In January 2020, there was a great deal of uncertainty around the emerging outbreak that would eventually become the COVID-19 pandemic. That month, a COVID-19 dashboard hosted by the Johns Hopkins University went online, tracking rising case numbers globally. Ever since, the dashboard has remained an important source of information for the media, policymakers, and the general public. Lauren M. Gardner, a professor of civil and systems engineering at the Johns Hopkins Whiting School of Engineering, is one of the creators of the dashboard. In recognition of her efforts in creating the dashboard and establishing a standard for disseminating public health data in real time, Gardner has won the 2022 Lasker-Bloomberg Public Service Award. Gardner spoke with PNAS about her award-winning accomplishment that has served as a crucial public health tool.

PNAS: What was your initial motivation for creating the COVID-19 dashboard?

Gardner: We started the dashboard for a couple different reasons. Firstly, I have a background in infectious disease modeling, specifically in emerging infectious diseases, which are novel pathogens that appear and have a lot of uncertainty around them. Because I've modeled these kinds of scenarios, I knew the importance of timely high-quality data for assessing the risk of disease spread at the earliest stages of an outbreak. Additionally, my PhD student Ensheng Dong was closely watching the newly reported cases of what were later declared COVID-19 coming out of his home country, China, and following that data on a Chinese website. My expertise and his personal vested interest in this particular situation were the reasons that we decided together to commit to continuing to collect the data that was available at the time and visualize and share it online. We committed to this plan in a coffee meeting on January 21, 2020, built the prototype of the dashboard that night, and I shared it on Twitter the next day with a link to the data that was populating the visualization (1).

PNAS: What was the biggest challenge in maintaining the dashboard?

Gardner: The biggest challenge was conducting such large-scale data collection in an environment that lacked standards with regard to the data being reported. When we started collecting reported COVID-19 data, there weren't web pages dedicated to COVID reporting hosted by local health authorities. The very initial data had to be manually collected from all sorts of different sources, including Twitter and Facebook posts and local news outlets. Over time, the set of sources reporting on COVID-19 grew, as there were dedicated web pages for COVID-19 that were stood up by public health authorities. While this change was welcome, it posed a real challenge on the data collection side, because each source varied in what and how they reported information. Websites used different mechanisms for publishing data. Some examples include PDF reports, ESRI dashboards, Power BI, and even infographics.

Further complicating this, the actual variables that were being provided—[for example], confirmed cases, probable cases, confirmed deaths, probable deaths, recoveries—varied in their definition, depending on the time and location of reporting. For example, sometimes probable cases would be bundled in with confirmed cases, sometimes they would be presented separately, and sometimes they wouldn't be included at all. To address these challenges, we had to carefully study each source to understand the reporting and design custom scraping technology to be able to collect and integrate the data accurately. We also regularly have to audit our set of sources and adapt the scraping technologies because the sources themselves are consistently changing how they report.

PNAS: Based on what you are seeing on the dashboard, what is your outlook on the current state of the pandemic?

Gardner: The first word that comes to my mind when you ask me about the future of COVID is “uncertainty.” I think that the one thing we know about COVID is that we can't predict it, and the one thing we can predict is that it's going to continue to surprise us. I actually use our data on a daily basis.
basis within mathematical models we build, to try and better understand what drives COVID outbreaks, when and where they happen, and to try to forecast COVID outbreak dynamics into the future. So, I feel like I can say with authority that it's a very challenging virus to predict, and there is still a lot for us to learn.

My personal opinion about why COVID is so hard to predict is because the outbreak dynamics are heavily dependent on human behavior, which we all know is very complicated, highly variable, and hard to predict. Behavior is driven by many complex and interacting factors, including policy, policy compliance, politics, risk perception, access to information, misinformation propagation, as well as other more traditional factors that affect viral dynamics. That being said, I would say I'm cautiously optimistic that the worst days of COVID, in terms of disease burden, are behind us. Moving forward, I expect that COVID is here to stay, and we will likely see many new versions of it. But, due to the hard work of so many scientists, we have many tools at our disposal to manage COVID risks, vaccines, in particular. We just have to be willing to use them.

**PNAS:** You published an article (2) in PNAS on how the duration of travel affects the spatial dynamics of infectious diseases. How does considering the duration of travel in addition to the number of trips affect outbreak modeling?

**Gardner:** Travel patterns, specifically the volume of human movements between locations, are well established as a critical factor in the spread of infectious disease. However, the amount of time people spend at a travel destination also plays an important role in the probability of onward disease transmission. Thus, expanding the current infectious disease modeling framework to account for trip duration is critical for improving our understanding of disease risk.

**PNAS:** More recently, you were part of a large consortium of nearly 300 authors who published an article (3) in PNAS on the accuracy of COVID-19 forecasting. What was your experience working with such a large team and what was your contribution to the article?

**Gardner:** This paper was led by the team that manages the COVID19 ForecastHub (https://covid19forecasthub.org/). My team, in the Center for Systems Science and Engineering, contributed to this work in two ways. We serve as one of the teams contributing weekly COVID-19 forecasts, which were evaluated independently in this analysis, and also contributed to the performance of the ensemble model. Perhaps more importantly, the John Hopkins University COVID-19 dashboard data served as the ground truth for the evaluation analysis presented in this paper. My experience collaborating with this large team of researchers is unique, in that I serve in the role of both a data provider and a data user. This experience has been extremely beneficial, as the regular communication with the modeling community provides an opportunity to translate data-related issues directly to the modeling teams and thus improve the performance of COVID-19 forecasting models and, simultaneously, we use feedback from the modeling teams to improve the quality of the dashboard data product.

1. L. Gardner, (@TexasDownUnder), “We are tracking the 2019-nCoV spread in real-time. Cases and locations can be viewed here; data available for download. #nCoV2019 @JHUSystems. https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6, (2020). https://twitter.com/TexasDownUnder/status/1220014483516592129?cxt=HHwWgsC07a6nu4hAAAA. Accessed 22 January 2020.
2. J. R. Giles et al., The duration of travel impacts the spatial dynamics of infectious diseases. Proc. Natl. Acad. Sci. U.S.A. 117, 22572-22579 (2020).
3. E. Y. Cramer et al., Evaluation of individual and ensemble probabilistic forecasts of COVID-19 mortality in the United States. Proc. Natl. Acad. Sci. U.S.A. 119, e2113561119 (2022).