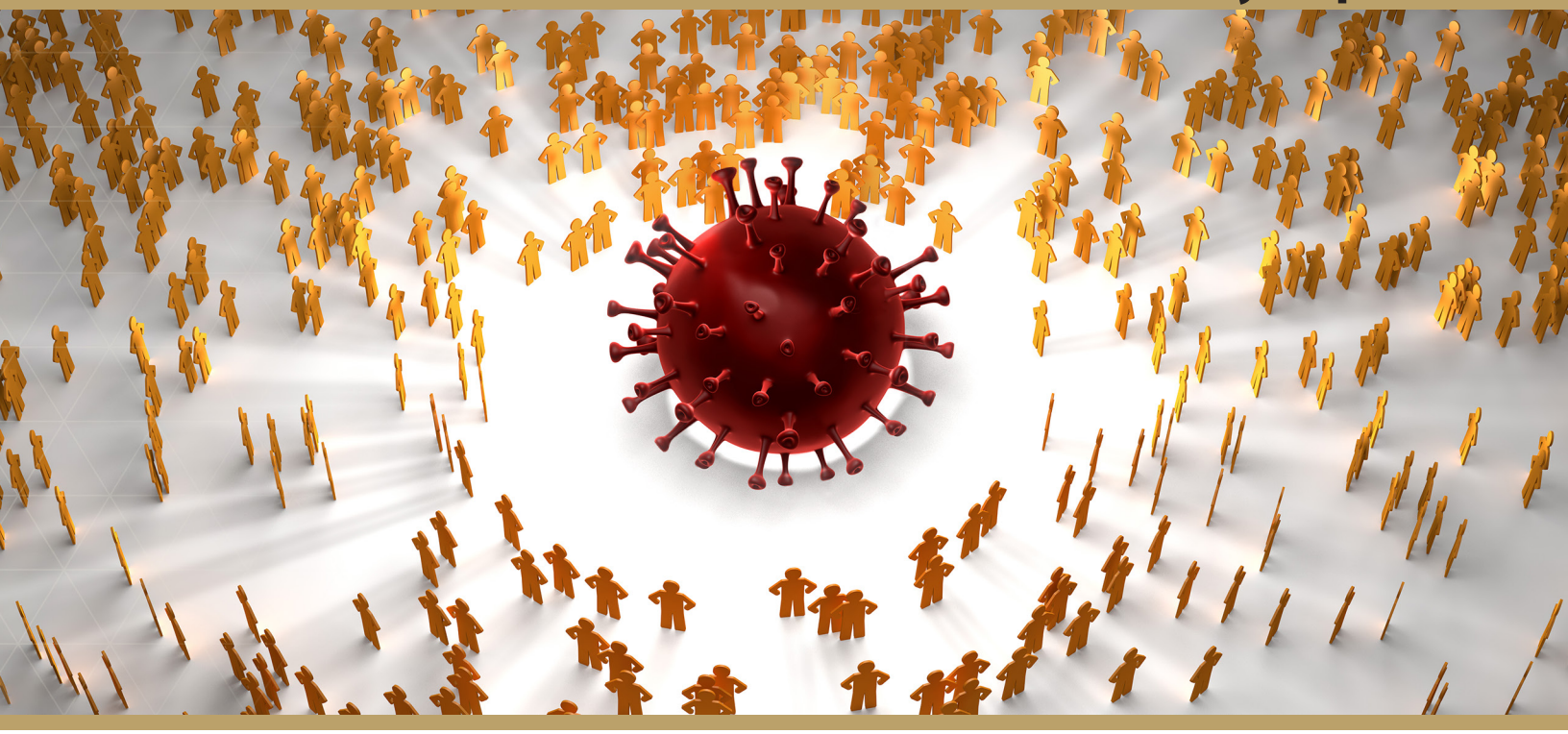
OPERATIONAL ANALYSIS FOR CORONAVIRUS TESTING

Recommendations for Practice

National Security Report



Alan Brown | Marc Mangel

JOHNS HOPKINS
APPLIED PHYSICS LABORATORY

OPERATIONAL ANALYSIS FOR CORONAVIRUS TESTING

Recommendations for Practice

Alan Brown

Marc Mangel

JOHNS HOPKINS
APPLIED PHYSICS LABORATORY

NSAD-R-21-014

Contents

Figures

Summary

Even though vaccines for coronavirus are increasingly available, it will be many months before sufficient herd immunity is achieved. Thus, testing remains a key tool for those managing health care and making policy decisions. Test errors, both false positive tests and false negative tests, mean that the surface positivity (the observed fraction of tests that are positive) does not accurately represent the incidence rate (the unobserved fraction of individuals infected with coronavirus). In this report, directed to individuals tasked with providing analytical advice to policymakers, we describe a method for translating from the surface positivity to a point estimate for the incidence rate, then to an appropriate range of values for the incidence rate, and finally to the risk (defined as the probability of including one infected individual) associated with groups of different sizes. The method is summarized in four equations that can be implemented in a spreadsheet or using a handheld calculator. We discuss limitations of the method and provide an appendix describing the underlying mathematical models.

Introduction

Even though coronavirus vaccines are increasingly available, it will be many months before sufficient herd immunity is achieved, so testing remains a key tool for those managing health care and making policy decisions.[1] In this report, we provide a method for individuals tasked with providing scientific advice concerning the interpretation of coronavirus tests.[2]

We focus on three questions. First, how does one go from surface positivity (the observed fraction of tests that are positive) to the incidence rate (the unobserved fraction of individuals infected with coronavirus), knowing that there are test errors?[3]

Second, how does one go from a point estimate for the incidence rate to a range of reasonably likely incidence rates? Third, how does one compute the risk (defined as the probability of including one infected individual) of coronavirus transmission in groups of different sizes, given the point estimate and range of values for the incidence rate?[4]

The Operational Situation

We assume that T coronavirus tests are administered to a population and, for each individual tested, the result is either a positive or negative test (Figure 1). A fraction f of these individuals are infected with coronavirus and are antigen positive. However, these individuals have a probability p_{FN} of a false negative result, where the test reports no infection when the individual is in fact infected. We assume that the probability of a false negative test is known.[5] The remaining individuals, a fraction $1 - f$ of the sample, are not infected (i.e., are antigen negative) but have a probability p_{FP} of a positive test result. We also assume that the probability of a false positive test is known.[6] It is likely that there is both medical and operational heterogeneity in the values of p_{FP} and p_{FN}. When heterogeneity can be characterized by distributions for the test, our method generalizes to account for it, as explained briefly in Appendix B here and more fully in a forthcoming report.[7]

Our goal is to begin with the test results, in this case P positive results out of T tests administered, giving surface positivity P/T, and obtain an estimate for the unobserved incidence rate of coronavirus. We denote this estimate by $\hat{f}$ and recognize that it can never be known precisely.

Given the unobserved incidence rate f, the probability of a false positive test p_{FP}, and the probability of a false negative test p_{FN}, the expected positivity rate $p_+(f)$ (the probability of a positive test) is composed of two terms: (1) the fraction of antigen-positive individuals tested whose test results are accurate plus (2) the fraction of

[1] Allen et al., *Roadmap to Pandemic Resilience*; and Thompson, "Testing Remains One of Our Best Tools to Fight COVID-19."

[2] Supporting analyses are described in Mangel and Brown, *Operational Analysis for COVID-19 Diagnostic Testing*. In Appendix A here, we describe the assumptions underlying our analysis.

[3] Test errors include false negatives (infected individuals give a negative test) (Oran and Topol, "Prevalence of Asymptomatic SARS-CoV-2 Infection"; and Watson, Whiting, and Brush, "Interpreting a Covid-19 Test Result") and false positives (noninfected individuals give a positive test) (He et al., "Diagnostic Performance between CT and Initial Real-Time RT-PCR"). False negative tests are unavoidable (Sethuraman, Jeremiah, and Ryo, "Interpreting Diagnostic Tests for SARS-CoV-2"); it may be possible to reduce the likelihood of false positive tests. What is important for our method is a recognition that these occur and can be at least approximately quantified.

[4] See Klein et al., *Stepping Back to School*, for the importance of this question in relation to reopening schools.

[5] Watson, Whiting, and Brush ("Interpreting a Covid-19 Test Result") report that in a clinical setting p_{FN} ranges from about 2% to 30%.

[6] He et al. ("Diagnostic Performance between CT and Initial Real-Time RT-PCR," Table 2) report p_{FP} a bit less than 0.1.

[7] Mangel and Brown, *Operational Analysis for COVID-19 Diagnostic Testing*.

antigen-negative individuals tested whose test results are inaccurate. That is

$$p_+(f) = f(1 - p_{FN}) + (1 - f)p_{FP}. \quad (1)$$

It is clear from this equation that the surface positivity rate is not equal to the incidence rate unless there are no test errors. The analytical challenge is that we observe the surface positivity but want to know the incidence rate.

From Surface Positivity to Incidence Rate

An estimate of the incidence rate $\hat{f}$ from the test results (P positive tests out of a total of T tests) with test errors p_{FN} and p_{FP} is

$$\hat{f} = \frac{P/T - p_{FP}}{1 - p_{FN} - p_{FP}}. \quad (2)$$

As explained in Appendix A, on average, the value of P will be $p_+(f)T$. Replacing P/T in Eq. 2 by $p_+(f)$ from Eq. 1 gives the result that $\hat{f} = f$. That is, on average the estimate in Eq. 2 will capture the true value of the incidence rate. However, we do not live in a world of on average. Rather, each testing event will give a single value of P/T that is randomly drawn from a distribution of possible values from which we construct the estimate $\hat{f}$. We would like to know a reasonable range for the estimate of the incidence rate.

From a Point Estimate of the Incidence Rate to a Range for the Incidence Rate

Eq. 2 provides a point estimate for the incidence rate given P/T and the test errors. Since P/T is a random variable, if one were to repeatedly sample the same population multiple times, the values of P/T would generally be different but around the expected positivity rate given in Eq. 1.

Variation in P/T generates variation in $\hat{f}$. Hence, the next step is to compute a range of possible values for $\hat{f}$. A range having 95% probability of including the true but unknown value of the incidence rate is[8]

$$Range(\hat{f}) = 3.92\sqrt{\frac{p_+(\hat{f})(1 - p_+(\hat{f}))}{T(1 - p_{FN} - p_{FP})^2}}. \quad (3)$$

Mangel and Brown[9] show via simulation testing that:

- $Range(\hat{f})$ is symmetrically distributed around the true range, $Range(f)$, which is obtained by replacing $\hat{f}$ by f in Eq. 3.
- The mean error between the two is a fraction of a percent, so that Eq. 3 is, on average, a very accurate characterization of the range.

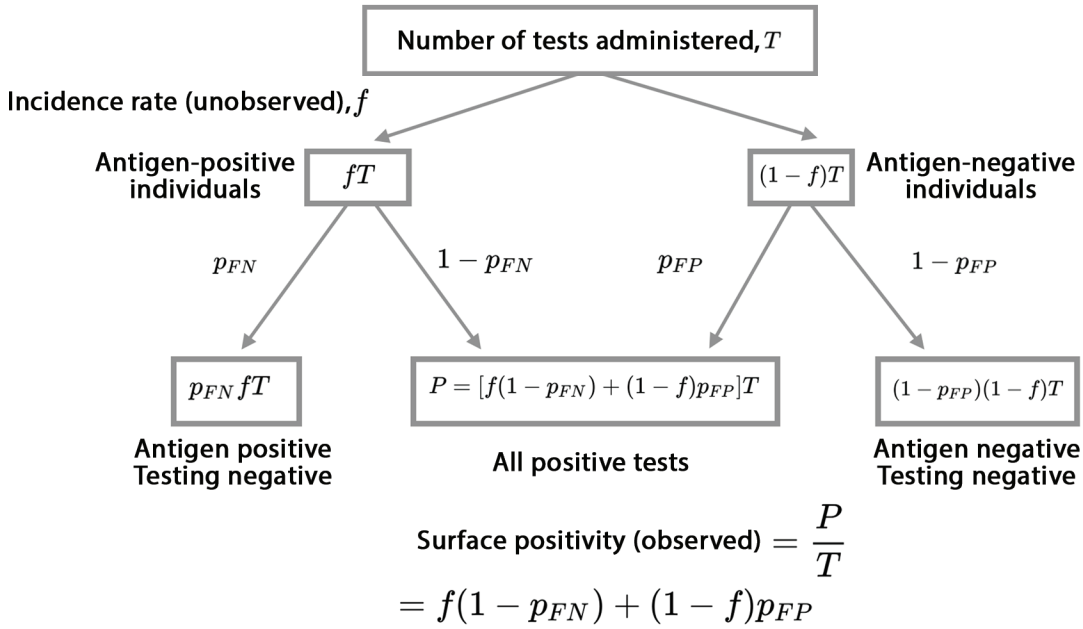
Thus, we are able to construct meaningful and reasonable lower and upper limits for the estimated infection rate as $\hat{f}_{lower} = \hat{f} - 0.5(Range(\hat{f}))$ and $\hat{f}_{upper} = \hat{f} + 0.5(Range(\hat{f}))$. In these equations, the factor of 0.5 accounts for $Range(\hat{f})$ being symmetrically distributed around $Range(f)$.

From Incidence Rate to Risk Associated with Groups of Different Sizes

Having gone from the surface positivity P/T to a point estimate $\hat{f}$ for the incidence rate, and then to a range $[\hat{f}_{lower}, \hat{f}_{upper}]$ for the incidence rate, we illustrate how these values can be used to compute

[8] In Appendix A, we give an explanation of the determination of range, which we interpret as a compatibility interval (McElreath, *Statistical Rethinking*, 54). It is a range of values for f that is compatible with the data and model and thus avoids the undesired implications of words such as *confidence* or *credible* (Morey et al., "Fallacy of Placing Confidence in Confidence Intervals"). See Mangel and Brown, *Operational Analysis for COVID-19 Diagnostic Testing,* for a fuller discussion.

[9] Mangel and Brown, *Operational Analysis for COVID-19 Diagnostic Testing.*



The surface positivity (fraction of positive tests) is observed, and from that we want to infer the unobserved incidence rate *f*. See text for further details.

Figure 1. Visual Representation of the Operational Situation

the risk associated with groups of different sizes, where we define risk as the probability that a group of g individuals contains at least one infected individual.

We denote this risk by $\mathcal{R}(g, \hat{f})$ to emphasize that it depends on both the size of the group and the point estimate for the incidence rate. If we replace $\hat{f}$ by $\hat{f}_{lower}$ or $\hat{f}_{upper}$, we obtain lower and upper bounds on the risk, respectively. We also emphasize, and will clarify in the example below, that the level of acceptable risk is a policy question, not an analytical question. That is, analysis can provide the relationships among group size, infection rate, and risk but cannot delineate the thresholds at which risk should be considered acceptable.

The risk associated with a group of size g when the point estimate of incidence rate is $\hat{f}$ is[10]

$$\mathcal{R}(g, \hat{f}) = 1 - (1 - \hat{f})^g. \qquad (4)$$

[10] In Appendix A we describe the underlying assumptions.

A Worked Example

To illustrate the application of these ideas, assume that 1,000 tests, with errors p_{FN} = 0.25 and p_{FP} = 0.05, were administered and that the number of positive tests was 112,[11] so that the surface positivity in the test pool was 0.112.

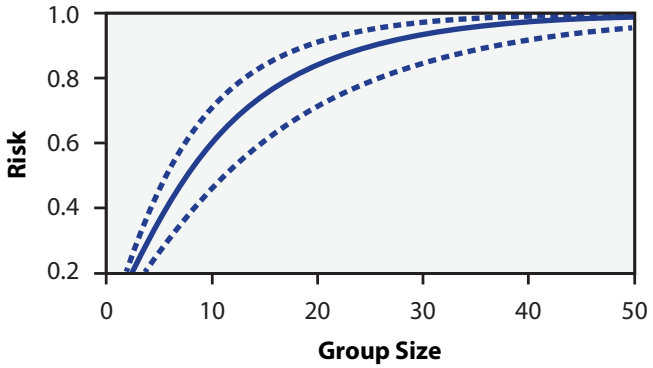
Using Eq. 2, we obtain the point estimate for the incidence rate $\hat{f}$ = 0.0886 from Eq. 2 and the lower and upper limits for the range $\hat{f}_{lower}$ = 0.0606 and $\hat{f}_{upper}$ = 0.1165 from Eq. 3 for the range and the equations given in the paragraph below Eq. 3.

Using Eq. 4 three times with $\hat{f}$, $\hat{f}_{lower}$, and $\hat{f}_{upper}$, respectively, we are able to make a plot of risk versus group size using the point, lower, and upper estimates for the incidence rate (Figure 2).

[11] We used a simulation with incidence rate f = 0.09 to generate the number of positive tests. R script for this simulation is available from the second author at marcmangel@protonmail.com. In an actual setting, the incidence rate remains forever unknown.

As described above, the construction of Figure 2 is scientific advice. That is, we developed analytical methods that produce a detailed description of the relationship between group size and risk (the probability of at least one coronavirus-infected person is in the group), but the choice of an acceptable level of risk is a policy decision.

For example, a decision-maker might consider a 50-50 risk acceptable. That person would then draw a horizontal line from 0.5 on the y axis to intersect the three curves, and from each intersection point draw a vertical line to the x axis to determine the group sizes associated with the 50-50 level of risk for the point, lower, and upper estimates of the infection rate. In such a case, groups of about 5–11 (using $\hat{f}_{upper}$, $\hat{f}$, and $\hat{f}_{lower}$ in Eq. 4) will be consistent with the chosen level of risk. A risk-prone decision-maker might consider an 80% risk acceptable and thus conclude that groups of around 12–25 are acceptable. A risk-averse decision-maker might consider a 20% risk acceptable, with groups of just a few individuals.



The methods embodied in Eqs. 1–4 allow one go to from the surface positivity to the risk of groups of different sizes, where risk is the probability that a group of size g includes at least one coronavirus-infected individual. These plots correspond to 1,000 tests, of which 112 are positive, when the test errors are p_{FP} = 0.05 and p_{FN} = 0.25. See text for further details on how the plots can be used.

Figure 2. Risk versus Group Size

Discussion

The method embodied in Eqs. 1–4 is straightforward, simple to use (it can be implemented in an Excel spreadsheet or using a handheld calculator), and has detailed justification.[12]

Although the discussion of group size and risk was written for policymakers, it also has implications for public health educators. For example, figures similar to Figure 2 can be used in a health education campaign to help individuals visualize what could happen if they are in contact with 5, 10, 30, 50, or 100 other individuals outside of their known circle. A visual tool such as a color-coded system will give individuals a way to assess the risk of their behaviors.

Finally, we note two limitations. First, this method only identifies the incidence rate from knowledge of surface positivity and the test errors. In particular, it does not allow us to separate symptomatic and asymptomatic infections. Doing so requires a more complicated version of the methods reported here.

Second, this method will not work all the time. For example, when the surface positivity P/T is less than the probability of a false positive test p_{FP}, Eq. 2 gives a negative estimate for the incidence rate. This is clearly a meaningless result. Surface positivity is most likely to be small in the latter stages of the pandemic, when the incidence rate is falling to 0 and random fluctuations cause P/T to be very small. At this latter stage of the pandemic, a very different method is needed.[13]

[12] Mangel and Brown, *Operational Analysis for COVID-19 Diagnostic Testing.*

[13] See Mangel and Brown, *Operational Analysis for COVID-19 Diagnostic Testing.*

Appendix A The Underlying Mathematical Models

Our analysis is based on two classical probability distributions. Given the probability of a positive test in Eq. 1, the number of positive tests follows a binomial distribution with parameters T and $p_+(f)$.[14]

The range computed in Eq. 3 is based on the Gaussian approximation to the binomial distribution,[15] which is highly accurate as long as the expected number of positive tests is bounded away from either 0 or T. Since we are considering test numbers in the thousands, the expected number of positive tests will be close to 0 or T when (1) the incidence rate is close to 0 (in which case, as discussed above, a different method is needed) or (2) the incidence rate is close to 1 (in which case everyone is infected and the incidence rate is not relevant to the current pandemic). The factor 3.92 in front of the square root in Eq. 3 comes from the 95% compatibility interval for a normal distribution with mean 0 and standard deviation 1.

The risk computed in Eq. 4 uses an assumption similar to the binomial distribution—that individuals in the group join independent of their infection status. Since $1 - \hat{f}$ is the probability that a single individual is not infected and probabilities are multiplied when they are independent, the probability that all individuals in a group of size g are uninfected is $1 - \hat{f}$ multiplied by itself g times.

[14] For example, Feller, *Introduction to Probability Theory and Its Applications,* 146 and following.

[15] Feller, *Introduction to Probability Theory and Its Applications,* 174 and following.

Appendix B Accounting for Heterogeneity in Test Errors

Eq. 2 requires single values of the test errors to estimate the incidence rate. If one knows the error probabilities for different kinds of tests and when those tests are done (so that a surface positivity is associated with single values for the test errors), Eq. 2 can be repeatedly applied.

An alternative is that one has distributions for the test errors and does not know which test corresponds to a given surface positivity but only that the test errors come from the distributions. In such a case, we can apply the delta-method[16] to generalize Eq. 2.

We assume that the means, variances, and covariance of test errors are known and denote these by $\overline{p}_{FN}$ and $\overline{p}_{FP}$ (mean test errors), $V_{p_{FN}}$, $V_{p_{FP}}$ (variance in test errors), and $Cov(p_{FN},p_{FP})$ (covariance between test errors), respectively.

In a forthcoming report,[17] we give the details of the derivation using the delta-method; here we report the result. For ease of notation, we let

$$\overline{f}(\overline{p}_{FN},\overline{p}_{FP}) = \frac{P/T - \overline{p}_{FP}}{1 - \overline{p}_{FN} - \overline{p}_{FP}}.$$

This is the estimate in Eq. 2 in which we use the mean values of the test errors.

The generalization of Eq. 2 is then

$$\begin{aligned}\mathcal{E}(\hat{f}) = \overline{f}(\overline{p}_{FN},\overline{p}_{FP}) &+ \frac{\overline{f}(\overline{p}_{FN},\overline{p}_{FP})}{(1-\overline{p}_{FN}-\overline{p}_{FP})^2}V_{p_{FN}} + \frac{\overline{f}(\overline{p}_{FN},\overline{p}_{FP})-1}{(1-\overline{p}_{FN}-\overline{p}_{FP})^2}V_{p_{FP}} \\ &+ \frac{2\overline{f}(\overline{p}_{FN},\overline{p}_{FP})-1}{(1-\overline{p}_{FN}-\overline{p}_{FP})^2}Cov(p_{FN},p_{FP}),\end{aligned} \tag{5}$$

where $\mathcal{E}$ denotes the expectation over the distribution of the test errors.

[16] Hilborn and Mangel, *Ecological Detective*, 58–59.

[17] Mangel and Brown, *Operational Analysis for COVID-19 Diagnostic Testing*.

Bibliography

Allen, Danielle, Sharon Block, Joshua Cohen, Peter Eckersley, M. Eifler, Lawrence Gostin, Darshan Goux, et al. *Roadmap to Pandemic Resilience. Massive Scale Testing, Tracing, and Supported Isolation (TTSI) as the Path to Pandemic Resilience for a Free Society*. Cambridge, MA: Edmond J. Safra Center for Ethics at Harvard University, April 20, 2020. https://ethics.harvard.edu/files/center-for-ethics/files/roadmap-topandemicresilience_updated_4.20.20_0.pdf.

Feller, W. *An Introduction to Probability Theory and Its Applications*. Vol. 1. New York: John Wiley & Sons, 1968.

He, Jian-Long, Lin Luo, Zhen-Dong Luo, Jian-Xun Lyu, Ming-Yen Ng, Xin-Ping Shen, and Zhibo Wen. "Diagnostic Performance between CT and Initial Real-Time RT-PCR for Clinically Suspected 2019 Coronavirus Disease (COVID-19) Patients outside Wuhan, China." *Respiratory Medicine* 168, article 105980 (2020): 1–5. https://doi.org/10.1016/j.rmed.2020.105980.

Hilborn, Ray, and Marc Mangel. *The Ecological Detective: Confronting Models with Data*. Princeton, NJ: Princeton University Press, 1997.

Klein, Daniel, Cliff Kerr, Dina Mistry, Edward Wenger, and Jamie Cohen. *Stepping Back to School. A Step-by-Step Look at COVID Introduction, Spread, and Exportation*. Institute for Disease Modeling (IDM), February 24, 2021. https://iazpvnewgrp01.blob.core.windows.net/source/archived/Stepping_Back_to_School.pdf.

Mangel, Marc, and Alan Brown. *Operational Analysis for COVID-19 Diagnostic Testing*. National Security Report. Laurel, MD: Johns Hopkins University Applied Physics Laboratory, forthcoming.

McElreath, Richard. *Statistical Rethinking. A Bayesian Course with Examples in R and Stan*. 2nd ed. Boca Raton, FL: CRC Press, 2020.

Morey, Richard D., Rink Hoekstra, Jeffrey N. Rouder, Michael D. Lee, and Eric-Jan Wagenmakers. "The Fallacy of Placing Confidence in Confidence Intervals." *Psychonomic Bulletin & Review* 23 (2016): 103–123. https://link.springer.com/article/10.3758/s13423-015-0947-8.

Oran, Daniel P., and Eric J. Topol. "Prevalence of Asymptomatic SARS-CoV-2 Infection. A Narrative Review." *Annals of Internal Medicine* 173, no. 5 (2020): 362–367. https://doi.org/10.7326/M20-3012.

Sethuraman, Nandini, Sundararaj Stanleyraj Jeremiah, and Akihide Ryo. "Interpreting Diagnostic Tests for SARS-CoV-2." *JAMA* 323, no. 22 (2021): 2249–2251. https://doi.org/10.1001/jama.2020.8259.

Thompson, Jennifer. "Testing Remains One of Our Best Tools to Fight COVID-19." Tacoma-Pierce County Health Department blog, February 12, 2021. https://www.tpchd.org/Home/Components/Blog/Blog/31919/333.

Watson, Jessica, Penny F. Whiting, and John E. Brush. "Interpreting a Covid-19 Test Result." *BMJ* 369, article m1808 (2020): 1–7. https://doi.org/10.1136/bmj.m1808.

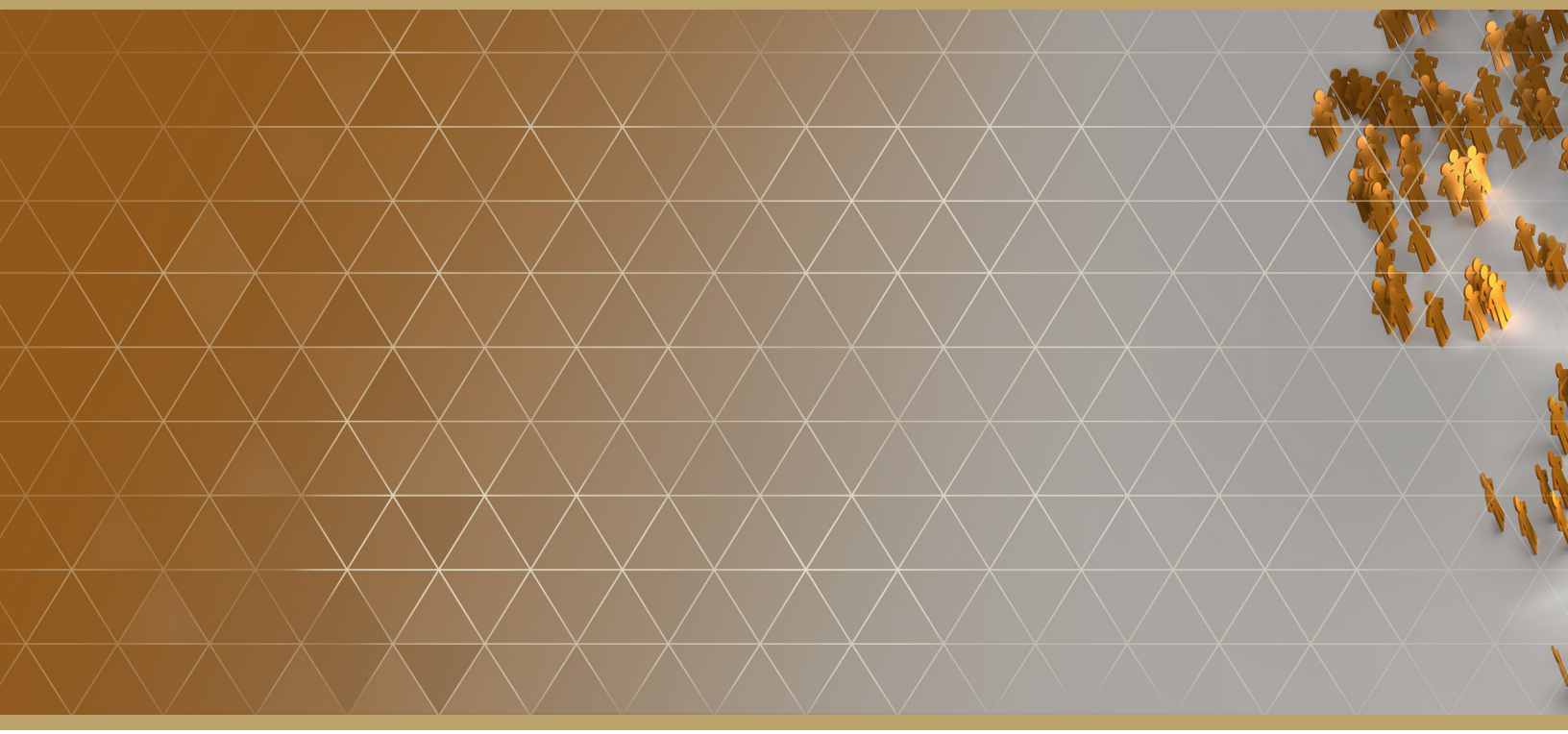
Acknowledgments

Marc Mangel was funded by a consulting contract with APL. We thank Jin Mou and Bethann Pflugeisen for comments on a previous version of the manuscript that improved presentation.

About the Authors

Dr. Alan Brown researches operational issues in APL's Asymmetric Operations Sector (AOS). Recently, he supported the National Response Coordination Center at FEMA, leading the future operations analysis cell during the first months of the COVID-19 pandemic. Currently, he leads analysis in support of development and execution of a serologic testing program for Howard County, Maryland. In addition to supporting these health-related analysis efforts, Dr. Brown conducts research on cyber issues as a member of the AOS Transformational Projects Office, serves as chief scientist for the AOS Mission Analysis Group, and was recently technical lead for systems engineering analysis and operational testing for the deputy secretary of defense special program Missile Defeat Time Critical Targeting. He joined APL in 2015 after thirty years at the Center for Naval Analyses (CNA), where his work combined analyses of capabilities, operations, resource requirements, and policies to inform senior-level decision-makers as they grappled with complex challenges. Dr. Brown led CNA's Operational Policy Team, which analyzed real-world operations and the policies that govern them, including mission-focused organization; war gaming; tactic, technique, and procedure development; and training for operational units. He holds a PhD from Brown University and a BS from Worcester Polytechnic Institute, both in physics.

Dr. Marc Mangel is distinguished professor emeritus at the University of California Santa Cruz and professor emeritus at the University of Bergen. He has applied operations and systems analysis widely, including search theory in agricultural pest control and fisheries, stochastic optimization for management of fisheries and understanding the life history patterns of salmon and southern ocean krill, and nonlinear dynamical systems in the study of disease. After receiving his PhD, Dr. Mangel worked in the Operations Evaluation Group (OEG) of the Center for Naval Analyses (CNA), including in a field position at Naval Air Station Whidbey Island. From there, he moved to the University of California Davis with the goal of doing OEG-style work with applications in fisheries and agriculture. In collaboration with colleagues at APL, he is now bringing ideas from population biology to describe variability, compromise, and recovery of cyber systems. During his career, Dr. Mangel has provided leadership and advice to many national and international organizations, most recently the Scientific Review Board of the International Pacific Halibut Commission. In 2013, he was the principal scientific expert for Australia in the International Court of Justice case *Whaling in the Antarctic: Australia v. Japan. New Zealand Intervening*. Dr. Mangel has authored and co-authored numerous journal articles and books, including *Decision and Control in Uncertain Resource Systems*, *Dynamic Modeling in Behavioral Ecology*, *The Ecological Detective: Confronting Models with Data*, and *The Theoretical Biologist's Toolbox*. He holds a BSc (physics) and an MSc (biophysics) from the University of Illinois and a PhD (applied mathematics and statistics) from the University of British Columbia; he was awarded a doctor of science *honoris causa* from the University of Guelph in 2014.



JOHNS HOPKINS
APPLIED PHYSICS LABORATORY