learned, for example, from the long-standing success of the arXiv e-print repository in the fields of physics, mathematics, and computer science, fueled by a combination of grants, in-kind support, and institutional memberships.

The struggle for control over information and knowledge looms large. When Berners-Lee created the World Wide Web, his intention was to enable researchers to share their work. Not only have our research communication tools and practices thus far fallen short of the decentralization that the Web made possible, but the evolution of the Web itself also reminds us that making vast amounts of linked data readily accessible to third parties can trigger a number of unintended consequences. The dominance of a limited number of social networks, shopping services, and search engines shows us how internet platforms based on data and analytics can tend toward monopoly. In the research information space, contracts are being negotiated establishing de facto terms and conditions for how data analytics services are being provided. Learned societies are being wooed. Research assessment metrics are being proposed. Building blocks for establishing discipline portals are being assembled. The time for the academic community to act in coordination is now. ■

REFERENCES AND NOTES

1. Press release, 8 April 2020; https://group.springernature.com/gp/group/media/ press-releases/springer-nature-plan-s/17877246.
2. K. Smith, Three Questions with Don Waters. *Council on Library and Information Resources* (*CLIR*) *Newsletter* 131 (2019); https://www.clir.org/2019/10/clir-issues-131/.
3. T. C. Bergstrom, P. N. Courant, R. P. McAfee, M. A. Williams, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 9425 (2014).
4. L. Hinchliffe, Transformative Agreements: A Primer. *The Scholarly Kitchen* (23 April 2019); https://scholarlykitchen.sspnet.org/2019/04/23/transformative-agreements/.
5. Leaked document on Elsevier negotiations sparks controversy, *ScienceGuide* (2019); www.scienceguide.nl/2019/11/leaked-document-on-elsevier-negotiations-sparks-controversy/.
6. B. Upton, Dutch universities agree 'new way of working' with Elsevier. Research Professional News. 12 December 2019; www.researchprofessionalnews.com/rr-news-europe-infrastructure-2019-12-dutch-universities-agree-new-way-of-working-with-elsevier/.
7. Elsevier Launches PracticeUpdate, a Free Information Portal for Physicians; www.elsevier.com/about/press-releases/clinical-solutions/elsevier-launches-practiceupdate,-a-free-information-portal-for-physicians.
8. Elsevier and Society of Petroleum Engineers Collaboration will Empower Geoscientists to Make More Informed and Confident Decisions; www.prnewswire.com/news-releases/elsevier-and-society-of-petroleum-engineers-collaboration-will-empower-geoscientists-to-make-more-informed-and-confident-decisions (2019).
9. SPARC Roadmap for Action (2019); https://sparcopen.org/our-work/roadmap-for-action/.
10. J. W. Maxwell *et al.*, Mind the Gap: A Landscape Analysis of Open Source Publishing Tools and Platforms, posted July 2019; https://mindthegap.pubpub.org/.

ACKNOWLEDGMENTS

We thank several anonymous reviewers. C.A. is a paid consultant to SPARC and was the lead author of (*9*).

10.1126/science.aba3763

SCIENCE AND DECISION-MAKING: COVID-19

Harnessing multiple models for outbreak management

Expert elicitation methods and a structured decision-making framework will help account for risk and uncertainty

By **Katriona Shea[1], Michael C. Runge[2], David Pannell[3], William J. M. Probert[4], Shou-Li Li[5], Michael Tildesley[6], Matthew Ferrari[1]**

The coronavirus disease 2019 (COVID-19) pandemic has triggered efforts by multiple modeling groups to forecast disease trajectory, assess interventions, and improve understanding of the pathogen. Such models can often differ substantially in their projections and recommendations, reflecting different policy assumptions and objectives, as well as scientific, logistical, and other uncertainty about biological and management processes (*1*). Disparate predictions during any outbreak can hinder intervention planning and response by policy-makers (*2, 3*), who may instead choose to rely on single trusted sources of advice, or on consensus where it appears. Thus, valuable insights and information from other models may be overlooked, limiting the opportunity for decision-makers to account for risk and uncertainty and resulting in more lives lost or resources used than necessary. We advocate a more systematic approach, by merging two well-established research fields. The first element involves formal expert elicitation methods applied to multiple models to deliberately generate, retain, and synthesize valuable individual model ideas and share important insights during group discussions, while minimizing various cognitive biases. The second element uses a decision-theoretic framework to capture and account for within- and between-model uncertainty as we evaluate actions in a timely manner to achieve management objectives.

EXPERT ELICITATION AND JUDGMENT

Formal methods for elicitation of information from individuals were developed to harness the collective knowledge of many minds while avoiding the frailties of individual experts (e.g., overconfidence) and the problems that arise in group interactions, such as agreeing with field "leaders" (dominance effects), focusing on suggestions raised early in the process to the detriment of other ideas (starting-point bias, groupthink, anchoring), the dominating effects of "loud voices," and overly rapid adoption of early ideas that might, on more careful consideration, be incorrect (*4, 5*). In these formal methods, idea generation and idea evaluation are deliberately separated, allowing a fuller range of possibilities to be explored and a wide range of uncertainties to be assessed. As one example, in the IDEA protocol for expert elicitation (*6*), once experts are clear about the questions, they individually provide initial best estimates and ranges, receive feedback on how their estimates compare with others, discuss the results, and then provide a final individual estimate. Some protocols, including IDEA, are designed to work remotely—an essential requirement in the present COVID-19 context.

To harness both the creativity of individuals and the insights of groups, variations on the Delphi method (developed by the RAND Corporation in the 1950s and included within the IDEA protocol) and the Nominal Group Technique (*7*) involve both independent and interactive stages in an iterative elicitation process (*8, 9*). The expert judgment literature shows that a failure to manage the elicitation process well can lead to generation of biased information and overconfidence (*4, 5*). Expert judgment approaches have been used for elicitation from individual experts in a wide range of relevant settings, such as development of clinical guidelines (*8*), and in conservation and ecology (*9*).

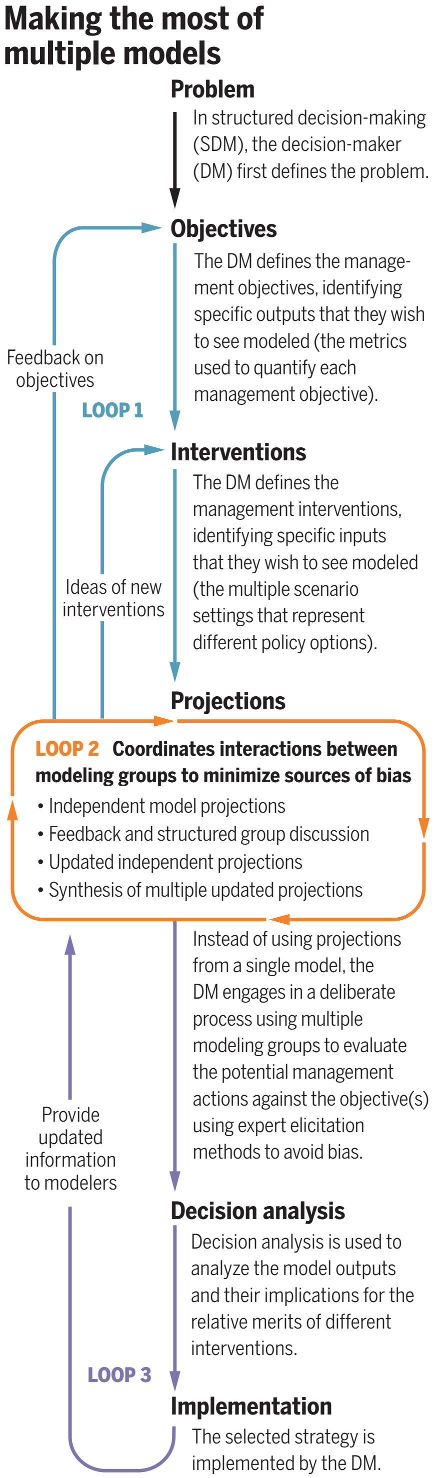
MULTIPLE MODELS

There are a number of existing approaches for dealing with multiple models in weather and climate research (*10*), fisheries (*11*), and disease forecasting (*2, 12*). Such forecasting efforts generally focus on statisti-

[1]Department of Biology, The Pennsylvania State University, University Park, PA, USA. [2]U.S. Geological Survey, Patuxent Wildlife Research Center, Laurel, MD, USA. [3]University of Western Australia, Perth WA 6009, Australia. [4]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, University of Oxford, Oxford, UK. [5]State Key Laboratory of Grassland Agro-ecosystems, Center for Grassland Microbiome, and College of Pastoral, Agriculture Science and Technology, Lanzhou University, Lanzhou, People's Republic of China. [6]Zeeman Institute for Systems Biology and Infectious Disease Epidemiology Research, Mathematics Institute and School of Life Sciences, University of Warwick, Coventry CV47AL, UK. Email: k-shea@psu.edu

cal averaging of model outputs (ensemble averaging, super-ensemble modeling) and are growing in popularity and impact. A few multiple-model protocols also address optimal management (*1, 3, 13, 14*), specifically asking "what should we do to most increase the benefits or reduce the costs?", instead of "what will happen?" In the current COVID-19 outbreak, there has been unprecedented sharing of data and models in curated discussion groups, on preprint servers, and in working groups coordinated by policy agencies such as the World Health Organization (WHO) and the U.S. Centers for Disease Control and Prevention (CDC). These efforts save duplication of research and allow for rapid dissemination of new information, and are essential in a crisis.

Unfortunately, there is a downside to rapid sharing. Poor, as well as good, information may spread rapidly; a failure to prevent this leads to bias. If premature consensus is reached, models predicated on poor assumptions will inevitably propagate bias. Such propagation of initial bias is well documented in the elicitation literature (*4, 5*) and is clearly a potential concern in an epidemiological context.



MODELING-GROUP ELICITATION

We present an elicitation process for multiple modeling groups based on a modified-Delphi approach to expert elicitation (*6*) embedded in a structured decision-making (SDM) process (*15*). The overall process is coordinated by the decision-maker (DM)—for instance, the public health policy agency with authority to act. The adoption of a formal protocol for SDM (see the figure for an overview of the SDM process) allows us to assess the same set of actions in different models to address key objectives in a timely manner, and with an appropriate expression of uncertainty to enable risk-based decision-making. We expand the SDM approach in loop 2 of the figure to explicitly involve multiple modeling groups in a modified Delphi expert elicitation process. In general overview, the modeling groups initially work alone, then together (coordinated by the DM), and finally alone again.

The DM first outlines the process, and the required information, for the modeling groups. This includes description of the management objective and associated metrics (e.g., minimization of total morbidity or mortality) and of the potential management interventions. Note that projecting spatial spread or caseload trajectory under some baseline scenario of no action is still a management intervention ("do nothing"). Thus, efforts to forecast and to assess interventions should be fully integrated and presented consistently. The DM also provides instructions about how within-model uncertainty should be documented (e.g., provide a full probability distribution of outcomes—not just a mean value—for intervention-objective projections, and document sources of variation). Uncertainty may be structural (e.g., should asymptomatic carriers be modeled explicitly?), or parametric with respect to the biology (e.g., what is the expected time between sequential cases in a chain of transmission?), or it may relate to the interventions (e.g., what is the expected impact of social distancing?). The DM must also provide background information and access to current relevant data, information on intervention efficacy (if known), and guidance on data curation.

The DM next must assemble modeling groups for the elicitation process. In the current COVID-19 outbreak, hundreds of research groups around the world are assisting national and international agencies with forecasts and management recommendations. Any type of new or existing model that encapsulates a scientific research group's best understanding of the situation effectively represents a hypothesis about the system, and should be eligible; however, restrictive criteria may be applied with justification in some situations (*3*). To constructively participate, all groups must agree to examine how the interventions of interest meet the DM's stated objectives. However, if some models cannot assess all interventions (e.g., if nonspatial models cannot address spatially explicit interventions), incomplete results may nevertheless be informative (*14*).

During a first round of analysis, individual modeling groups, working independently, project the outcomes for each of the interventions, capturing their own within-model uncertainty. The DM invites a first report of results in a short period of time; this may be facilitated by providing a template format for model outputs. The DM then compiles and compares the results, provides feedback to the individual groups about their projections, and provides results (anonymized, to reduce peer pressure) from all participants to the whole group.

All the modeling groups then participate in a formal, structured discussion to compare results, assess common features, discuss what caused differences, identify valuable information that might not have been available to all groups, generate important insights about the nature of the disease and its dynamics, and identify important insights to share. This structured discussion is an important step, allowing modelers and the DM to assess why models disagree.

After these discussions, individual modeling groups, again working independently, update their models and projections based on the insights from the whole-group discussion. All groups have the opportunity to revise and rerun their models, using their judgments and data, taking account of the all-group discussion to the extent they think it is warranted, but not asking for consensus. The DM then compiles the second round of results, reporting both the central tendencies and the uncertainties within and across models. The Round 2 results are also then

GRAPHIC: N. CARY/SCIENCE

combined into an ensemble projection for each management strategy (*1*, *14*). For example, an ensemble projection for different "reopening" scenarios could show the best estimate of COVID-19 case load across models, for different policies for reinstating social interaction, including full expression of the uncertainty captured by the model set.

The full set of results can then be used to guide policy deliberation. Two particular techniques from the field of decision analysis are relevant here: first, risk analysis (which assesses the probability of poor outcomes for different interventions, and judges the tolerability of such risk) (*15*); and second, value of information analysis (which estimates the value to the DM of resolving one or more uncertainties prior to the implementation of a decision) (*13*). Central to these approaches is the recognition that projections may be wrong; by documenting uncertainty, a DM can evaluate which interventions may be most robust to uncertainty and can allocate research effort to reduce the most consequential uncertainty.

Two rounds of modeling (instead of one) are key to the modeling-group elicitation process. Current sharing of early ideas through preprint servers and curated discussion groups permits communication of independent ideas that might otherwise not be shared; however, broad sharing also runs the risk of losing independent idea generation (*3*). The Round 1 projections, undertaken before the groups start collaborating with each other, are essential to avoid starting-point bias and groupthink, by encouraging independent idea generation and creative thinking. Asking modeling teams to formally review collective results and determine the reasons for differences in model projections highlights key uncertainties and helps modeling groups to detect flaws in their assumptions or modeling approach. Nevertheless, the inclusion of a second, postdiscussion round of modeling need not delay the DM from taking action: The first round of results can be used for initial policy recommendations if time is of the essence.

There are major advantages to embedding the model group elicitation process in a structured decision-making framework (*15*), relative to the ad hoc decision processes that are often employed. First, one iteration of the whole elicitation process can be used by the DM to decide on an initial course of action to meet a clearly stated objective. Second, an important advance with expert elicitation is to elicit a relatively unbiased and full expression of uncertainty. The process provides policy-makers with two crucial pieces of information: a sense of the central tendency of the projections across models, and an understanding of the underlying uncertainty, as captured by the range of projections for different interventions. The results can be analyzed to identify which uncertainties most strongly affect the choice of action. It may be that one intervention is ranked best by all models despite uncertainty, or it may be that the top-ranked intervention is highly sensitive to a particular uncertainty, indicating the need for research on that particular factor. Third, if the outbreak is ongoing, and if models disagree (so that uncertainty is an impediment to choosing a course of action), and if new information on dynamics or management outcomes can be incorporated into the process, an adaptive management (AM) approach, involving management with a plan for learning, is warranted (*13*).

It is also possible for policy-makers to tailor the degree to which the elicitation exercise itself feeds into targeted research or management decisions. Scientists can contribute to the process by providing input on new or better-specified objectives to policy-makers and by providing suggestions for additional or modified interventions for the whole group to assess. For example, in loop 1, if a policy-maker requests advice on the optimal intervention to "control" a disease, modeling groups may demonstrate that a more precise objective statement is needed (e.g., while zero cases equates to zero deaths, a small outbreak of COVID-19 in a nursing home may lead to more deaths than a large outbreak in a university setting; reducing caseload and mortality are not equivalent objectives). Similarly, in the 2014 Ebola outbreak, model forecasts ranged wildly, yet a focus on minimizing the number of cases brought surprising consensus on the best approach to intervention (*1*). In loop 1, modeling work by one group may also suggest potentially fruitful interventions that all groups could evaluate (e.g., earlier intervention triggers than proposed by the DM, or previously unconsidered interventions).

The proposed process encourages a healthy conversation between scientists and decision-makers and engenders a stronger integration of science and policy, enabling policy agencies to more effectively achieve their management goals. Furthermore, it helps the DM to embrace uncertainty, rather than hastening to a premature consensus that could derail or deflect management efforts (*9*).

Adoption of elicitation methods for multiple modeling groups should be relatively straightforward; all structures to support and facilitate such a process are already in place [e.g., the "forecasting challenges" exemplified by (*4*, *12*), and channels for communication between DMs and modeling groups]. Thus, additional costs of this approach relative to traditional approaches should be minimal. A strength of this approach is that individual modeling groups may preserve their autonomy and unilaterally conduct additional analyses and publish their work independently, as before, yet gain additional benefits from being involved formally in the group exercise, both in terms of access to information and feedback, and in terms of contributing to the greater good. As a result, high participation of modeling groups should be achievable. We suggest that this strategy will prompt better outbreak management outcomes for similar effort by deliberately leveraging more value from the modeling process. As such, the risks and stresses inherent in implementing the approach are also minimized and should not be deterrents.

Leveraging the contributions of multiple modeling groups is likely to pay dividends in preventing morbidity and death. In short, managing any disease, including a disease outbreak, requires that policy objectives be well defined, the decision-making process be carefully structured, and multiple contributing modeling groups be handled in a manner that shares knowledge while avoiding bias. Intertwining expert judgment methods with multiple model comparisons, in a planned and deliberate manner, will thus increase the benefits derived from multiple groups' efforts to model outbreaks. ■

REFERENCES AND NOTES

1. S. L. Li *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **114**, 5659 (2017).
2. C. J. Carlson, E. Dougherty, M. Boots, W. Getz, S. J. Ryan, *Sci. Rep.* **8**, 4921 (2018).
3. S. den Boon *et al.*, *BMC Med.* **17**, 163 (2019).
4. P. E. Tetlock, D. Gardner, Superforecasting: The Art and Science of Prediction (Random House, 2016).
5. M. A. Burgman, Trusting Judgements: How to Get the Best out of Experts (Cambridge Univ. Press, 2015).
6. V. Hemming, M. A. Burgman, A. M. Hanea, M. F. McBride, B. C. Wintle, *Methods Ecol. Evol.* **9**, 169 (2018).
7. A. Delbecq, A. Van de Ven, *J. Appl. Behav. Sci.* **7**, 466 (1971).
8. M. K. Murphy *et al.*, *Health Technol. Assess.* **2**, 1 (1998).
9. M. C. Runge, S. J. Converse, J. E. Lyons, *Biol. Conserv.* **144**, 1214 (2011).
10. T. N. Krishnamurti *et al.*, *Science* **285**, 1548 (1999).
11. S. C. Anderson *et al.*, *Fish Fish.* **18**, 732 (2017).
12. C. Viboud *et al.*, *Epidemics* **22**, 13 (2018).
13. K. Shea, M. J. Tildesley, M. C. Runge, C. J. Fonnesbeck, M. J. Ferrari, *PLOS Biol.* **12**, e1001970 (2014).
14. W. J. M. Probert *et al.*, *Epidemics* **15**, 10 (2016).
15. R. Gregory *et al.*, Structured Decision Making: A Practical Guide for Environmental Management Choices (Wiley-Blackwell, 2012).

ACKNOWLEDGMENTS

M.F., M.T., and K.S. are supported by an award from the NSF Ecology and Evolution of Infectious Diseases (EEID) program. K.S. is supported by an NSF COVID-19 Rapid Response Research (RAPID) award. K.S. is supported by the Huck Institutes for the Life Sciences at The Pennsylvania State University. W.J.M.P. is funded by the Li Ka Shing Foundation. The authors thank E. Howerton and C. Viboud. Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

10.1126/science.abb9934

Harnessing multiple models for outbreak management

Katriona Shea, Michael C. Runge, David Pannell, William J. M. Probert, Shou-Li Li, Michael Tildesley and Matthew Ferrari

Science **368** (6491), 577-579.
DOI: 10.1126/science.abb9934

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.