular data. Preprint. Available at arXiv:2004.06826.

CAPPELLO, L., KIM, J., LIU, S. and PALACIOS, J. A. (2022). Supplement to "Statistical Challenges in Tracking the Evolution of SARS-CoV-2." https://doi.org/10.1214/22-STS853SUPP

CHOI, S. C. (2020). A phylodynamic analysis of epidemiological situation of East Asia due to the coronavirus disease of 2019. *The Microbiological Society of Korea* **56** 241–253.

CHOPIN, N. and PAPASPILIOPOULOS, O. (2020). *An Introduction to Sequential Monte Carlo*. *Springer Series in Statistics*. Springer, Cham. MR4215639 https://doi.org/10.1007/978-3-030-47845-2

COLIJN, C. and PLAZZOTTA, G. (2017). A metric on phylogenetic tree shapes. *Syst. Biol.* **67** 113–126. https://doi.org/10.1093/sysbio/syx046

CYBIS, G. B., SINSHEIMER, J. S., BEDFORD, T., MATHER, A. E., LEMEY, P. and SUCHARD, M. A. (2015). Assessing phenotypic correlation through the multivariate phylogenetic latent liability model. *Ann. Appl. Stat.* **9** 969–991. MR3371344 https://doi.org/10.1214/15-AOAS821

DAVIES, N. G., ABBOTT, S., BARNARD, R. C., JARVIS, C. I., KUCHARSKI, A. J., MUNDAY, J. D., PEARSON, C. A., RUSSELL, T. W., TULLY, D. C. et al. (2021). Estimated transmissibility and impact of Sars-CoV-2 lineage B. 1.1. 7 in England. *Science* **372**.

DEARLOVE, B. and WILSON, D. J. (2013). Coalescent inference for infectious disease: Meta-analysis of hepatitis C. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* **368** 20120314. https://doi.org/10.1098/rstb.2012.0314

DELLICOUR, S., DURKIN, K., HONG, S. L., VANMECHELEN, B., MARTÍ-CARRERAS, J., GILL, M. S., MEEX, C., BONTEMS, S., ANDRÉ, E. et al. (2021). A phylodynamic workflow to rapidly gain insights into the dispersal history and dynamics of Sars-CoV-2 lineages. *Mol. Biol. Evol.* **38** 1608–1613.

DENG, X., GU, W., FEDERMAN, S., DU PLESSIS, L., PYBUS, O. G., FARIA, N. R., WANG, C., YU, G., BUSHNELL, B. et al. (2020). Genomic surveillance reveals multiple introductions of Sars-CoV-2 into northern California. *Science* **369** 582–587.

DIGGLE, P. J., MENEZES, R. and SU, T. (2010). Geostatistical inference under preferential sampling. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **59** 191–232. MR2744471 https://doi.org/10.1111/j.1467-9876.2009.00701.x

DINH, V., DARLING, A. E. and IV, F. A. M. (2018). Online Bayesian phylogenetic inference: Theoretical foundations via sequential Monte Carlo. *Syst. Biol.* **67** 503–517. https://doi.org/10.1093/sysbio/syx087

DRUMMOND, A. and RODRIGO, A. G. (2000). Reconstructing genealogies of serial samples under the assumption of a molecular clock using serial-sample UPGMA. *Mol. Biol. Evol.* **17** 1807–1815.

DU PLESSIS, L., MCCRONE, J. T., ZAREBSKI, A. E., HILL, V., RUIS, C., GUTIERREZ, B., RAGHWANI, J., ASHWORTH, J., COLQUHOUN, R. et al. (2021). Establishment and lineage dynamics of the Sars-CoV-2 epidemic in the UK. *Science* **371** 708–712.

EFRON, B., HALLORAN, E. and HOLMES, S. (1996). Bootstrap confidence levels for phylogenetic trees. *Proc. Natl. Acad. Sci. USA* **93** 7085–7090.

FAULKNER, J. R., MAGEE, A. F., SHAPIRO, B. and MININ, V. N. (2020). Horseshoe-based Bayesian nonparametric estimation of effective population size trajectories. *Biometrics* (in press).

FEATHERSTONE, L. A., DI GIALLONARDO, F., HOLMES, E. C., VAUGHAN, T. G. and DUCHÊNE, S. (2021). Infectious disease phylodynamics with occurrence data. *Methods Ecol. Evol.* **12** 1498–1507. https://doi.org/10.1111/2041-210X.13620

FELSENSTEIN, J. (1985). Phylogenies and the comparative method. *Amer. Nat.* **125** 1–15.

FELSENSTEIN, J. (2004). *Inferring Phylogenies* **2**. Sinauer associates Sunderland, MA.

FELSENSTEIN, J. (2005). Using the quantitative genetic threshold model for inferences between and within species. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* **360** 1427–1434.

FELSENSTEIN, J. (2012). A comparative method for both discrete and continuous characters using the threshold model. *Amer. Nat.* **179** 145–156.

FELSENSTEIN, J. and RODRIGO, A. G. (1999). Coalescent approaches to HIV population genetics. In *The Evolution of HIV* 233–272. Johns Hopkins Univ. Press, Baltimore.

FITCH, W. M. (1971). Toward defining the course of evolution: Minimum change for a specific tree topology. *Syst. Biol.* **20** 406–416.

FOURMENT, M., CLAYWELL, B. C., DINH, V., MCCOY, C., IV, F. A. M. and DARLING, A. E. (2018). Effective online Bayesian phylogenetics via sequential Monte Carlo with guided proposals. *Syst. Biol.* **67** 490–502. https://doi.org/10.1093/sysbio/syx090

FRICK, K., MUNK, A. and SIELING, H. (2014). Multiscale change point inference. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76** 495–580. MR3210728 https://doi.org/10.1111/rssb.12047

FROST, S. D. W., PYBUS, O. G., GOG, J. R., VIBOUD, C., BONHOEFFER, S. and BEDFORD, T. (2015). Eight challenges in phylodynamic inference. *Epidemics* **10** 88–92. https://doi.org/10.1016/j.epidem.2014.09.001

GEIDELBERG, L., BOYD, O., JORGENSEN, D., SIVERONI, I., NASCIMENTO, F. F., JOHNSON, R., RAGONNET-CRONIN, M., FU, H., WANG, H. et al. (2021). Genomic epidemiology of a densely sampled Covid-19 outbreak in China. *Virus Evolution* **7**. veaa102. https://doi.org/10.1093/ve/veaa102

GERNHARD, T. (2008). The conditioned reconstructed process. *J. Theoret. Biol.* **253** 769–778. MR2964590 https://doi.org/10.1016/j.jtbi.2008.04.005

GILL, M. S., LEMEY, P., SUCHARD, M. A., RAMBAUT, A. and BAELE, G. (2020). Online Bayesian phylodynamic inference in BEAST with application to epidemic reconstruction. *Mol. Biol. Evol.* **37** 1832–1842. https://doi.org/10.1093/molbev/msaa047

GRAFEN, A. (1989). The phylogenetic regression. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* **326** 119–157.

GRENFELL, B. T., PYBUS, O. G., GOG, J. R., WOOD, J. L. N., DALY, J. M., MUMFORD, J. A. and HOLMES, E. C. (2004). Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* **303** 327–332.

GRIFFITHS, R. C. and TAVARE, S. (1994). Sampling theory for neutral alleles in a varying environment. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* **344** 403–410.

GRUBAUGH, N. D., GANGAVARAPU, K., QUICK, J., MATTESON, N. L., DE JESUS, J. G., MAIN, B. J., TAN, A. L., PAUL, L. M., BRACKNEY, D. E. et al. (2019). An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol.* **20** 1–19.

GUPTA, A., MANCEAU, M., VAUGHAN, T., KHAMMASH, M. and STADLER, T. (2020). The probability distribution of the reconstructed phylogenetic tree with occurrence data. *J. Theoret. Biol.* **488** 110115, 10. MR4051870 https://doi.org/10.1016/j.jtbi.2019.110115

HADFIELD, J., MEGILL, C., BELL, S. M., HUDDLESTON, J., POTTER, B., CALLENDER, C., SAGULENKO, P., BEDFORD, T. and NEHER, R. A. (2018). Nextstrain: Real-time tracking of pathogen evolution. *Bioinformatics* **34** 4121–4123. https://doi.org/10.1093/bioinformatics/bty407

HASEGAWA, M., KISHINO, H. and YANO, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **2** 160–164.

HELED, J. and DRUMMOND, A. J. (2008). Bayesian inference of population size history from multiple loci. *BMC Evol. Biol.* **8** 289. https://doi.org/10.1186/1471-2148-8-289

HODCROFT, E. B., ZUBER, M. et al. (2021). Spread of a Sars-CoV-2 variant through Europe in the summer of 2020. *Nature* **595** 707–712. https://doi.org/10.1038/s41586-021-03677-y

HOFFMAN, M. D., BLEI, D. M., WANG, C. and PAISLEY, J. (2013). Stochastic variational inference. *J. Mach. Learn. Res.* **14** 1303–1347. MR3081926

HUANG, Z. and GELMAN, A. (2005). Sampling for Bayesian computation with large datasets. Available at SSRN 1010107.

HUELSENBECK, J. P. and RANNALA, B. (2003). Detecting correlation between characters in a comparative analysis with uncertain phylogeny. *Evolution* **57** 1237–1247.

JONES, B. R., KINLOCH, N. N., HORACSEK, J., GANASE, B., HARRIS, M., HARRIGAN, P. R., JONES, R. B., BROCKMAN, M. A., JOY, J. B. et al. (2018). Phylogenetic approach to recover integration dates of latent HIV sequences within-host. *Proc. Natl. Acad. Sci. USA* **115** E8958–E8967. https://doi.org/10.1073/pnas.1802028115

JORDAN, M. I., GHAHRAMANI, Z., JAAKKOLA, T. S. and SAUL, L. K. (1999). An introduction to variational methods for graphical models. *Mach. Learn.* **37** 183–233.

KARCHER, M. D., PALACIOS, J. A., BEDFORD, T., SUCHARD, M. A. and MININ, V. N. (2016). Quantifying and mitigating the effect of preferential sampling on phylodynamic inference. *PLoS Comput. Biol.* **12** e1004789.

KARCHER, M. D., SUCHARD, M. A., DUDAS, G. and MININ, V. N. (2020). Estimating effective population size changes from preferentially sampled genetic sequences. *PLoS Comput. Biol.* **in press**.

KERMACK, W. O., MCKENDRICK, A. G. and WALKER, G. T. (1927). A contribution to the mathematical theory of epidemics. *Proc. R. Soc. Lond. Ser. A, Contain. Pap. Math. Phys. Character* **115** 700–721. https://doi.org/10.1098/rspa.1927.0118

KIM, J., ROSENBERG, N. A. and PALACIOS, J. A. (2020). Distance metrics for ranked evolutionary trees. *Proc. Natl. Acad. Sci. USA* **117** 28876–28886. https://doi.org/10.1073/pnas.1922851117

KINGMAN, J. F. C. (1982a). The coalescent. *Stochastic Process. Appl.* **13** 235–248. MR0671034 https://doi.org/10.1016/0304-4149(82)90011-4

KINGMAN, J. F. C. (1982b). On the genealogy of large populations. *J. Appl. Probab.* **19A** 27–43. MR0633178

KOELLE, K. and RASMUSSEN, D. A. (2012). Rates of coalescence for common epidemiological models at equilibrium. *J. R. Soc. Interface* **9** 997–1007. https://doi.org/10.1098/rsif.2011.0495

KÜHNERT, D., STADLER, T., VAUGHAN, T. G. and DRUMMOND, A. J. (2016). Phylodynamics with migration: A computational framework to quantify population structure from genomic data. *Mol. Biol. Evol.* **33** 2102–2116. https://doi.org/10.1093/molbev/msw064

LAN, S., PALACIOS, J. A., KARCHER, M., MININ, V. N. and SHAHBABA, B. (2015). An efficient Bayesian inference framework for coalescent-based nonparametric phylodynamics. *Bioinformatics* **31** 3282–3289.

LEMEY, P., HONG, S. L., HILL, V., BAELE, G., POLETTO, C., COLIZZA, V., O'TOOLE, Á., MCCRONE, J. T., ANDERSEN, K. G. et al. (2020). Accommodating individual travel history and unsampled diversity in Bayesian phylogeographic inference of Sars-CoV-2. *Nat. Commun.* **11** 1–14.

LEMEY, P., RUKTANONCHAI, N., HONG, S. L., COLIZZA, V., POLETTO, C., DEN BROECK, F. V., GILL, M. S., JI, X., LEVASSEUR, A. et al. (2021). Untangling introductions and persistence in Covid-19 resurgence in Europe. *Nature* **595** 713–717. https://doi.org/10.1038/s41586-021-03754-2

LEUNG, K., SHUM, M. H., LEUNG, G. M., LAM, T. T. and WU, J. T. (2021). Early transmissibility assessment of the N501Y mutant strains of Sars-CoV-2 in the United Kingdom, October to November 2020. *Euro Surveill.* **26** 2002106.

LEWIS, P. O., XIE, W., CHEN, M.-H., FAN, Y. and KUO, L. (2014). Posterior predictive Bayesian phylogenetic model selection. *Syst. Biol.* **63** 309–321.

LI, L. M., GRASSLY, N. C. and FRASER, C. (2017). Quantifying transmission heterogeneity using both pathogen phylogenies and incidence time series. *Mol. Biol. Evol.* **34** 2982–2995. https://doi.org/10.1093/molbev/msx195

LINTUSAARI, J., GUTMANN, M. U., DUTTA, R., KASKI, S. and CORANDER, J. (2016). Fundamentals and recent developments in approximate Bayesian computation. *Syst. Biol.* **66** e66–e82. https://doi.org/10.1093/sysbio/syw077

LOPEZ BERNAL, J., ANDREWS, N., GOWER, C., GALLAGHER, E., SIMMONS, R., THELWALL, S., STOWE, J., TESSIER, E., GROVES, N. 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

LOUCA, S., MCLAUGHLIN, A., MACPHERSON, A., JOY, J. B. and PENNELL, M. W. (2021). Fundamental identifiability limits in molecular epidemiology. *Mol. Biol. Evol.* **38** 4010–4024. https://doi.org/10.1093/molbev/msab149

MACCANNELL, T., BATSON, J., BONIN, B., KC, A., QUENELLE, R., STRONG, B., LIN, W., RUDMAN, S. L., DYNERMAN, D. et al. (2021). Genomic epidemiology and transmission dynamics of Sars-CoV-2 in congregate healthcare facilities in Santa Clara County, California. *Clin. Infect. Dis.*.

MACPHERSON, A., LOUCA, S., MCLAUGHLIN, A., JOY, J. B. and PENNELL, M. W. (2021). Unifying phylogenetic birth-death models in epidemiology and macroevolution. *Syst. Biol.*. syab049. https://doi.org/10.1093/sysbio/syab049

MANCEAU, M., GUPTA, A., VAUGHAN, T. and STADLER, T. (2021). The probability distribution of the ancestral population size conditioned on the reconstructed phylogenetic tree with occurrence data. *J. Theoret. Biol.* **509** Paper No. 110400, 18. MR4171763 https://doi.org/10.1016/j.jtbi.2020.110400

MARJORAM, P. and TAVARÉ, S. (2006). Modern computational approaches for analysing molecular genetic variation data. *Nat. Rev. Genet.* **7** 759–770.

MARJORAM, P., MOLITOR, J., PLAGNOL, V. and TAVARÉ, S. (2003). Markov chain Monte Carlo without likelihoods. *Proc. Natl. Acad. Sci. USA* **100** 15324–15328. https://doi.org/10.1073/pnas.0306899100

MAURANO, M. T., RAMASWAMI, S., ZAPPILE, P., DIMARTINO, D., BOYTARD, L., RIBEIRO-DOS SANTOS, A. M., VULPESCU, N. A., WESTBY, G., SHEN, G. et al. (2020). Sequencing identifies multiple early introductions of Sars-CoV-2 to the New York City region. *Genome Res.* **30** 1781–1788.

MILLER, D., MARTIN, M. A., HAREL, N., TIROSH, O., KUSTIN, T., MEIR, M., SOREK, N., GEFEN-HALEVI, S., AMIT, S. et al. (2020). Full genome viral sequences inform patterns of Sars-CoV-2 spread into and within Israel. *Nat. Commun.* **11** 5518. https://doi.org/10.1038/s41467-020-19248-0

MINH, B. Q., SCHMIDT, H. A., CHERNOMOR, O., SCHREMPF, D., WOODHAMS, M. D., VON HAESELER, A. and LANFEAR, R. (2020). IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37** 1530–1534.

MININ, V. N., BLOOMQUIST, E. W. and SUCHARD, M. A. (2008). Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. *Mol. Biol. Evol.* **25** 1459–1471.

MOREL, B., BARBERA, P., CZECH, L., BETTISWORTH, B., HÜBNER, L., LUTTEROPP, S., SERDARI, D., KOSTAKI, E.-G., MAMAIS, I. et al. (2021). Phylogenetic analysis of Sars-CoV-2 data is difficult. *Mol. Biol. Evol.* **38** 1777–1791.

MORENO, G. K., BRAUN, K. M., RIEMERSMA, K. K., MARTIN, M. A., HALFMANN, P. J., CROOKS, C. M., PRALL, T., BAKER, D., BACZENAS, J. J. et al. (2020). Revealing fine-scale spatiotemporal differences in Sars-CoV-2 introduction and spread. *Nat. Commun.* **11** 5558. https://doi.org/10.1038/s41467-020-19346-z

MÜLLER, N. F. and BOUCKAERT, R. R. (2020). Adaptive Metropolis-coupled MCMC for BEAST 2. *PeerJ* **8** e9473. https://doi.org/10.7717/peerj.9473

MÜLLER, N. F., KISTLER, K. E. and BEDFORD, T. (2022). Recombination patterns in coronaviruses. *BioRxiv*. https://doi.org/10.1101/2021.04.28.441806

MÜLLER, N. F., RASMUSSEN, D. A. and STADLER, T. (2017). The structured coalescent and its approximations. *Mol. Biol. Evol.* **34** 2970–2981. https://doi.org/10.1093/molbev/msx186

MÜLLER, N. F., WAGNER, C., FRAZAR, C. D., ROYCHOUDHURY, P., LEE, J., MONCLA, L. H., PELLE, B., RICHARDSON, M., RYKE, E. et al. (2021). Viral genomes reveal patterns of the Sars-CoV-2 outbreak in Washington state. *Sci. Transl. Med.* **13**.

NADEAU, S. A., VAUGHAN, T. G., SCIRE, J., HUISMAN, J. S. and STADLER, T. (2021). The origin and early spread of Sars-CoV-2 in Europe. *Proc. Natl. Acad. Sci. USA* **118**. https://doi.org/10.1073/pnas.2012008118

NEAL, R. M. (2001). Annealed importance sampling. *Stat. Comput.* **11** 125–139. MR1837132 https://doi.org/10.1023/A:1008923215028

NEE, S., MAY, R. M. and HARVEY, P. H. (1994). The reconstructed evolutionary process. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* **344** 305–311.

NEISWANGER, W., WANG, C. and XING, E. (2013). Asymptotically exact, embarrassingly parallel MCMC. Preprint. Available at arXiv:1311.4780.

PAGEL, M. (1994). Detecting correlated evolution on phylogenies: A general method for the comparative analysis of discrete characters. *Proc. R. Soc. Lond., B Biol. Sci.* **255** 37–45.

PALACIOS, J. A. and MININ, V. N. (2012). Integrated nested Laplace approximation for Bayesian nonparametric phylodynamics. In *Proceedings of the Twenty-Eighth Conference on Uncertainty in Artificial Intelligence. UAI'*12 726–735. AUAI Press, Arlington, VA, United States.

PALACIOS, J. A. and MININ, V. N. (2013). Gaussian process-based Bayesian nonparametric inference of population size trajectories from gene genealogies. *Biometrics* **69** 8–18. MR3058047 https://doi.org/10.1111/biom.12003

PALACIOS, J. A., GILL, M. S., SUCHARD, M. A. and MININ, V. N. (2014). Bayesian nonparametric phylodynamics.

PALACIOS, J. A., VÉBER, A., CAPPELLO, L., WANG, Z., WAKELEY, J. and RAMACHANDRAN, S. (2019). Bayesian estimation of population size changes by sampling Tajima's trees. *Genetics* **213** 967–986.

PARAG, K. V., DU PLESSIS, L. and PYBUS, O. G. (2020). Jointly inferring the dynamics of population size and sampling intensity from molecular sequences. *Mol. Biol. Evol.* **37** 2414–2429.

PARIKH, V. N., IOANNIDIS, A., JIMENEZ-MORALES, D. et al. (2021). Multi-omic surveillance disambiguates social and biological determinants of COVID19 severity. *in Preparation*.

PARKER, J., RAMBAUT, A. and PYBUS, O. G. (2008). Correlating viral phenotypes with phylogeny: Accounting for phylogenetic uncertainty. *Infect. Genet. Evol.* **8** 239–246.

POON, A. F. Y. (2015). Phylodynamic inference with kernel ABC and its application to HIV epidemiology. *Mol. Biol. Evol.* **32** 2483–2495. https://doi.org/10.1093/molbev/msv123

POPINGA, A., VAUGHAN, T., STADLER, T. and DRUMMOND, A. J. (2015). Inferring epidemiological dynamics with Bayesian coalescent inference: The merits of deterministic and stochastic models. *Genetics* **199** 595–607. https://doi.org/10.1534/genetics.114.172791

RAGONNET-CRONIN, M., BOYD, O., GEIDELBERG, L., JORGENSEN, D., NASCIMENTO, F. F., SIVERONI, I., JOHNSON, R. A., BAGUELIN, M., CUCUNUBÁ, Z. M. et al. (2021). Genetic evidence for the association between Covid-19 epidemic severity and timing of non-pharmaceutical interventions. *Nat. Commun.* **12** 2188. https://doi.org/10.1038/s41467-021-22366-y

RAJANALA, S. and PALACIOS, J. A. (2021). Statistical summaries of unlabelled evolutionary trees and ranked hierarchical clustering trees. Preprint. Available at arXiv:2106.02724.

RASMUSSEN, D. A., RATMANN, O. and KOELLE, K. (2011). Inference for nonlinear epidemiological models using genealogies and time series. *PLoS Comput. Biol.* **7** e1002136, 11. MR2845064 https://doi.org/10.1371/journal.pcbi.1002136

RATMANN, O., DONKER, G., MEIJER, A., FRASER, C. and KOELLE, K. (2012). Phylodynamic inference and model assessment with approximate Bayesian computation: Influenza as a case study. *PLoS Comput. Biol.* **8** 1–14. https://doi.org/10.1371/journal.pcbi.1002835

ROBINSON, D. F. and FOULDS, L. R. (1981). Comparison of phylogenetic trees. *Math. Biosci.* **53** 131–147. MR0613619 https://doi.org/10.1016/0025-5564(81)90043-2

RODRIGO, A. G. and FELSENSTEIN, J. (1999). Coalescent approaches to HIV-1 population genetics. In *Molecular Evolution of HIV* (K. A. Crandell, ed.) Johns Hopkins Univ. Press, Baltimore, MD.

ROSENBERG, N. A. and NORDBORG, M. (2002). Genealogical trees, coalescent theory and the analysis of genetic polymorphisms. *Nat. Rev. Genet.* **3** 380–390.

RUE, H., MARTINO, S. and CHOPIN, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 319–392. MR2649602 https://doi.org/10.1111/j.1467-9868.2008.00700.x

SAGULENKO, P., PULLER, V. and NEHER, R. A. (2018). TreeTime: Maximum-likelihood phylodynamic analysis. *Virus Evol.* **4** vex042. https://doi.org/10.1093/ve/vex042

SAINUDIIN, R., STADLER, T. and VÉBER, A. (2015). Finding the best resolution for the Kingman-Tajima coalescent: Theory and applications. *J. Math. Biol.* **70** 1207–1247. MR3323594 https://doi.org/10.1007/s00285-014-0796-5

SAN, J. E., NGCAPU, S., KANZI, A. M., TEGALLY, H., FONSECA, V., GIANDHARI, J., WILKINSON, E., NELSON, C. W., SMIDT, W. et al. (2021). Transmission dynamics of Sars-CoV-2 within-host diversity in two major hospital outbreaks in South Africa. *Virus Evolution* **7**. veab041. https://doi.org/10.1093/ve/veab041

SAULNIER, E., GASCUEL, O. and ALIZON, S. (2017). Inferring epidemiological parameters from phylogenies using regression-ABC: A comparative study. *PLoS Comput. Biol.* **13** 1–31. https://doi.org/10.1371/journal.pcbi.1005416

SCIRE, J., BARIDO-SOTTANI, J., KÜHNERT, D., VAUGHAN, T. G. and STADLER, T. (2020). Improved multi-type birth-death phylodynamic inference in BEAST 2. *BioRxiv*.

SEEMANN, T., LANE, C. R., SHERRY, N. L., DUCHENE, S., GONÇALVES DA SILVA, A., CALY, L., SAIT, M., BALLARD, S. A., HORAN, K. et al. (2020). Tracking the Covid-19 pandemic in Australia using genomics. *Nat. Commun.* **11** 4376. https://doi.org/10.1038/s41467-020-18314-x

SHU, Y. and MCCAULEY, J. (2017). GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill.* **22**. https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494

SIMPER, M. and PALACIOS, J. A. (2020). An adjacent-swap Markov chain on coalescent trees. Preprint. Available at arXiv:2012.08030.

SLATKIN, M. and HUDSON, R. R. (1991). Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* **129** 555–562.

SRIVASTAVA, S., CEVHER, V., DINH, Q. and DUNSON, D. (2015). WASP: Scalable Bayes via barycenters of subset posteriors. In *Artificial Intelligence and Statistics* 912–920. PMLR.

STADLER, T. (2009). On incomplete sampling under birth-death models and connections to the sampling-based coalescent. *J. Theoret. Biol.* **261** 58–66. MR2980272 https://doi.org/10.1016/j.jtbi.2009.07.018

STADLER, T. (2010). Sampling-through-time in birth-death trees. *J. Theoret. Biol.* **267** 396–404. MR2974417 https://doi.org/10.1016/j.jtbi.2010.09.010

STADLER, T. and BONHOEFFER, S. (2013). Uncovering epidemiological dynamics in heterogeneous host populations using phylogenetic methods. *Philos. Trans. R. Soc. Lond. B, Biol. Sci.* **368** 20120198. https://doi.org/10.1098/rstb.2012.0198

STADLER, T., KÜHNERT, D., BONHOEFFER, S. and DRUMMOND, A. J. (2013). Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proc. Natl. Acad. Sci. USA* **110** 228–233. https://doi.org/10.1073/pnas.1207965110

SUCHARD, M. A., LEMEY, P., BAELE, G., AYRES, D. L., DRUMMOND, A. J. and RAMBAUT, A. (2018). Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4** vey016. https://doi.org/10.1093/ve/vey016

TANG, M., DUDAS, G., BEDFORD, T. and MININ, V. N. (2019). Fitting stochastic epidemic models to gene genealogies using linear noise approximation. Preprint. Available at arXiv:1902.08877 [q-bio.PE].

TAVARÉ, S. (2004). *Ancestral Inference in Population Genetics. Lectures on Probability Theory and Statistics: Ecole D'Eté de Probabilités de Saint-Flour XXXI*-2001. Springer, Berlin.

THOMPSON, E. A. (1975). *Human Evolutionary Trees*. Cambridge Univ. Press, Cambridge.

THORNLOW, B., YE, C., DE MAIO, N., MCBROOME, J., HINRICHS, A. S., LANFEAR, R., TURAKHIA, Y. and CORBETT-DETIG, R. (2021). Online phylogenetics using parsimony produces slightly better trees and is dramatically more efficient for large SARS-CoV-2 phylogenies than de novo and maximum-likelihood approaches. BioRxiv.

TILLETT, R. L., SEVINSKY, J. R., HARTLEY, P. D., KERWIN, H., CRAWFORD, N., GORZALSKI, A., LAVERDURE, C., VERMA, S. C., ROSSETTO, C. C. et al. (2021). Genomic evidence for reinfection with Sars-CoV-2: A case study. *Lancet Infect. Dis.* **21** 52–58. https://doi.org/10.1016/S1473-3099(20)30764-7

TRUCCHI, E., GRATTON, P., MAFESSONI, F., MOTTA, S., CICCONARDI, F., MANCIA, F., BERTORELLE, G., D'ANNESSA, I. and DI MARINO, D. (2021). Population dynamics and structural effects at short and long range support the hypothesis of the selective advantage of the G614 Sars-CoV-2 spike variant. *Mol. Biol. Evol.* **38** 1966–1979.

TURAKHIA, Y., DE MAIO, N., THORNLOW, B., GOZASHTI, L., LANFEAR, R., WALKER, C. R., HINRICHS, A. S., FERNANDES, J. D., BORGES, R. et al. (2020). Stability of Sars-CoV-2 phylogenies. *PLoS Genet.* **16** e1009175.

VAN DORP, L., RICHARD, D., TAN, C. C., SHAW, L. P., ACMAN, M. and BALLOUX, F. (2020). No evidence for increased transmissibility from recurrent mutations in Sars-CoV-2. *Nat. Commun.* **11** 1–8.

VAUGHAN, T. G., LEVENTHAL, G. E., RASMUSSEN, D. A., DRUMMOND, A. J., WELCH, D. and STADLER, T. (2019). Estimating epidemic incidence and prevalence from genomic data. *Mol. Biol. Evol.* **36** 1804–1816. https://doi.org/10.1093/molbev/msz106

VOLZ, E. M. (2012). Complex population dynamics and the coalescent under neutrality. *Genetics* **190** 187–201. https://doi.org/10.1534/genetics.111.134627

VOLZ, E. M. and FROST, S. D. W. (2014). Sampling through time and phylodynamic inference with coalescent and birth & death models. *J. R. Soc. Interface* **11** 20140945. https://doi.org/10.1098/rsif.2014.0945

VOLZ, E. M., KOELLE, K. and BEDFORD, T. (2013). Viral phylodynamics. *PLoS Comput. Biol.* **9** e1002947, 12. MR3048921 https://doi.org/10.1371/journal.pcbi.1002947

VOLZ, E. M., ROMERO-SEVERSON, E. and LEITNER, T. (2017). Phylodynamic inference across epidemic scales. *Mol. Biol. Evol.* **34** 1276–1288. https://doi.org/10.1093/molbev/msx077

VOLZ, E. M. and SIVERONI, I. (2018). Bayesian phylodynamic inference with complex models. *PLoS Comput. Biol.* **14** 1–15. https://doi.org/10.1371/journal.pcbi.1006546

VOLZ, E. M., KOSAKOVSKY POND, S. L., WARD, M. J., LEIGH BROWN, A. J. and FROST, S. D. W. (2009). Phylodynamics of infectious disease epidemics. *Genetics* **183** 1421–1430. https://doi.org/10.1534/genetics.109.106021

VOLZ, E., BAGUELIN, M., BHATIA, S., BOONYASIRI, A., CORI, A., CUCUNUBÁ, Z., CUOMO-DANNENBURG, G., DONNELLY, C. A., DORIGATTI, I. et al. (2020). Phylogenetic analysis of SARS-CoV-2. Imperial College London (15-02-2020). https://doi.org/10.25561/77169

VOLZ, E., HILL, V., MCCRONE, J. T., PRICE, A., JORGENSEN, D., O'TOOLE, Á., SOUTHGATE, J., JOHNSON, R., JACKSON, B.

et al. (2021a). Evaluating the effects of Sars-CoV-2 spike mutation D614G on transmissibility and pathogenicity. *Cell* **184** 64–75.

VOLZ, E., MISHRA, S., CHAND, M., BARRETT, J. C., JOHNSON, R., GEIDELBERG, L., HINSLEY, W. R., LAYDON, D. J., DABRERA, G. et al. (2021b). Assessing transmissibility of Sars-CoV-2 lineage B. 1.1. 7 in England. *Nature* **593** 266–269.

VOZNICA, J., ZHUKOVA, A., BOSKOVA, V., SAULNIER, E., LEMOINE, F., MOSLONKA-LEFEBVRE, M. and GASCUEL, O. (2021). Deep learning from phylogenies to uncover the transmission dynamics of epidemics. *BioRxiv*. https://doi.org/10.1101/2021.03.11.435006

WAKELEY, J. (2009). *Coalescent Theory*: *An Introduction*. Roberts and Co, Greenwood Village, CO.

WAKELEY, J. (2020). Developments in coalescent theory from single loci to chromosomes. *Theor. Popul. Biol.* **133** 56–64. https://doi.org/10.1016/j.tpb.2020.02.002

WAKELEY, J. and ALIACAR, N. (2001). Gene genealogies in a metapopulation *Genetics* **159** 893–905.

WAKELEY, J. and SARGSYAN, O. (2009). Extensions of the coalescent effective population size. *Genetics* **181** 341–345.

WANG, L., BOUCHARD-CÔTÉ, A. and DOUCET, A. (2015). Bayesian phylogenetic inference using a combinatorial sequential Monte Carlo method. *J. Amer. Statist. Assoc.* **110** 1362–1374. MR3449032 https://doi.org/10.1080/01621459.2015.1054487

WANG, L., WANG, S. and BOUCHARD-CÔTÉ, A. (2020). An annealed sequential Monte Carlo method for Bayesian phylogenetics. *Syst. Biol.* **69** 155–183.

WANG, T. H., DONALDSON, Y. K., BRETTLE, R. P., BELL, J. E. and SIMMONDS, P. (2001). Identification of shared populations of human immunodeficiency virus type 1 infecting microglia and tissue macrophages outside the central nervous system. *J. Virol.* **75** 11686–11699. https://doi.org/10.1128/JVI.75.23.11686-11699.2001

WHIDDEN, C. and MATSEN IV, F. A. (2015). Quantifying MCMC exploration of phylogenetic tree space. *Syst. Biol.* **64** 472–491.

WU, S. L., MERTENS, A. N., CRIDER, Y. S., NGUYEN, A., POKPONGKIAT, N. N., DJAJADI, S., SETH, A., HSIANG, M. S., COLFORD, J. M. et al. (2020). Substantial underestimation of Sars-CoV-2 infection in the United States. *Nat. Commun.* **11** 1–10.

YANG, Z. (2014). *Molecular Evolution*: *A Statistical Approach*. Oxford Univ. Press, London.

ZAREBSKI, A. E., DU PLESSIS, L., PARAG, K. V. and PYBUS, O. G. (2021). A computationally tractable birth-death model that combines phylogenetic and epidemiological data. *BioRxiv*. https://doi.org/10.1101/2020.10.21.349068

ZHANG, C. and MATSEN IV, F. A. (2019). Variational Bayesian phylogenetic inference. In *International Conference on Learning Representations*.

ZHANG, X., TAN, Y., LING, Y., LU, G., LIU, F., YI, Z., JIA, X., WU, M., SHI, B. et al. (2020). Viral and host factors related to the clinical outcome of Covid-19. *Nature* **583** 437–440.

ZHANG, Z., NISHIMURA, A., BASTIDE, P., JI, X., PAYNE, R. P., GOULDER, P., LEMEY, P. and SUCHARD, M. A. (2021). Large-scale inference of correlation among mixed-type biological traits with phylogenetic multivariate probit models. *Ann. Appl. Stat.* **15** 230–251. MR4255276 https://doi.org/10.1214/20-aoas1394

ZHOU, D., DEJNIRATTISAI, W., SUPASA, P., LIU, C., MENTZER, A. J., GINN, H. M., ZHAO, Y., DUYVESTEYN, H. M. E., TUEKPRAKHON, A. et al. (2021). Evidence of escape of Sars-CoV-2 variant B.1.351 from natural and vaccine-induced sera. *Cell* **184** 2348–2361.e6. https://doi.org/10.1016/j.cell.2021.02.037

Statistical Science
2022, Vol. 37, No. 2, 183–206
https://doi.org/10.1214/22-STS854

Interoperability of Statistical Models in Pandemic Preparedness: Principles and Reality

George Nicholson, Marta Blangiardo, Mark Briers, Peter J. Diggle, Tor Erlend Fjelde, Hong Ge, Robert J. B. Goudie, Radka Jersakova, Ruairidh E. King, Brieuc C. L. Lehmann, Ann-Marie Mallon, Tullia Padellini, Yee Whye Teh, Chris Holmes and Sylvia Richardson

Abstract. We present *interoperability* as a guiding framework for statistical modelling to assist policy makers asking multiple questions using diverse datasets in the face of an evolving pandemic response. Interoperability provides an important set of principles for future pandemic preparedness, through the joint design and deployment of adaptable systems of statistical models for disease surveillance using probabilistic reasoning. We illustrate this through case studies for inferring and characterising spatial-temporal prevalence and reproduction numbers of SARS-CoV-2 infections in England.

Key words and phrases: Bayesian graphical models, Bayesian melding, COVID-19, evidence synthesis, interoperability, modularization, multisource inference.

REFERENCES

[1] ADES, A. E. and SUTTON, A. J. (2006). Multiparameter evidence synthesis in epidemiology and medical decision-making: Current approaches. *J. Roy. Statist. Soc. Ser. A* **169** 5–35. MR2222010 https://doi.org/10.1111/j.1467-985X.2005.00377.x

[2] ANDERSON, R., DONNELLY, C., HOLLINGSWORTH, D., KEELING, M., VEGVARI, C., BAGGALEY, R. and MADDREN, R. (2020). Reproduction number (R) and growth rate (r) of the COVID-19 epidemic in the UK: Methods of estimation, data sources, causes of heterogeneity, and use as a guide in policy formulation Technical Report London, UK: Royal Society.

[3] BI, Q. et al. (2020). Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: A retrospective cohort study. *Lancet Infect. Dis.* **20** 911–919.

[4] BIRRELL, P., BLAKE, J., VAN LEEUWEN, E., GENT, N. and DE ANGELIS, D. (2021). Real-time nowcasting and forecasting of COVID-19 dynamics in England: The first wave. *Philos. Trans. - R. Soc., Biol. Sci.* **376** 20200279. https://doi.org/10.1098/rstb.2020.0279

[5] BRACHER, J., RAY, E. L., GNEITING, T. and REICH, N. G. (2021). Evaluating epidemic forecasts in an interval format. *PLoS Comput. Biol.* **17** e1008618. https://doi.org/10.1371/journal.pcbi.1008618

[6] BRAUER, F., VAN DEN DRIESSCHE, P. and WU, J., eds. (2008) In *Mathematical Epidemiology*. *Lecture Notes in Math.* **1945**. Springer, Berlin. MR2452129 https://doi.org/10.1007/978-3-540-78911-6

[7] BRAUNER, J. M., MINDERMANN, S., SHARMA, M., JOHNSTON, D., SALVATIER, J., GAVENČIAK, T., STEPHENSON, A. B., LEECH, G., ALTMAN, G. et al. (2021). Inferring the effectiveness of government interventions against COVID-19. *Science* **371**. https://doi.org/10.1126/science.abd9338

George Nicholson is Senior Researcher, University of Oxford, Oxford, UK (e-mail: george.nicholson@stats.ox.ac.uk). Marta Blangiardo is Professor, MRC Centre for Environment and Health, Dept. of Epidemiology and Biostatistics, Imperial College London, London, UK. Mark Briers is Honorary Turing Fellow, The Alan Turing Institute, London, UK. Peter J. Diggle is Distinguished Professor, CHICAS, Lancaster Medical School, Lancaster University, Lancaster, UK. Tor Erlend Fjelde is PhD student, Department of Engineering, University of Cambridge, Cambridge, UK. Hong Ge is Senior Research Fellow, Department of Engineering, University of Cambridge, Cambridge, UK. Robert J. B. Goudie is Senior Investigator Statistician, MRC Biostatistics Unit, University of Cambridge, Cambridge, UK. Radka Jersakova is Senior Research Data Scientist, The Alan Turing Institute, London, UK. Ruairidh E. King is Data Wrangler, MRC Harwell Institute, Harwell, UK. Brieuc C. L. Lehmann is Assistant Professor, UCL, London, UK. Ann-Marie Mallon is Head of Bioinformatics, MRC Harwell Institute, Harwell, UK. Tullia Padellini is Research Associate, MRC Centre for Environment and Health, Dept. of Epidemiology and Biostatistics, Imperial College London, London, UK. Yee Whye Teh is Professor, University of Oxford, Oxford, UK. Chris Holmes is Professor, University of Oxford, Oxford, UK, Programme Director for Health and Medical Sciences, The Alan Turing Institute, London, UK, and Programme Leader, MRC Harwell Institute, Harwell, UK. Sylvia Richardson is Professor of Biostatistics and Emeritus Director, MRC Biostatistics Unit, University of Cambridge, Cambridge, UK (e-mail: sylvia.richardson@mrc-bsu.cam.ac.uk).

[8] CARMONA, C. U. and NICHOLLS, G. K. (2020). Semi-modular inference: Enhanced learning in multi-modular models by tempering the influence of components. Available at arXiv:2003.06804.

[9] CHEN, P. M., LEE, E. K., GIBSON, G. A., KATZ, R. H. and PATTERSON, D. A. (1994). RAID: High-performance, reliable secondary storage. *ACM Comput. Surv.* **26** 145–185. https://doi.org/10.1145/176979.176981

[10] COVID-19 INFECTION SURVEY—OFFICE FOR NATIONAL STATISTICS. Available at https://www.ons.gov.uk/surveys/informationforhouseholdsandindividuals/householdandindividualsurveys/covid19infectionsurvey.

[11] COVID-19 TASK FORCE. Available at https://rss.org.uk/policy-campaigns/policy-groups/covid-19-task-force/.

[12] DANIELS, M. J. and KASS, R. E. (1998). A note on first-stage approximation in two-stage hierarchical models. *Sankhya, Ser. B* **60** 19–30. MR1717073

[13] DAVISON, A. C. (2003). *Statistical Models. Cambridge Series in Statistical and Probabilistic Mathematics* **11**. Cambridge Univ. Press, Cambridge. MR1998913 https://doi.org/10.1017/CBO9780511815850

[14] DAWID, A. P. (1985). Probability, symmetry and frequency. *British J. Philos. Sci.* **36** 107–128. MR0915922 https://doi.org/10.1093/bjps/36.2.107

[15] DAWID, A. P. and LAURITZEN, S. L. (1993). Hyper-Markov laws in the statistical analysis of decomposable graphical models. *Ann. Statist.* **21** 1272–1317. MR1241267 https://doi.org/10.1214/aos/1176349260

[16] DE ANGELIS, D., PRESANIS, A. M., BIRRELL, P. J., TOMBA, G. S. and HOUSE, T. (2015). Four key challenges in infectious disease modelling using data from multiple sources. *Epidemics* **10** 83–87. https://doi.org/10.1016/j.epidem.2014.09.004

[17] DEPARTMENT OF HEALTH AND SOCIAL CARE (UK). COVID-19 testing data: Methodology note. Available at https://www.gov.uk/government/publications/coronavirus-covid-19-testing-data-methodology/covid-19-testing-data-methodology-note.

[18] DEPARTMENT OF HEALTH AND SOCIAL CARE GUIDANCE. REPRODUCTION NUMBER (R) AND GROWTH RATE: METHODOLOGY. Available at https://www.gov.uk/government/publications/reproduction-number-r-and-growth-rate-methodology.

[19] DOMINICI, F., SAMET, J. M. and ZEGER, S. L. (2000). Combining evidence on air pollution and daily mortality from the 20 largest US cities: A hierarchical modelling strategy. *J. Roy. Statist. Soc. Ser. A* **163** 263–302.

[20] FLAXMAN, S., MISHRA, S., GANDY, A., UNWIN, H. J. T., MELLAN, T. A., COUPLAND, H., WHITTAKER, C., ZHU, H., BERAH, T. et al. (2020). Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* **584** 257–261.

[21] GE, H., XU, K. and GHAHRAMANI, Z. (2018). Turing: A language for flexible probabilistic inference. In *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics* (A. Storkey and F. Perez-Cruz, eds.). *Proceedings of Machine Learning Research* **84** 1682–1690. PMLR, Playa Blanca, Lanzarote, Canary Islands.

[22] GIT. Available at https://git-scm.com/.

[23] GOUDIE, R. J. B., HOVORKA, R., MURPHY, H. R. and LUNN, D. (2015). Rapid model exploration for complex hierarchical data: Application to pharmacokinetics of insulin aspart. *Stat. Med.* **34** 3144–3158. MR3402581 https://doi.org/10.1002/sim.6536

[24] GOUDIE, R. J. B., PRESANIS, A. M., LUNN, D., DE ANGELIS, D. and WERNISCH, L. (2019). Joining and splitting models with Markov melding. *Bayesian Anal.* **14** 81–109. MR3910039 https://doi.org/10.1214/18-BA1104

[25] GREEN, P. J., HJORT, N. L. and RICHARDSON, S. (2003). *Highly Structured Stochastic Systems. Oxford Statistical Science Series*. Oxford University Press, Oxford, New York.

[26] HELLEWELL, J., RUSSELL, T. W., THE SAFER INVESTIGATORS AND FIELD STUDY TEAM, THE CRICK COVID-19 CONSORTIUM, CMMID COVID-19 WORKING GROUP, BEALE, R., KELLY, G., HOULIHAN, C., NASTOULI, E. et al. (2020). Estimating the effectiveness of routine asymptomatic PCR testing at different frequencies for the detection of SARS-CoV-2 infections. *MedRxiv* 2020.11.24.20229948. https://doi.org/10.1101/2020.11.24.20229948

[27] HINTON, G. E. (2002). Training products of experts by minimizing contrastive divergence. *Neural Comput.* **14** 1771–1800. https://doi.org/10.1162/089976602760128018

[28] HOOTEN, M. B., JOHNSON, D. S. and BROST, B. M. (2021). Making recursive Bayesian inference accessible. *Amer. Statist.* **75** 185–194. MR4256123 https://doi.org/10.1080/00031305.2019.1665584

[29] JACOB, P. E., MURRAY, L. M., HOLMES, C. C. and ROBERT, C. P. (2017). Better together? Statistical learning in models made of modules. Available at arXiv:1708.08719.

[30] JACOB, P. E., O'LEARY, J. and ATCHADÉ, Y. F. (2020). Unbiased Markov chain Monte Carlo methods with couplings. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **82** 543–600. https://doi.org/10.1111/rssb.12336

[31] JOHNSON, D. S., BROST, B. M. and HOOTEN, M. B. (2020). Greater than the sum of its parts: Computationally flexible Bayesian hierarchical modeling. Available at arXiv:2010.12568.

[32] THE JULIA PROGRAMMING LANGUAGE. Available at https://julialang.org/.

[33] LEE, D. and SARRAN, C. (2015). Controlling for unmeasured confounding and spatial misalignment in long-term air pollution and health studies. *Environmetrics* **26** 477–487. MR3415567 https://doi.org/10.1002/env.2348

[34] LINDSTEN, F., JOHANSEN, A. M., NAESSETH, C. A., KIRKPATRICK, B., SCHÖN, T. B., ASTON, J. A. D. and BOUCHARD-CÔTÉ, A. (2017). Divide-and-conquer with sequential Monte Carlo. *J. Comput. Graph. Statist.* **26** 445–458. MR3640200 https://doi.org/10.1080/10618600.2016.1237363

[35] LIU, F., BAYARRI, M. J. and BERGER, J. O. (2009). Modularization in Bayesian analysis, with emphasis on analysis of computer models. *Bayesian Anal.* **4** 119–150. MR2486241 https://doi.org/10.1214/09-BA404

[36] LIU, Y. and GOUDIE, R. J. B. (2021). Generalized geographically weighted regression model within a modularized Bayesian framework. Available at arXiv:2106.00996.

[37] LIU, Y. and GOUDIE, R. J. B. (2022). Stochastic approximation cut algorithm for inference in modularized Bayesian models. *Stat. Comput.* **32** Paper No. 7. MR4350200 https://doi.org/10.1007/s11222-021-10070-2

[38] LUNN, D., BARRETT, J., SWEETING, M. and THOMPSON, S. (2013). Fully Bayesian hierarchical modelling in two stages, with application to meta-analysis. *J. R. Stat. Soc. Ser. C. Appl. Stat.* **62** 551–572. MR3083911 https://doi.org/10.1111/rssc.12007

[39] LUNN, D., BEST, N., SPIEGELHALTER, D., GRAHAM, G. and NEUENSCHWANDER, B. (2009). Combining MCMC with 'sequential' PKPD modelling. *J. Pharmacokinet. Pharmacodyn.* **36** 19–38. https://doi.org/10.1007/s10928-008-9109-1

[40] MAISHMAN, T., SCHAAP, S., SILK, D. S., NEVITT, S. J., WOODS, D. C. and BOWMAN, V. E. (2021). Statistical methods used to combine the effective reproduction number, $R(t)$,

and other related measures of COVID-19 in the UK. Available at arXiv:2103.01742.

[41] MANDERSON, A. A. and GOUDIE, R. J. B. (2021). Combining chains of Bayesian models with Markov melding. Available at arXiv:2111.11566.

[42] MASSA, M. S. and LAURITZEN, S. L. (2010). Combining statistical models. In *Algebraic Methods in Statistics and Probability II* (M. A. G. Viana and H. P. Wynn, eds.). *Contemp. Math.* **516** 239–259. Amer. Math. Soc., Providence, RI. MR2730753 https://doi.org/10.1090/conm/516/10179

[43] MATHUR, R., RENTSCH, C., MORTON, C., HULME, W., SCHULTZE, A., MACKENNA, B., EGGO, R., BHASKARAN, K., WONG, A. et al. (2021). Ethnic differences in SARS-CoV-2 infection and COVID-19-related hospitalisation, intensive care unit admission, and death in 17 million adults in England: An observational cohort study using the OpenSAFELY platform. *Lancet* **397** 1711–1724.

[44] MAUCORT-BOULCH, D., FRANCESCHI, S., PLUMMER, M. and IARC HPV PREVALENCE SURVEYS STUDY GROUP (2008). International correlation between human papillomavirus prevalence and cervical cancer incidence. *Cancer Epidemiol. Biomark. Prev.* **17** 717–720. https://doi.org/10.1158/1055-9965.EPI-07-2691

[45] MAUFF, K., STEYERBERG, E., KARDYS, I., BOERSMA, E. and RIZOPOULOS, D. (2020). Joint models with multiple longitudinal outcomes and a time-to-event outcome: A corrected two-stage approach. *Stat. Comput.* **30** 999–1014. MR4108688 https://doi.org/10.1007/s11222-020-09927-9

[46] MINISTRY OF HOUSING, COMMUNITIES & LOCAL GOVERNMENT. English indices of deprivation 2019. Available at https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019.

[47] MORALES, D. R. and ALI, S. N. (2021). COVID-19 and disparities affecting ethnic minorities. *Lancet* **397** 1684–1685. https://doi.org/10.1016/S0140-6736(21)00949-1

[48] MUGGLIN, A. S., CARLIN, B. P. and GELFAND, A. E. (2000). Fully model-based approaches for spatially misaligned data. *J. Amer. Statist. Assoc.* **95** 877–887. https://doi.org/10.1080/01621459.2000.10474279

[49] NICHOLLS, G. K., LEE, J. E., WU, C.-H. and CARMONA, C. U. (2022). Valid belief updates for prequentially additive loss functions arising in semi-modular inference. Available at arXiv:2201.09706.

[50] NICHOLSON, G., BLANGIARDO, M., BRIERS, M., DIGGLE, P. J., FJELDE, T. E., GE, H., GOUDIE, R. J. B., JERSAKOVA, R., KING, R. E. et al. (2022). Supplement to "Interoperability of statistical models in pandemic preparedness: Principles and reality." https://doi.org/10.1214/22-STS854SUPP

[51] NICHOLSON, G., LEHMANN, B., PADELLINI, T., POUWELS, K. B., JERSAKOVA, R., LOMAX, J., KING, R. E., MALLON, A.-M., DIGGLE, P. J. et al. (2022). Improving local prevalence estimates of SARS-CoV-2 infections using a causal debiasing framework. *Nat. Microbiol.* **7** 97–107. https://doi.org/10.1038/s41564-021-01029-0

[52] PADELLINI, T., JERSAKOVA, R., DIGGLE, P. J., HOLMES, C., KING, R. E., LEHMANN, B. C. L., MALLON, A.-M., NICHOLSON, G., RICHARDSON, S. et al. (2022). Time varying association between deprivation, ethnicity and SARS-CoV-2 infections in England: A population-based ecological study. *The Lancet Regional Health—Europe* **15** 100322. https://doi.org/10.1016/j.lanepe.2022.100322

[53] PIRANI, M., MASON, A. J., HANSELL, A. L., RICHARDSON, S. and BLANGIARDO, M. (2020). A flexible hierarchical framework for improving inference in area-referenced environmental health studies. *Biom. J.* **62** 1650–1669. MR4184022 https://doi.org/10.1002/bimj.201900241

[54] PLUMMER, M. (2015). Cuts in Bayesian graphical models. *Stat. Comput.* **25** 37–43. MR3304902 https://doi.org/10.1007/s11222-014-9503-z

[55] POMPE, E. and JACOB, P. E. (2021). Asymptotics of cut distributions and robust modular inference using Posterior Bootstrap. Available at arXiv:2110.11149.

[56] POOLE, D. and RAFTERY, A. E. (2000). Inference for deterministic simulation models: The Bayesian melding approach. *J. Amer. Statist. Assoc.* **95** 1244–1255. MR1804247 https://doi.org/10.2307/2669764

[57] POUWELS, K. B., HOUSE, T., PRITCHARD, E., ROBOTHAM, J. V., BIRRELL, P. J., GELMAN, A., VIHTA, K.-D., BOWERS, N., BOREHAM, I. et al. (2021). Community prevalence of SARS-CoV-2 in England from April to November, 2020: Results from the ONS Coronavirus Infection Survey. *The Lancet Public Health* **6** e30–e38.

[58] PRESANIS, A. M., OHLSSEN, D., SPIEGELHALTER, D. J. and DE ANGELIS, D. (2013). Conflict diagnostics in directed acyclic graphs, with applications in Bayesian evidence synthesis. *Statist. Sci.* **28** 376–397. MR3135538 https://doi.org/10.1214/13-STS426

[59] RIEBLER, A., SØRBYE, S. H., SIMPSON, D. and RUE, H. (2016). An intuitive Bayesian spatial model for disease mapping that accounts for scaling. *Stat. Methods Med. Res.* **25** 1145–1165. MR3541089 https://doi.org/10.1177/0962280216660421

[60] RILEY, S., AINSLIE, K. E., EALES, O., JEFFREY, B., WALTERS, C. E., ATCHISON, C. J., DIGGLE, P. J., ASHBY, D., DONNELLY, C. A. et al. (2020). Community prevalence of SARS-CoV-2 virus in England during May 2020: REACT study. *MedRxiv*.

[61] ROSE, T. C., MASON, K., PENNINGTON, A., MCHALE, P., TAYLOR-ROBINSON, D. C. and BARR, B. (2020). Inequalities in COVID19 mortality related to ethnicity and socioeconomic deprivation. *MedRxiv*.

[62] RUE, H., MARTINO, S. and CHOPIN, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **71** 319–392. MR2649602 https://doi.org/10.1111/j.1467-9868.2008.00700.x

[63] SCOTT, S. L. (2002). Bayesian methods for hidden Markov models: Recursive computing in the 21st century. *J. Amer. Statist. Assoc.* **97** 337–351. MR1963393 https://doi.org/10.1198/016214502753479464

[64] SIMPSON, D., RUE, H., RIEBLER, A., MARTINS, T. G. and SØRBYE, S. H. (2017). Penalising model component complexity: A principled, practical approach to constructing priors. *Statist. Sci.* **32** 1–28. MR3634300 https://doi.org/10.1214/16-STS576

[65] SØRBYE, S. H. and RUE, H. (2014). Scaling intrinsic Gaussian Markov random field priors in spatial modelling. *Spat. Stat.* **8** 39–51. MR3326820 https://doi.org/10.1016/j.spasta.2013.06.004

[66] TEH, Y. W., BHOOPCHAND, A., DIGGLE, P., ELESEDY, B., HE, B., HUTCHINSON, M., PAQUET, U., READ, J., TOMASEV, N. et al. (2021). Efficient Bayesian inference of instantaneous re-production numbers at fine spatial scales, with an application to mapping and nowcasting the Covid-19 epidemic in british local authorities. Technical report. To be published in Journal of the Royal Statistical Society, Series A (Statistics in Society). Available at https://localcovid.info/assets/docs/localcovid-writeup.pdf.

[67] UK DATA SERVICE CENSUS. Available at https://ukdataservice.ac.uk/learning-hub/census/.

[68] WELTON, N. J., SUTTON, A. J., COOPER, N. J., ABRAMS, K. R. and ADES, A. E. (2012). *Evidence Synthesis for Decision Making in Healthcare*. Wiley, Chichester.

[69] YU, B. and KUMBIER, K. (2020). Veridical data science. *Proc. Natl. Acad. Sci. USA* **117** 3920–3929. MR4075122 https://doi.org/10.1073/pnas.1901326117

[70] YU, X., NOTT, D. J. and SMITH, M. S. (2021). Variational inference for cutting feedback in misspecified models. Available at arXiv:2108.11066.

[71] ZHANG, L., BEAL, S. L. and SHEINER, L. B. (2003). Simultaneous vs. sequential analysis for population PK/PD data I: Best-case performance. *J. Pharmacokinet. Pharmacodyn.* **30** 387–404. https://doi.org/10.1023/b:jopa.0000012998.04442.1f

Statistical Science
2022, Vol. 37, No. 2, 207–228
https://doi.org/10.1214/22-STS856

Real-Time Estimation of COVID-19 Infections: Deconvolution and Sensor Fusion

Maria Jahja, Andrew Chin and Ryan J. Tibshirani

Abstract. We propose, implement, and evaluate a method to estimate the daily number of new symptomatic COVID-19 infections, at the level of individual U.S. counties, by deconvolving daily reported COVID-19 case counts using an estimated symptom-onset-to-case-report delay distribution. Importantly, we focus on estimating infections in real-time (rather than retrospectively), which poses numerous challenges. To address these, we develop new methodology for both the distribution estimation and deconvolution steps, and we employ a sensor fusion layer (which fuses together predictions from models that are trained to track infections based on auxiliary surveillance streams) in order to improve accuracy and stability.

Key words and phrases: COVID-19, nowcasting, deconvolution, sensor fusion.

REFERENCES

ABBOTT, S., HELLEWELL, J., THOMPSON, R. N., SHERRATT, K., GIBBS, H. P., BOSSE, N. I., MUNDAY, J. D., MEAKIN, S., DOUGHTY, E. L. et al. (2020). Estimating the time-varying reproduction number of Sars-CoV-2 using national and subnational case counts. *Wellcome Open Research* **5**.

ACKLEY, A. F., PILEWSKI, S., PETROVIC, V. S., WORDEN, L., MURRAY, E. and PORCO, T. C. (2020). Assessing the utility of a smart thermometer and mobile application as a surveillance tool for influenza and influenza-like illness. *Health Inform. J.* **26** 2148–2158.

BAVADEKAR, S., DAI, A., DAVIS, J., DESFONTAINES, D., ECKSTEIN, I., EVERETT, K., FABRIKANT, A., FLORES, G., GABRILOVICH, E. et al. (2020). Google COVID-19 search trends symptoms dataset: Anonymization process description. arXiv:2009.01265.

BETTENCOURT, L. M. A. and RIBEIRO, R. M. (2008). Real time Bayesian estimation of the epidemic potential of emerging infectious diseases. *PLoS ONE* **3** e2185. https://doi.org/10.1371/journal.pone.0002185

BROOKS, L. C. (2020). Pancasting: Forecasting epidemics from provisional data Ph.D. thesis Carnegie Mellon Univ.

BROWNSTEIN, J. S., FREIFELD, C. C. and MADOFF, L. C. (2009). Digital disease detection — harnessing the web for public health surveillance. *N. Engl. J. Med.* **360** 2153–2157.

CARLSON, S. J., DALTON, C. B., BUTLER, M. T., FEJSA, J., ELVIDGE, E. and DURRHEIM, D. N. (2013). Flutracking weekly online community survey of influenza-like illness annual report 2011 and 2012. *Communicable Diseases Intelligence Quarterly Report* **37** E398–406.

CHARU, V. (2017). Human mobility and the spatial transmission of influenza in the United States. *PLoS Comput. Biol.* **13** 1–23.

CHITWOOD, M. H., RUSSI, M., GUNASEKERA, K., HAVUMAKI, J., PITZER, V. E., SALOMON, J. A., SWARTWOOD, N., WARREN, J. L., WEINBERGER, D. M. et al. (2021). Reconstructing the course of the COVID-19 epidemic over 2020 for US states and counties: Results of a Bayesian evidence synthesis model. medRxiv. https://doi.org/10.1101/2020.06.17.20133983

CORI, A., FERGUSON, N. M., FRASER, C. and CAUCHEMEZ, S. (2013). A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am. J. Epidemiol.* **178** 1505–1512.

DEBEYE, H. W. J. and VAN RIEL, P. (1990). L_p-norm deconvolution. *Geophys. Prospect.* **38** 381–403.

DONG, E., DU, H. and GARDNER, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20** 533–544.

FARROW, D. C. (2016). Modeling the past, present, and future of influenza Ph.D. thesis Carnegie Mellon Univ.

CENTERS FOR DISEASE CONTROL (2020a). COVID-19 Case Surveillance Public Use Data. https://data.cdc.gov/Case-Surveillance/COVID-19-Case-Surveillance-Public-Use-Data/vbim-akqf. Data accessed on November 3, 2021.

CENTERS FOR DISEASE CONTROL (2020b). COVID-19 Case Surveillance Restricted Access Detailed Data. https://data.cdc.gov/Case-Surveillance/COVID-19-Case-Surveillance-Restricted-Access-Detai/mbd7-r32t. Data accessed on November 3, 2021.

GINSBERG, J., MOHEBBI, M. H., PATEL, R. S., BRAMMER, L., SMOLINSKI, M. S. and BRILLIANT, L. (2009). Detecting in-

Maria Jahja is Ph.D. Candidate, Department of Statistics & Data Science, Machine Learning Department, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA (e-mail: maria@stat.cmu.edu). Andrew Chin is Statistical Developer, Machine Learning Department, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA (e-mail: achin23@jhu.edu). Ryan J. Tibshirani is Professor, Department of Statistics & Data Science, Machine Learning Department, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA (e-mail: ryantibs@cmu.edu).

fluenza epidemics using search engine query data. *Nature* **457** 1012–1014.

GOLDSTEIN, E., DUSHOFF, J., MA, J., PLOTKIN, J. B., EARN, D. J. and LIPSITCH, M. (2009a). Reconstructing influenza incidence by deconvolution of daily mortality time series. *Proc. Natl. Acad. Sci. USA* **106** 21825–21829.

GOSTIC, K. M. (2020). Practical considerations for measuring the effective reproductive number, R_t. *PLoS Comput. Biol.* **16** 1–21.

HAWRYLUK, I., HOELTGEBAUM, H., MISHRA, S., MISCOURIDOU, X., SCHNEKENBERG, R. P., WHITTAKER, C., VOLLMER, M., FLAXMAN, S., BHATT, S. et al. (2021). Gaussian process nowcasting: application to COVID-19 mortality reporting. In *Conference on Uncertainty in Artificial Intelligence*.

JAHJA, M., CHIN, A. and TIBSHIRANI, R. J (2022). Supplement to "Real-Time Estimation of COVID-19 Infections: Deconvolution and Sensor Fusion." https://doi.org/10.1214/22-STS856SUPP

JAHJA, M., FARROW, D., ROSENFELD, R. and TIBSHIRANI, R. J. (2019). Kalman filter, sensor fusion, and constrained regression: Equivalences and insights. In *Advances in Neural Information Processing Systems*.

JOHNSON, N. A. (2013). A dynamic programming algorithm for the fused lasso and L_0-segmentation. *J. Comput. Graph. Statist.* **22** 246–260. MR3173713 https://doi.org/10.1080/10618600.2012.681238

KAPLAN, E. L. and MEIER, P. (1958). Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* **53** 457–481. MR0093867

KASS-HOUT, T. A. and ALHINNAWI, H. (2013). Social media in public health. *Br. Med. Bull.* **108** 5–24. https://doi.org/10.1093/bmb/ldt028

KASS-HOUT, T. A. and ZHANG, X. (2011). *Biosurveillance: Methods and Case Studies*. CRC Press, Boca Raton.

LEUBA, S. I. (2020). Tracking and predicting U.S. influenza activity with a real-time surveillance network. *PLoS Comput. Biol.* **16** 1–14.

MCDONALD, D. J., BIEN, J., GREEN, A., HU, A. J., DEFRIES, N., HYUN, S., OLIVEIRA, N. L., SHARPNACK, J., TANG, J. et al. (2021). Can auxiliary indicators improve COVID-19 forecasting and hotspot prediction? e2111453118. To appear, PNAS.

MCGOUGH, S. F., JOHANSSON, M. A., LIPSITCH, M. and MENZIES, N. A. (2020). Nowcasting by Bayesian smoothing: A flexible, generalizable model for real-time epidemic tracking. *PLoS Comput. Biol.* **16** e1007735. https://doi.org/10.1371/journal.pcbi.1007735

MCIVER, D. J. and BROWNSTEIN, J. S. (2014). Wikipedia usage estimates prevalence of influenza-like illness in the United States in near real-time. *PLoS Comput. Biol.* **10** e1003581. https://doi.org/10.1371/journal.pcbi.1003581

OPPENHEIM, A. V. and VERGHESE, G. C. (2017). *Signals, Systems and Inference*. Pearson, Upper Saddle River.

PAUL, M. J. and DREDZE, M. (2017). Social monitoring for public health. *Synth. Lect. Inf. Concept. Retr. Services* **9** 1–183.

RADIN, J. M., WINEINGER, N. E., TOPOL, E. J. and STEINHUBL, S. R. (2020). Harnessing wearable device data to improve state-level real-time surveillance of influenza-like illness in the USA: A population-based study. *The Lancet Digital Health* **2** e85–e93.

RAMDAS, A. and TIBSHIRANI, R. J. (2016). Fast and flexible ADMM algorithms for trend filtering. *J. Comput. Graph. Statist.* **25** 839–858. MR3533641 https://doi.org/10.1080/10618600.2015.1054033

REICH LAB (2020). The COVID-19 Forecast Hub. https://covid19forecasthub.org.

REINHART, A., BROOKS, L., JAHJA, M., RUMACK, A., TANG, J., AGRAWAL, S., SAEED, W. A., ARNOLD, T., BASU, A. et al. (2021). An open repository of real-time COVID-19 indicators. *Proc. Natl. Acad. Sci. USA* **51** e2111452118.

ROSENFELD, R. and TIBSHIRANI, R. J. (2021). Epidemic tracking and forecasting: Lessons learned from a tumultuous year. *Proc. Natl. Acad. Sci. USA* **51** e2111456118.

RUDIN, L. I. and OSHER, S. (1994). Total variation based image restoration with free local constraints. In *International Conference on Image Processing* **1** 31–35.

SALATHÉ, M., BENGTSSON, L., BODNAR, T. J., BREWER, D. D., BROWNSTEIN, J. S., BUCKEE, C., CAMPBELL, E. M., CATTUTO, C., KHANDELWAL, S. et al. (2012). Digital epidemiology. *PLoS Comput. Biol.* **8** 1–3.

SALOMON, J. A., REINHART, A., BILINSKI, A., CHUA, E. J., LA MOTTE-KERR, W., RÖNN, M. M., REITSMA, M., MORRIS, K. A., LAROCCA, S. et al. (2021). The COVID-19 trends and impact survey: Continuous real-time measurement of COVID-19 symptoms, risks, protective behaviors, testing and vaccination. *Proc. Natl. Acad. Sci. USA* **51** e2111454118.

SANTILLANA, M., NGUYEN, A. T., DREDZE, M., PAUL, M. J., NSOESIE, E. O. and BROWNSTEIN, J. S. (2015). Combining search, social media, and traditional data sources to improve influenza surveillance. *PLoS Comput. Biol.* **11** e1004513. https://doi.org/10.1371/journal.pcbi.1004513

SANTILLANA, M., NGUYEN, A. T., LOUIE, T., ZINK, A., GRAY, J., SUNG, I. and BROWNSTEIN, J. S. (2016). Cloud-based electronic health records for real-time, region-specific influenza surveillance. *Sci. Rep.* **6** 1–8.

SMOLINSKI, M. S., CRAWLEY, A. W., BALTRUSAITIS, K., CHUNARA, R., OLSEN, J. M., WÓJCIK, O., SANTILLANA, M., NGUYEN, A. and BROWNSTEIN, J. S. (2015). flu near you: Crowdsourced symptom reporting spanning 2 influenza seasons. *Am. J. Publ. Health* **105** 2124–2130.

SYSTROM, K., VLADEK, T. and KRIEGER, M. (2020). Rt.live. https://github.com/rtcovidlive/covid-model.

TAYLOR, H. L., BANKS, S. C. and MCCOY, J. F. (1979). Deconvolution with the ℓ_1 norm. *Geophysics* **44** 39–52.

THOMPSON, R. N., STOCKWIN, J. E., VAN GAALEN, R. D., POLONSKY, J. A., KAMVAR, Z. N., DEMARSH, P. A., DAHLQWIST, E., LI, S., MIGUEL, E. et al. (2019). Improved inference of time-varying reproduction numbers during infectious disease outbreaks. *Epidemics* **29** 100356.

TIBSHIRANI, R. J. (2014). Adaptive piecewise polynomial estimation via trend filtering. *Ann. Statist.* **42** 285–323. MR3189487 https://doi.org/10.1214/13-AOS1189

TIBSHIRANI, R. J. (2020). Divided differences, falling factorials, and discrete splines: Another look at trend filtering and related problems. arXiv:2003.03886.

VIBOUD, C. (2014). Demonstrating the use of high-volume electronic medical claims data to monitor local and regional influenza activity in the US. *PLoS ONE* **9** 1–12.

WALLINGA, J. and LIPSITCH, M. (2007). How generation intervals shape the relationship between growth rates and reproductive numbers. *Proc. R. Soc. Lond., B Biol. Sci.* **274** 599–604.

WIENER, N. (1964). *Extrapolation, Interpolation, and Smoothing of Stationary Time Series*. MIT Press, Cambridge.

YANG, S., SANTILLANA, M. and KOU, S. C. (2015). Accurate estimation of influenza epidemics using Google search data via ARGO. *Proc. Natl. Acad. Sci. USA* **112** 14473–14478.

YANG, C.-Y., CHEN, R.-J., CHOU, W.-L., LEE, Y.-J. and LO, Y.-S. (2019). An integrated influenza surveillance framework based on national influenza-like illness incidence and multiple hospital electronic medical records for early prediction of influenza epidemics: Design and evaluation. *J. Med. Internet Res.* **21** e12341.

Statistical Science
2022, Vol. 37, No. 2, 229–250
https://doi.org/10.1214/22-STS857

Statistical Modeling for Practical Pooled Testing During the COVID-19 Pandemic

Saskia Comess, Hannah Wang, Susan Holmes and Claire Donnat

Abstract. Pooled testing offers an efficient solution to the unprecedented testing demands of the COVID-19 pandemic, despite their potentially lower sensitivity and increased costs to implementation in certain settings. Assessments of this trade-off typically assume the underlying infection statuses of pooled specimens to be independent and identically distributed. Yet, in the context of COVID-19, these assumptions are often violated: testing done on networks (housemates, spouses, co-workers) captures individuals with correlated infection statuses and risk, while infection risk varies substantially across time, place and individuals. Neglecting dependencies and heterogeneity may bias established optimality grids and induce a sub-optimal implementation of the procedure. As a lesson learned from this pandemic, this paper highlights the necessity of integrating field sampling information with statistical modeling to efficiently optimize pooled testing. Using real data, we show that (a) greater gains can be achieved at low logistical cost by exploiting natural correlations (nonindependence) between samples—allowing improvements in sensitivity and efficiency of up to 30% and 90%, respectively; and (b) these gains are robust despite substantial heterogeneity across pools (nonidentical). Our modeling results complement and extend the observations of Barak et al. (*Sci. Transl. Med.* **13** (2021) 1–8) who report an empirical sensitivity well beyond expectations. Finally, we provide an interactive tool for selecting an optimal pool size using contextual information.[1]

Key words and phrases: COVID-19, pooled testing, correlations, heterogeneity.

REFERENCES

ABDALHAMID, B., BILDER, C. R., MCCUTCHEN, E. L., HINRICHS, S. H., KOEPSELL, S. A. and IWEN, P. C. (2020). Assessment of specimen pooling to conserve SARS CoV-2 testing resources. *Am. J. Clin. Pathol.* **153** 715–718. https://doi.org/10.1093/ajcp/aqaa064

ADAM, D. C., WU, P., WONG, J. Y., LAU, E. H. Y., TSANG, T. K., CAUCHEMEZ, S., LEUNG, G. M. and COWLING, B. J. (2020). Clustering and superspreading potential of SARS-CoV-2 infections in Hong Kong. *Nat. Med.* **26** 1714–1719.

ASM (2021). Supply Shortages Impacting COVID-19 and Non-COVID Testing. [Online; posted 19-January-2021].

BARAK, N., AMI, R. B., SIDO, T., PERRI, A., SHTOYER, A., RIVKIN, M., LICHT, T., PERETZ, A., MAGENHEIM, J. et al. (2021). Lessons from applied large-scale pooling of 133,816 SARS-CoV-2 RT-PCR tests. *Sci. Transl. Med.* **13** 1–8.

BERKE, E. M., NEWMAN, L. M., JEMSBY, S., HYDE, B., BHALLA, N., SHEILS, N. E., OOMMAN, N., REPPAS, J., VERMA, P. et al. (2021). Pooling in a pod: A strategy for COVID-19 testing to facilitate a safe return to school. *Public Health Reports* **136** 663–670.

BILDER, C. R. and TEBBS, J. M. (2012). Pooled-testing procedures for screening high volume clinical specimens in heterogeneous populations. *Stat. Med.* **31** 3261–3268. MR3041806 https://doi.org/10.1002/sim.5334

BILDER, C. R., TEBBS, J. M. and CHEN, P. (2010). Informative retesting. *J. Amer. Statist. Assoc.* **105** 942–955. MR2752591 https://doi.org/10.1198/jasa.2010.ap09231

BILDER, C. R., TEBBS, J. M. and MCMAHAN, C. S. (2019). Informative group testing for multiplex assays. *Biometrics* **75** 278–288. MR3953728 https://doi.org/10.1111/biom.12988

CAPS (2020). Pooled Testing: Guidance from the CAP's Microbiology Committee. Available at www.cap.org/covid-19/pooled-

Saskia Comess is a PhD student, Emmett Interdisciplinary Program in Environment and Resources, Stanford University, Stanford, California (e-mail: saskiaco@stanford.edu). Hannah Wang is a resident physician, Department of Anatomic and Clinical Pathology, Stanford University School of Medicine, Stanford, California (e-mail: wangh1@stanford.edu). Susan Holmes is a Professor, Department of Statistics, Stanford University, Stanford, California (e-mail: susan@stat.stanford.edu). Claire Donnat is an Assistant Professor, Department of Statistics, The University of Chicago, Chicago, Illinois (e-mail: cdonnat@uchicago.edu).

testing-guidance-from-cap-microbiology-committee. Accessed: 2021-01-06.

CAVE, E. (2020). COVID-19 super-spreaders: Definitional quandaries and implications. *Asian Bioethics Review* 1.

CENSUS. GOV (2021). State Population Totals and Components of Change: 2010–2019. Available at www.census.gov/data/tables/time-series/demo/popest/2010s-state-total.html.

CHEN, P., TEBBS, J. M. and BILDER, C. R. (2009). Group testing regression models with fixed and random effects. *Biometrics* **65** 1270–1278. MR2756515 https://doi.org/10.1111/j.1541-0420.2008.01183.x

CHITWOOD, M. H., RUSSI, M., GUNASEKERA, K., HAVUMAKI, J., KLAASSEN, F., PITZER, V. E., SALOMON, J. A., SWARTWOOD, N. A., WARREN, J. L. et al. (2021). Reconstructing the course of the COVID-19 epidemic over 2020 for US states and counties: Results of a Bayesian evidence synthesis model. *MedRxiv*. https://doi.org/10.1101/2020.06.17.20133983

CIRILLO, P. and TALEB, N. N. (2020). Tail risk of contagious diseases. *Nat. Phys.* **16** 606–613.

COLLINS, K. (2020). Is Your State Doing Enough Coronavirus Testing? Available at www.nytimes.com/interactive/2020/us/coronavirus-testing.html. [Online; posted 1-November-2020].

COMESS, S., WANG, H., HOLMES, S. and DONNAT, C. (2022). Supplement to "Statistical Modeling for Practical Pooled Testing During the COVID-19 Pandemic." https://doi.org/10.1214/22-STS857SUPP

CRAMER, E. Y., LOPEZ, V. K., NIEMI, J., GEORGE, G. E., CEGAN, J. C., DETTWILLER, I. D., ENGLAND, W. P., FARTHING, M. W., HUNTER, R. H. et al. (2021). Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the US. *MedRxiv* 2021.02.03.21250974.

CURMEI, M., ILYAS, A., EVANS, O. and STEINHARDT, J. (2020). Estimating household transmission of SARS-CoV-2. *medRxiv* 1–24. https://doi.org/10.1101/2020.05.23.20111559

DECKERT, A., BÄRNIGHAUSEN, T. and KYEI, N. N. (2020). Simulation of pooled-sample analysis strategies for COVID-19 mass testing. *Bull. World Health Organ.* **98** 590–598. https://doi.org/10.2471/BLT.20.257188

DHILLON, R. S., SRIKRISHNA, D., GARRY, R. F. and CHOWELL, G. (2015). Ebola control: Rapid diagnostic testing. *Lancet Infect. Dis.* **15** 147–148.

DONNAT, C. and HOLMES, S. (2021). Modeling the heterogeneity in COVID-19's reproductive number and its impact on predictive scenarios. *J. Appl. Stat.* 1–29.

DONNAT, C., MIOLANE, N., BUNBURY, F. and KREINDLER, J. (2020). A Bayesian Hierarchical Network for Combining Heterogeneous Data Sources in Medical Diagnoses. Available at arXiv:2007.13847.

DONNAT, C., BUNBURY, F., KREINDLER, J., FILIPPIDIS, F. T., EL-OSTA, A., ESKO, T. and HARRIS, M. (2021). A Predictive Modelling Framework for COVID-19 Transmission to Inform the Management of Mass Events. *MedRxiv* 2021.05.13.21256857.

DORFMAN, R. (1943). The detection of defective members of large populations. *Ann. Math. Stat.* **14** 436–440.

EUNJUNG CHA, A. (2021). The future of coronavirus testing is in Greenville, N.C. Available at www.washingtonpost.com/health/2021/04/28/new-coronavirus-testing-strategy-home-kits/. [Online; posted 28-April-2021].

FDA (2020). Coronavirus (COVID-19) Update: FDA Issues First Emergency Authorization for Sample Pooling in Diagnostic Testing. Available at www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-issues-first-emergency-authorization-sample-pooling-diagnostic. [Online; posted 18-July-2020].

FRAYER, L. (2021). Oxygen Rationing, Test Shortages: India Caught Unprepared In COVID-19 Crisis. Available at www.npr.org/2021/04/24/990544555/oxygen-rationing-test-shortages-india-caught-unprepared-in-covid-19-crisis. [Online; posted 24-April-2021].

GANDHI, M., YOKOE, D. S. and HAVLIR, D. V. (2020). Asymptomatic transmission, the Achilles' heel of current strategies to control COVID-19. *N. Engl. J. Med.* **382** 36–39.

GASTWIRTH, J. L. (2000). The efficiency of pooling in the detection of rare mutations. *Am. J. Hum. Genet.* **67** 1036–1039.

GAYDOS, C. A. (2005). Nucleic acid amplification tests for gonorrhea and chlamydia: Practice and applications. *Infectious Disease Clinics* **19** 367–386.

GIRI, B., PANDEY, S., SHRESTHA, R., POKHAREL, K., LIGLER, F. S. and NEUPANE, B. B. (2021). Review of analytical performance of COVID-19 detection methods. *Anal. Bioanal. Chem.* **413** 35–48. https://doi.org/10.1007/s00216-020-02889-x

GÓMEZ-CARBALLA, A., BELLO, X., PARDO-SECO, J., MARTINÓN-TORRES, F. and SALAS, A. (2020). Mapping genome variation of SARS-CoV-2 worldwide highlights the impact of COVID-19 super-spreaders. *Genome Res.* **30** 1434–1448. https://doi.org/10.1101/gr.266221.120

HE, J., GUO, Y., MAO, R. and ZHANG, J. (2020). Proportion of asymptomatic coronavirus disease 2019: A systematic review and meta-analysis. *Journal of Medical Virology*.

HEILWEIL, R. (2021). How omicron broke Covid-19 testing. Available at www.vox.com/recode/2021/12/21/22848286/omicron-rapid-test-covid-19-antigen. [Online; posted 21 December 2021].

JOACHIM, A., DEWALD, F., SUÁREZ, I., ZEMLIN, M., LANG, I., STUTZ, R., MARTHALER, A., BOSSE, H. M., LÜBKE, N. et al. (2021). Pooled RT-qPCR testing for SARS-CoV-2 surveillance in schools—a cluster randomised trial. *EClinicalMedicine* **39** 101082.

JONES, C. (2021). 'Pool testing' to combat Covid on campus grows popular in California schools. Available at edsource.org/2021/pool-testing-to-combat-covid-on-campus-grows-popular-in-california-schools/661144. [Online; posted 15 September 2021].

JOSEPH, L. and BELISLE, P. (2017). Version 1.3 beta.parms.from.quantiles [R] Computing Beta distribution parameters. Available at www.medicine.mcgill.ca/epidemiology/Joseph/PBelisle/BetaParmsFromQuantiles.html. [Online; updated February 2017].

KIM, H.-Y., HUDGENS, M. G., DREYFUSS, J. M., WESTREICH, D. J. and PILCHER, C. D. (2007). Comparison of group testing algorithms for case identification in the presence of test error. *Biometrics* **63** 1152–1163, 1313. MR2414593 https://doi.org/10.1111/j.1541-0420.2007.00817.x

KOH, W. C., NAING, L., CHAW, L., ROSLEDZANA, M. A., ALIKHAN, M. F., JAMALUDIN, S. A., AMIN, F., OMAR, A., SHAZLI, A. et al. (2020). What do we know about SARS-CoV-2 transmission? A systematic review and meta-analysis of the secondary attack rate and associated risk factors. *PLoS ONE* **15** 1–23.

LARREMORE, D. B., WILDER, B., LESTER, E., SHEHATA, S., BURKE, J. M., HAY, J. A., TAMBE, M., MINA, M. J. and PARKER, R. (2020). Test sensitivity is secondary to frequency and turnaround time for COVID-19 screening. *Sci. Adv.*

LIN, Y.-J., YU, C.-H., LIU, T.-H., CHANG, C.-S. and CHEN, W.-T. (2020). Positively Correlated Samples Save Pooled Testing Costs. Available at arXiv:2011.09794.

MA, Q., LIU, J., LIU, Q., KANG, L., LIU, R., JING, W., WU, Y. and LIU, M. (2021). Global percentage of asymptomatic SARS-CoV-2 infections among the tested population and individuals with confirmed COVID-19 diagnosis: A systematic review and meta-analysis. *JAMA Network Open* **4**.

MADEWELL, Z.J., YANG, Y., LONGINI, I.M., HALLORAN, M.E. and DEAN, N.E. (2020). Household transmission of SARS-CoV-2: A systematic review and meta-analysis. *JAMA Network Open* **3** e2031756. https://doi.org/10.1001/jamanetworkopen.2020.31756

MASS. GOV (2021). Baker-Polito Administration's First in the Nation COVID-19 Pooled Testing Initiative Finds 0.7% Positivity Rate in Schools Throughout Commonwealth. Available at www.mass.gov/news/baker-polito-administrations-first-in-the-nation-covid-19-pooled-testing-initiative-finds-07-positivity-rate-in-schools-throughout-commonwealth. [Online; posted 29 March 2021].

MASSACHUSSETS, DEPARTMENT OF EDUCATION (2022). COVID-19 Testing Program. Available at www.doe.mass.edu/covid19/testing/default.html. [Online; updated 5 January 2022].

MCMAHAN, C. S., TEBBS, J. M. and BILDER, C. R. (2012a). Informative Dorfman screening. *Biometrics* **68** 287–296. MR2909885 https://doi.org/10.1111/j.1541-0420.2011.01644.x

MCMAHAN, C. S., TEBBS, J. M. and BILDER, C. R. (2012b). Two-dimensional informative array testing. *Biometrics* **68** 793–804. MR3055184 https://doi.org/10.1111/j.1541-0420.2011.01726.x

MINA, M. J., PARKER, R. and LARREMORE, D. B. (2020). Rethinking Covid-19 test sensitivity—a strategy for containment. *N. Engl. J. Med.* **383** e120.

MWAI, P. (2021). Coronavirus in Africa: Concern growing over third wave of Covid-19 infections. Available at bbc.com/news/world-africa-53181555. [Online; posted 7-June-2021].

NOUVELLET, P., GARSKE, T., MILLS, H. L., NEDJATI-GILANI, G., HINSLEY, W., BLAKE, I. M., VAN KERKHOVE, M. D., CORI, A., DORIGATTI, I. et al. (2015). The role of rapid diagnostics in managing Ebola epidemics. *Nature* **528** S109–S116.

ORAN, D. P. and TOPOL, E. J. (2020). Prevalence of asymptomatic SARS-CoV-2 infection : A narrative review. *Ann. Intern. Med.* **173** 362–367. https://doi.org/10.7326/M20-3012

PIETSCH, B. (2021). More coronavirus tests will be available next month, Fauci says, as U.S. struggles with shortage. Available at www.washingtonpost.com/health/2021/12/27/omicron-covid-test-shortage-fauci/. [Online; posted 27 December 2021].

POLLOCK, A. M. and LANCASTER, J. (2020). Asymptomatic transmission of Covid-19. *BMJ* **371** m4851.

POLLOCK, N. R., BERLIN, D., SMOLE, S. C., MADOFF, L. C., BROWN, C., HENDERSON, K., LARSEN, E., HAY, J., GABRIEL, S. et al. (2021). Implementation of SARS-Cov2 screening in K-12 schools using in-school pooled molecular testing and deconvolution by rapid antigen test. *Journal of Clinical Microbiology* **59** 1–7.

PUBLIC HEALTH ENGLAND (2020). Understanding cycle threshold (Ct) in SARS-CoV-2 RT-PCR: A guide for health protection teams. Available at www.gov.uk/government/publications/cycle-threshold-ct-in-sars-cov-2-rt-pcr.

RANNAN-ELIYA, R. P., WIJEMUNIGE, N., GUNAWARDANA, J., AMARASINGHE, S. N., SIVAGNANAM, I., FONSEKA, S., KAPUGE, Y. and SIGERA, C. P. (2021). Increased intensity of PCR testing reduced COVID-19 transmission within countries during the first pandemic wave: Study examines increased intensity of reverse transcription–polymerase chain reaction (PCR) testing and its impact on COVID-19 transmission. *Health Aff.* 10–1377.

REWLEY, J. (2020). Specimen pooling to conserve additional testing resources when persons' infection status is correlated: A simulation study. *Epidemiology* **31** 832–835. https://doi.org/10.1097/EDE.0000000000001244

RITCHIE, H., MATHIEU, E., RODÉS-GUIRAO, L., APPEL, C., GIATTINO, C., ORTIZ-OSPINA, E., HASELL, J., MACDONALD, B., BELTEKIAN, D. et al. (2020). Coronavirus pandemic (COVID-19). *Our World in Data*. ourworldindata.org/coronavirus.

SIMAS, A. M., CROTT, J. W., SEDORE, C., ROHRBACH, A., MONACO, A. P., GABRIEL, S. B., LENNON, N., BLUMENSTIEL, B. and GENCO, C. A. (2021). Pooling for SARS-CoV2 surveillance: Validation and strategy for implementation in K-12 schools. *Frontiers in Public Health* **9** 1–7.

SPIELBERGER, B. D., GOERNE, T., GEWENIGER, A., HENNEKE, P. and ELLING, R. (2021). Intra-household and close-contact SARS-CoV-2 transmission among children—A systematic review. *Frontiers in Pediatrics* **9**. https://doi.org/10.3389/fped.2021.613292

STEVENS, J. P., HORNG, S., O'DONOGHUE, A., MORAVICK, S. and WEISS, A. (2021). How one Boston hospital built a Covid-19 forecasting system. *Harvard Business Review*.

TOM, M. R. and MINA, M. J. (2020). To interpret the SARS-CoV-2 test, consider the cycle threshold value. *Clin. Infect. Dis.* **71** 2252–2254. https://doi.org/10.1093/cid/ciaa619

TSO, C. F., GARIKIPATI, A., GREEN-SAXENA, A., MAO, Q. and DAS, R. (2021). Correlation of population SARS-CoV-2 cycle threshold values to local disease dynamics: Exploratory observational study. *JMIR Public Health and Surveillance* **7**.

TU, X. M., LITVAK, E. and PAGANO, M. (1995). On the informativeness and accuracy of pooled testing in estimating prevalence of a rare disease: Application to HIV screening. *Biometrika* **82** 287–297. MR1354229 https://doi.org/10.1093/biomet/82.2.287

WANG, H., HOGAN, C. A., MILLER, J. A., SAHOO, M. K., HUANG, C. H., MFUH, K. O., SIBAI, M., ZEHNDER, J., HICKEY, B. et al. (2021). Performance of nucleic acid amplification tests for detection of severe acute respiratory syndrome coronavirus 2 in prospectively pooled specimens. *Emerg. Infec. Dis.* **27** 92–103.

WEIN, L. M. and ZENIOS, S. A. (1996). Pooled testing for HIV screening: Capturing the dilution effect. *Oper. Res.* **44** 543–569.

YAMAMURA, K. and HINO, A. (2007). Estimation of the proportion of defective units by using group testing under the existence of a threshold of detection. *Comm. Statist. Simulation Comput.* **36** 949–957. MR2415696 https://doi.org/10.1080/03610910701539278

ZHANG, Y., LI, Y., WANG, L., LI, M. and ZHOU, X. (2020). Evaluating transmission heterogeneity and super-spreading event of COVID-19 in a Metropolis of China. *Int. J. Environ. Res. Public Health* **17** 3705.

ZHANG, J., TIAN, S., LOU, J. and CHEN, Y. (2020). Familial cluster of COVID-19 infection from an asymptomatic. *Critical Care* **24** 7–9.

ZHOU, Y., WANG, L., ZHANG, L., SHI, L., YANG, K., HE, J., ZHAO, B., OVERTON, W., PURKAYASTHA, S. et al. (2020). A spatiotemporal epidemiological prediction model to inform county-level COVID-19 risk in the United States. *Harvard Data Science Review*.

Statistical Science
2022, Vol. 37, No. 2, 251–265
https://doi.org/10.1214/22-STS861

Learning and Predicting from Dynamic Models for COVID-19 Patient Monitoring

Zitong Wang, Mary Grace Bowring, Antony Rosen, Brian Garibaldi, Scott Zeger and Akihiko Nishimura

Abstract. COVID-19 has challenged health systems to learn how to learn. This paper describes the context, methods and challenges for learning to improve COVID-19 care at one academic health center. Challenges to learning include: (1) choosing a right clinical target; (2) designing methods for accurate predictions by borrowing strength from prior patients' experiences; (3) communicating the methodology to clinicians so they understand and trust it; (4) communicating the predictions to the patient at the moment of clinical decision; and (5) continuously evaluating and revising the methods so they adapt to changing patients and clinical demands.

To illustrate these challenges, this paper contrasts two statistical modeling approaches—prospective longitudinal models in common use and retrospective analogues complementary in the COVID-19 context—for predicting future biomarker trajectories and major clinical events. The methods are applied to and validated on a cohort of 1678 patients who were hospitalized with COVID-19 during the early months of the pandemic. We emphasize graphical tools to promote physician learning and inform clinical decision making.

Key words and phrases: Longitudinal data analysis, prediction, inverse regression, decision support, statistical graphics.

REFERENCES

[1] ANDRINOPOULOU, E.-R., HARHAY, M. O., RATCLIFFE, S. J. and RIZOPOULOS, D. (2021). Reflections on modern methods: Dynamic prediction using joint models of longitudinal and time-to-event data. *Int. J. Epidemiol.*

[2] ANDRINOPOULOU, E.-R., RIZOPOULOS, D., TAKKENBERG, J. J. M. and LESAFFRE, E. (2014). Joint modeling of two longitudinal outcomes and competing risk data. *Stat. Med.* **33** 3167–3178. MR3260535 https://doi.org/10.1002/sim.6158

[3] BATES, J. M. and GRANGER, C. W. (1969). The combination of forecasts. *J. Oper. Res. Soc.* **20** 451–468.

[4] BENNETT, T. D., MOFFITT, R. A., HAJAGOS, J. G., AMOR, B., ANAND, A., BISSELL, M. M., BRADWELL, K. R., BREMER, C., BYRD, J. B. et al. (2021). Clinical characterization and prediction of clinical severity of SARS-CoV-2 infection among US adults using data from the US national COVID cohort collaborative. *Journal of the American Medical Association Network Open* **4** e2116901–e2116901.

[5] BERWICK, D. M. and HACKBARTH, A. D. (2012). Eliminating waste in US health care. *J. Am. Med. Assoc.* **307** 1513–1516.

[6] BOWRING, M. G., WANG, Z., XU, Y., BETZ, J., MUSCHELLI, J., GARIBALDI, B. T. and ZEGER, S. L. (2021). Outcome-stratified analysis of biomarker trajectories for patients infected with severe acute respiratory syndrome Coronavirus 2. *Am. J. Epidemiol.* **190** 2094–2106.

[7] BUDRIONIS, A. and BELLIKA, J. G. (2016). The learning healthcare system: Where are we now? A systematic review. *J. Biomed. Inform.* **64** 87–92. https://doi.org/10.1016/j.jbi.2016.09.018

[8] CHEN, X., GAO, W., LI, J., YOU, D., YU, Z., ZHANG, M., SHAO, F., WEI, Y., ZHANG, R. et al. (2021). A predictive

Zitong Wang is a PhD student, Department of Biostatistics, The Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland, USA (e-mail: zwang238@jhmi.edu). Mary Grace Bowring is an MD/PhD student, Departments of Biomedical Engineering and Biostatistics, The Johns Hopkins University School of Medicine, Baltimore, Maryland, USA (e-mail: mbowrin1@jhmi.edu). Antony Rosen is Mary Betty Stevens Professor of Medicine and Vice Dean for Research, The Johns Hopkins University School of Medicine, Baltimore, Maryland, USA (e-mail: arosen@jhmi.edu). Brian Garibaldi is Associate Professor, Department of Medicine, The Johns Hopkins University School of Medicine, Baltimore, Maryland, USA (e-mail: bgariba1@jhmi.edu). Scott L. Zeger (co-senior author) is John C. Malone Professor, Department of Biostatistics and Medicine, The Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland, USA (e-mail: sz@jhu.edu). Akihiko (Aki) Nishimura (co-senior author) is Assistant Professor, Department of Biostatistics, The Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland, USA (e-mail: aki.nishimura@jhu.edu).

paradigm for COVID-19 prognosis based on the longitudinal measure of biomarkers. *Brief. Bioinform.*

[9] CHI, Y.-Y. and IBRAHIM, J. G. (2006). Joint models for multivariate longitudinal and multivariate survival data. *Biometrics* **62** 432–445. MR2227491 https://doi.org/10.1111/j.1541-0420.2005.00448.x

[10] COVID-19 DASHBOARD BY THE CENTER FOR SYSTEMS SCIENCE AND ENGINEERING (CSSE) AT JOHNS HOPKINS UNIVERSITY (JHU). https://coronavirus.jhu.edu/map.html. Accessed: 2021-10-14.

[11] DABBAH, M. A., REED, A. B., BOOTH, A. T. C., YASSAEE, A., DESPOTOVIC, A., KLASMER, B., BINNING, E., ARAL, M., PLANS, D. et al. (2021). Machine learning approach to dynamic risk modeling of mortality in COVID-19: A UK Biobank study. *Sci. Rep.* **11**.

[12] DUAN, N. and LI, K.-C. (1991). Slicing regression: A link-free regression method. *Ann. Statist.* **19** 505–530. MR1105834 https://doi.org/10.1214/aos/1176348109

[13] FDA. Clinical decision support software. Accessed 2022-01-12.

[14] GALLO MARIN, B., AGHAGOLI, G., LAVINE, K., YANG, L., SIFF, E. J., CHIANG, S. S., SALAZAR-MATHER, T. P., DUMENCO, L., SAVARIA, M. C. et al. (2021). Predictors of COVID-19 severity: A literature review. *Reviews in Medical Virology* **31** 1–10.

[15] GARIBALDI, B. T., FIKSEL, J., MUSCHELLI, J., ROBINSON, M. L., ROUHIZADEH, M., PERIN, J., SCHUMOCK, G., NAGY, P., GRAY, J. H. et al. (2021). Patient trajectories among persons hospitalized for COVID-19: A cohort study. *Ann. Intern. Med.* **174** 33–41.

[16] GELMAN, A. and HILL, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge Univ. Press, Cambridge.

[17] HEAGERTY, P., LUMLEY, T. and PEPE, M. (2000). Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics* **56** 337–44.

[18] HOETING, J. A., MADIGAN, D., RAFTERY, A. E. and VOLINSKY, C. T. (1999). Bayesian model averaging: A tutorial. *Statist. Sci.* **14** 382–417. MR1765176 https://doi.org/10.1214/ss/1009212519

[19] JIANG, C.-R., YU, W. and WANG, J.-L. (2014). Inverse regression for longitudinal data. *Ann. Statist.* **42** 563–591. MR3210979 https://doi.org/10.1214/13-AOS1193

[20] KARNIK, K. (2014). FDA regulation of clinical decision support software. *Journal of Law and the Biosciences* **1** 202–208.

[21] LIFE EXPECTANCY AT 65. https://data.oecd.org/healthstat/life-expectancy-at-65.htm#indicator-chart. Accessed: 2021-10-14.

[22] LITTLE, R. J. (1993). Pattern-mixture models for multivariate incomplete data. *J. Amer. Statist. Assoc.* **88** 125–134.

[23] MICHIELS, B., MOLENBERGHS, G. and LIPSITZ, S. R. (1999). Selection models and pattern-mixture models for incomplete data with covariates. *Biometrics* **55** 978–983.

[24] OLSEN, L., AISNER, D. and MCGINNIS, J. M. (2007). *The Learning Healthcare System: Workshop Summary*. The National Academies Press, Washington, DC.

[25] POURHOMAYOUN, M. and SHAKIBI, M. (2021). Predicting mortality risk in patients with COVID-19 using machine learning to help medical decision-making. *Smart Health* (*Amsterdam, Netherlands*) **20** 100178.

[26] RIZOPOULOS, D. (2012). *Joint Models for Longitudinal and Time-to-Event Data: With Applications in R*. CRC Press, Boca Raton.

[27] ROSEN, A. and ZEGER, S. L. (2019). Precision medicine: Discovering clinically relevant and mechanistically anchored disease subgroups at scale. *J. Clin. Invest.* **129** 944–945. https://doi.org/10.1172/JCI126120

[28] SCHUEMIE, M. J., RYAN, P. B., DUMOUCHEL, W., SUCHARD, M. A. and MADIGAN, D. (2014). Interpreting observational studies: Why empirical calibration is needed to correct p-values. *Stat. Med.* **33** 209–218. MR3146759 https://doi.org/10.1002/sim.5925

[29] SEIB, G. F. (2008). In crisis, opportunity for Obama. Accessed: 2021-10-14.

[30] SHORTLIFFE, E. H. and SEPÚLVEDA, M. J. (2018). Clinical decision support in the era of artificial intelligence. *J. Amer. Statist. Assoc.* **320** 2199–2200.

[31] TUKEY, J. W. (1963). Borrowing strength in a diversified situation Technical Report Princeton Univ. Working Paper, Statistical Techniques Research Group.

[32] TUTZ, G. and SCHMID, M. (2016). *Modeling Discrete Time-to-Event Data. Springer Series in Statistics*. Springer, Cham. MR3497009 https://doi.org/10.1007/978-3-319-28158-2

[33] WANG, J. M., LIU, W., CHEN, X., MCRAE, M. P., MCDEVITT, J. T. and FENYÖ, D. (2021). Predictive modeling of morbidity and mortality in patients hospitalized with COVID-19 and its clinical implications: Algorithm development and interpretation. *J. Med. Internet Res.* **23** e29514.

[34] WOLLENSTEIN-BETECH, S., CASSANDRAS, C. G. and PASCHALIDIS, I. C. (2020). Personalized predictive models for symptomatic COVID-19 patients using basic preconditions: Hospitalizations, mortality, and the need for an ICU or ventilator. *International Journal of Medical Informatics* **142**.

[35] WONGVIBULSIN, S., GARIBALDI, B. T., ANTAR, A. A. R., WEN, J., WANG, M.-C., GUPTA, A., BOLLINGER, R., XU, Y., WANG, K. et al. (2021). Development of severe COVID-19 adaptive risk predictor (SCARP), a calculator to predict severe disease or death in hospitalized patients with COVID-19. *Ann. Intern. Med.* **174** 777–785. https://doi.org/10.7326/M20-6754

[36] WONGVIBULSIN, S., WU, K. C. and ZEGER, S. L. (2020). Clinical risk prediction with random forests for survival, longitudinal, and multivariate (RF-SLAM) data analysis. *BMC Med. Res. Methodol.* **20** 1–14.

[37] YAEGER, K. A., MARTINI, M., YANIV, G., OERMANN, E. K. and COSTA, A. B. (2019). United States regulatory approval of medical devices and software applications enhanced by artificial intelligence. *Health Policy and Technology* **8** 192–197.

[38] ZHANG, Y. et al. (2018). Interpretable dynamic treatment regimes. *J. Amer. Statist. Assoc.* **113** 1541–1549. https://doi.org/10.1080/01621459.2017.1345743

Statistical Science
2022, Vol. 37, No. 2, 266–269
https://doi.org/10.1214/22-STS855

Seven Principles for Rapid-Response Data Science: Lessons Learned from Covid-19 Forecasting

Bin Yu and Chandan Singh

Abstract. In this article, we take a step back to distill seven principles out of our experience in the spring of 2020, when our 12-person rapid-response team used skills of data science and beyond to help distribute 340,000+ units of Covid PPE. This process included tapping into domain knowledge of epidemiology and medical logistics chains, curating a relevant data repository, developing models for short-term county-level death forecasting in the US, and building a website for sharing visualization (an automated AI machine). The principles are described in the context of working with Response4Life, a then-new nonprofit organization, to illustrate their necessity. Many of these principles overlap with those in standard data-science teams, but an emphasis is put on dealing with problems that require rapid response, often resembling agile software development. The technical work from this rapid response project resulted in a paper (Altieri et al. (2021)); see also this interview for more background (Yu and Meng (2021)).

Key words and phrases: Coronavirus, forecasting, county-level, data-science.

REFERENCES

[1] ALTIERI, N., BARTER, R. L., DUNCAN, J., DWIVEDI, R., KUMBIER, K., LI, X., NETZORG, R., PARK, B., SINGH, C. et al. (2021). Curating a Covid-19 data repository and forecasting county-level death counts in the United States. *Harvard Data Science Review*. https://hdsr.mitpress.mit.edu/pub/p6isyf0g. https://doi.org/10.1162/99608f92.1d4e0dae

[2] COCKBURN, A. and HIGHSMITH, J. (2001). Agile software development, the people factor. *Computer* **34** 131–133.

[3] PEDREGOSA, F., VAROQUAUX, G., GRAMFORT, A. et al. (2011). Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12** 2825–2830. MR2854348

[4] SCHULLER, G. D., YU, B., HUANG, D. and EDLER, B. (2002). Perceptual audio coding using adaptive pre-and post-filters and lossless compression. *IEEE Trans. Speech Audio Process.* **10** 379–390.

[5] SEABOLD, S. and PERKTOLD, J. (2010). Statsmodels: Econometric and statistical modeling with python. In *Proceedings of the 9th Python in Science Conference, Austin, TX* **57** 61.

[6] SINGH, C., NASSERI, K., TAN, Y. S., TANG, T. and YU, B. (2021). (2021). imodels: a python package for fitting interpretable models. *Journal of Open Source Software* **6** 3192. https://doi.org/10.21105/joss.03192

[7] VOVK, V., GAMMERMAN, A. and SHAFER, G. (2005). *Algorithmic Learning in a Random World*. Springer, New York. MR2161220

[8] YU, B. and MENG, X.-L. (2021). An interview with Bin Yu. *Harvard Data Science Review*.

Bin Yu is Professor, Statistics Department and EECS Department, University of California, Berkeley, Berkeley, California 94720-3860, USA (e-mail: binyu@berkeley.edu). Chandan Singh is a PhD Candidate, EECS Department, University of California, Berkeley, Berkeley, California 94720-3860, USA (e-mail: cs1@berkeley.edu).

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

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.

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 (Long-COVIDKidsDK): 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

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).

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

Statistical Science
2022, Vol. 37, No. 2, 278–283
https://doi.org/10.1214/22-STS860

Lessons Learned from the COVID-19 Pandemic: A Statistician's Reflection

Xihong Lin

Abstract. In this article, I will discuss my experience as a statistician involved in COVID-19 research in multiple capacities in the last two years, especially in the early phase of the pandemic. I will reflect on the challenges and the lessons I have learned in pandemic research regarding data collection and access, epidemic modeling and data analysis, open science and real time dissemination of research findings, implementation science, media and public communication, and partnerships between academia, government, industry and civil society. I will also make several recommendations on navigating the next stage of the pandemic and preparing for future pandemics.

Key words and phrases: COVID-19 research, data collection and access, data analysis, epidemic modeling, implementation science, open science, media and public communication.

REFERENCES

ALLEN, E., ALTAE-TRAN, H., BRIGGS, J., JIN, X., MCGEE, G., SHI, A., RAGHAVAN, R., KAMARIZA, M., NOVA, N., PERETA, A., OTHERS, O., KING, G., PROBASCO, R., CHENG, R., SILBERMANN, B., ZHANG, F. and LIN, X. (2020). Population-scale longitudinal mapping of COVID-19 symptoms, behaviour and testing. *Nat. Hum. Behav.*. **4** 972–982.

ADAM, D. (2020). A guide to R—the pandemic's misunderstood metric. *Nat. News*. Available at https://www.nature.com/articles/d41586-020-02009-w.

BRADLEY, V. C., KURIWAKI, S., ISAKOV, M., SEJDINOVIC, D., MENG, X.-L. and FLAXMAN, S. (2021). Unrepresentative big surveys significantly overestimated US vaccine uptake. *Nature* **600** 695–700. https://doi.org/10.1038/s41586-021-04198-4

CLARK, G. (2020). COVID-19 pandemic: some lessons learned so far. Available at https://publications.parliament.uk/pa/cm5801/cmselect/cmsctech/correspondence/200518-Chair-to-Prime-Minister-re-COVID-19-pandemic-some-lessons-learned-so-far.pdf.

DEAN, N. (2022). Tracking COVID-19 infections: Time for change. Available at https://www.nature.com/articles/d41586-022-00336-8.

FRASER, N., BRIERLEY, L., DEY, G., POLKA, J. K., PÁLFY, M., NANNI, F. and COATES, J. A. (2021). The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape. *PLoS Biol.* **19** e3000959.

HAO, X., CHENG, S., WU, D., WU, T., LIN, X. and WANG, C. (2020). Reconstruction of the full transmission dynamics of COVID-19 in Wuhan. *Nature* **584** 420–424.

JEWELL, N. P. (2021). Statistical Models for COVID-19 Incidence, Cumulative Prevalence, and R_t. *J. Amer. Statist. Assoc.* **116** 1578–1582. MR4353695 https://doi.org/10.1080/01621459.2021.1983436

LARSEN, D. A. and WIGGINTON, K. R. (2020). Tracking COVID-19 with wastewater. *Nat. Biotechnol.* **38** 1151–1153.

NATURE EDITORIAL TEAM (2021). How epidemiology has shaped the COVID pandemic. Available at https://www.nature.com/articles/d41586-021-00183-z.

NATURE EDITORIAL TEAM (2022). This is no time to stop tracking COVID-19. Available at https://www.nature.com/articles/d41586-022-00788-y.

PAN, A., LIU, L., WANG, C., GUO, H., HAO, X., WANG, Q., HUANG, J., HE, N., YU, H. et al. (2020). Association of public health interventions with the epidemiology of the COVID-19 outbreak in Wuhan, China. *J. Am. Med. Assoc.* **323** 1915–1923.2.

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

RYAN, M. (2020). https://7news.com.au/lifestyle/health-wellbeing/speed-trumps-perfection-boss-of-world-health-organisation-issues-blunt-coronavirus-warning-c-747300.

SHI, A., GAYNOR, S. M., DEY, R., ZHANG, H., QUICK, C. and LIN, X. (2022). COVID-19 Spread Mapper: A multi-resolution, unified framework and open-source tool. *Bioinformatics* btac129.

WANG, C., LIU, L., HAO, X., GUO, H., WANG, Q., HUANG, J., HE, N., YU, H., LIN, X. et al. (2020). Evolving epidemiology and impact of non-pharmaceutical interventions on the outbreak of coronavirus disease 2019 in Wuhan, China. *MedRxiv*. https://doi.org/10.1101/2020.03.03.20030593

Xihong Lin is Professor of Biostatistics at Harvard T.H. Chan School of Public Health, Boston, Massachusetts 02115, USA, and Professor of Statistics at Harvard University, Cambridge, Massachusetts 02138-2901, USA (e-mail: xlin@hsph.harvard.edu).

Statistical Science
2022, Vol. 37, No. 2, 284–288
https://doi.org/10.1214/22-STS858

Data, Science, and Global Disasters

John M. Chambers

Abstract. The spread and impact of COVID-19 have disrupted human activities and energized a response of scientific activity on a remarkable, nearly unprecedented scale. This has somewhat distracted attention from a broad range of less immediate but fundamentally more serious global threats resulting from human actions. These can be collectively labelled the *anthropocene disasters*.

Science cannot itself prevent or mitigate them. To do so requires a global policy resolve not currently existing. When and if that resolve emerges, science will be essential for guiding action. This science will be radically data-intensive, global and inclusive. Teams will be required that include the best and most motivated individuals from all relevant scientific disciplines, plus members knowledgable about implementing likely policy recommendations. Such participants must be attracted to join and then properly supported and rewarded–not likely with current academic structures. Some insights can be gained from the recent experience with COVID-19 and the much less recent example of research at Bell Labs.

Key words and phrases: Data science, climate change, biodiversity, pandemic, Bell Labs.

REFERENCES

[1] CHAMBERS, J. M. (2020). S, R, and data science. In *Proc. ACM Program. Lang.* **4** Article 84. Association for Computing Machinery, New York, NY, United States. https://doi.org/10.1145/3386334

[2] DAVIES, J. (2016). *The Birth of the Anthropocene*. Univ. California Press, Oakland, CA, USA.

[3] DÍAZ, S. (2022). A fabric of life view of the world. *Science* **375** 1204. https://doi.org/10.1126/science.abp8336

[4] GERTNER, J. (2013). *The Idea Factory*: *Bell Labs and the Great Age of American Innovation*. Penguin, Baltimore.

[5] IPCC (2022). Sixth assessment report. International Panel on Climate Change.

[6] JAHJA, M., CHIN, A. and TIBSHIRANI, R. J. (2022). Real-time estimation of Covid-19 infections: Deconvolution and sensor fusion. *Statist. Sci.* **37** 207–228.

[7] JHA, P., DESHMUKH, Y., TUMBE, C., SURAWEERA, W., BHOWMICK, A., SHARMA, S., NOVOSAD, P., FU, S. H., NEWCOMBE, L. 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

[8] KEITT, T. H. and ABELSON, E. S. (2021). Ecology in the age of automation. *Science* **373** 858–859. https://doi.org/10.1126/science.abi4692

[9] LUGHADHA, E. N., BACHMAN, S. P., LEÃO, T. C. C., FOREST, F., HALLEY, J. M., MOAT, J. et al. (2020). Extinction risk and threats to plants and fungi. *Plants, People, Planet* **2** 389–408. https://doi.org/10.1002/ppp3.10146

[10] MACLEOD, M., ARP, H. P. H., TEKMAN, M. B. and JAHNKE, A. (2021). The global threat from plastic pollution. *Science* **373** 61–65. https://doi.org/10.1126/science.abg5433

[11] MUKHERJEE, B. (2022). Being a public health statistician during a global pandemic. *Statist. Sci.* **37** 270–277.

[12] NARDUCCI, J., QUINTAS-SORIANO, C., CASTRO, A., SOM-CASTELLANO, R. and BRANDT, J. S. (2019). Implications of urban growth and farmland loss for ecosystem services in the western United States. *Land Use Policy* **86** 1–11.

[13] NICHOLSON, G., BLANGIARDO, M., BRIERS, M., DIGGLE, P. J., FJELDE, T. E., GE, H., GOUDIE, R. J. B., JERSAKOVA, R., KING, R. E. et al. (2022). Interoperability of statistical models in pandemic preparedness: Principles and reality. *Statist. Sci.* **37** 183–206.

[14] ROBINSON, S. K., THOMPSON III, F. R., DONOVAN, T. M., WHITEHEAD, D. R. and FAABORG, J. (1995). Regional forest fragmentation and the nesting success of migratory birds. *Science* **267** 1987–1990.

[15] SÁNCHEZ-BAYO, F. and WYCKHUYS, K. A. (2019). Worldwide decline of the entomofauna: A review of its drivers. *Biol. Conserv.* **232** 8–27.

[16] SCHIMANSKI, L. A. and ALPERIN, J. P. (2018). The evaluation of scholarship in academic promotion and tenure processes: Past, present, and future. F1000Research 7.

[17] STAFFORD, R. and JONES, P. J. (2019). Viewpoint–ocean plastic pollution: A convenient but distracting truth? *Mar. Policy* **103** 187–191.

John M. Chambers is Adjunct Professor in Statistics and Senior Advisor for Data Science, Stanford University, Stanford, California 94305-4020, USA. He was in Bell Labs Research from 1966 to 2005 (e-mail: jmc4@stanford.edu).

INSTITUTE OF MATHEMATICAL STATISTICS

(Organized September 12, 1935)

The purpose of the Institute is to foster the development and dissemination of the theory and applications of statistics and probability.

IMS OFFICERS

President: Krzysztof Burdzy, Department of Mathematics, University of Washington, Seattle, Washington 98195-4350, USA

President-Elect: Peter Bühlmann, Seminar für Statistik, ETH Zürich, 8092 Zürich, Switzerland

Past President: Regina Y. Liu, Department of Statistics, Rutgers University, Piscataway, New Jersey 08854-8019, USA

Executive Secretary: Edsel Peña, Department of Statistics, University of South Carolina, Columbia, South Carolina 29208-001, USA

Treasurer: Zhengjun Zhang, Department of Statistics, University of Wisconsin, Madison, Wisconsin 53706-1510, USA

Program Secretary: Annie Qu, Department of Statistics, University of California, Irvine, Irvine, CA 92697-3425, USA

IMS EDITORS

The Annals of Statistics. *Editors*: Enno Mammen, Institute for Applied Mathematics, Heidelberg University, 69120 Heidelberg, Germany. Lan Wang, Miami Herbert Business School, University of Miami, Coral Gables, FL 33124, USA

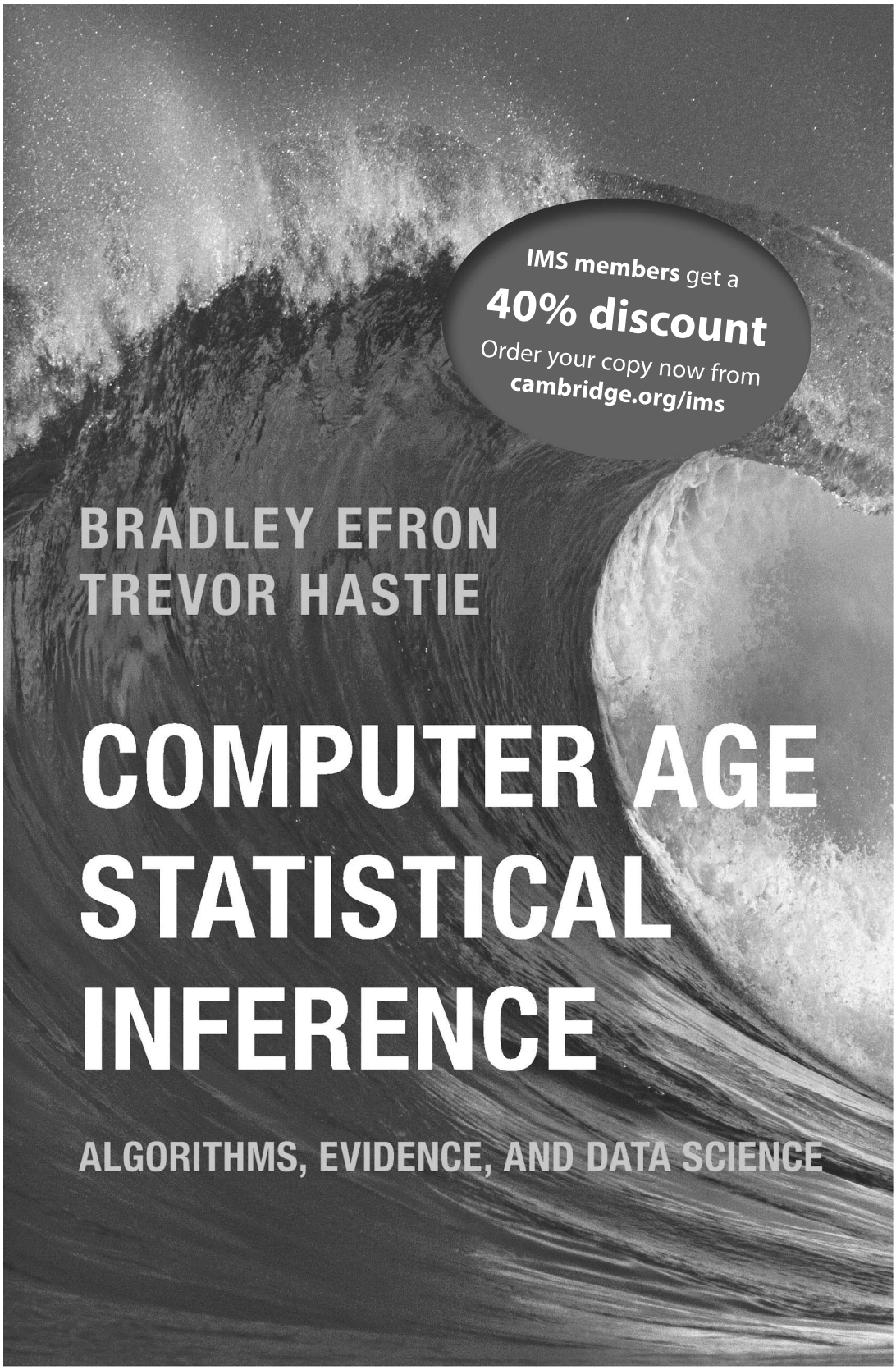
The Annals of Applied Statistics. *Editor-in-Chief*: Ji Zhu, Department of Statistics, University of Michigan, Ann Arbor, MI 48109, USA

The Annals of Probability. *Editors*: Alice Guionnet, Unité de Mathématiques Pures et Appliquées, ENS de Lyon, Lyon, France. Christophe Garban, Institut Camille Jordan, Université Claude Bernard Lyon 1, 69622 Villeurbanne, France

The Annals of Applied Probability. *Editors*: Kavita Ramanan, Division of Applied Mathematics, Brown University, Providence, RI 02912, USA. Qi-Man Shao, Department of Statistics and Data Science, Southern University of Science and Technology, Shenzhen, Guangdong 518055, P.R. China

Statistical Science. *Editor*: Sonia Petrone, Department of Decision Sciences, Università Bocconi, 20100 Milano MI, Italy

The IMS Bulletin. *Editor*: Vlada Limic, UMR 7501 de l'Université de Strasbourg et du CNRS, 7 rue René Descartes, 67084 Strasbourg Cedex, France


IMS members get a
40% discount
Order your copy now from
cambridge.org/ims
BRADLEY EFRON
TREVOR HASTIE
COMPUTER AGE
STATISTICAL
INFERENCE
ALGORITHMS, EVIDENCE, AND DATA SCIENCE