Towards recovery: Scientists with better ratings of their institution's response to the COVID-19 pandemic have more optimistic forecasts about their future research

Kyle R. Myers*[1,2], Karim R. Lakhani[1,2,3], Dashun Wang[4,5,6,7]

*Corresponding author: kmyers@hbs.edu
[1]Harvard Business School, Harvard University, Boston, MA, USA
[2]Laboratory for Innovation Science at Harvard, Harvard University, Boston, MA, USA
[3]Institute for Quantitative Social Science, Harvard University, Boston, MA, USA
[4]Center for Science of Science and Innovation, Northwestern University, Evanston, IL, USA
[5]Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL, USA
[6]McCormick School of Engineering, Northwestern University, Evanston, IL, USA
[7]Kellogg School of Management, Northwestern University, Evanston, IL, USA

Using a survey with randomized elements and truth-telling incentives, we document scientists' beliefs about current and future disruptions to their research and how these beliefs depend on the eventual length of the COVID-19 pandemic. Overall, scientists with more favorable ratings of their institution's response to the pandemic have more optimistic forecasts, even when controlling for their current level of disruptions. This relationship holds even amongst groups with the most pessimistic forecasts, who should be prioritized as science policy responses continue to develop.

Many of the most pressing questions in science policy today concern the support and recovery from the dramatic disruptions caused by the COVID-19 pandemic. Research institutions around the world have implemented, and continue to develop, policies and initiatives in hope of alleviating the setbacks caused by the pandemic. An understanding how well institutions' efforts may be able to mitigate the pandemic's effects is not only immediately useful to ongoing policy debates, but it may also prove important for the long-term vitality of the scientific enterprise.

Initial work has used data on article pre-prints, grant applications, and time-use surveys to investigate the immediate effects of this pandemic, finding very heterogeneous experiences across different groups of scientists.[i,ii,iii] But it remains unclear whether these early data points are good predictors of long-term outcomes, or whether institutions have the capabilities to mitigate these disruptions given the sweeping scale of the pandemic. For example, research universities and colleges have granted tenure clock extensions, funding agencies have loosened regulations surrounding no-cost extensions on grants, and national academies have convened panels to collect and disseminate new knowledge.[iv,v] But are these efforts helping? And if so, whom are they helping?

Surveying scientists in real-time

Here, we present some of the first systematic evidence that relates research institutions' pandemic responses to scientists' forecasts of their research productivity in the coming years. To provide timely data on how the pandemic has and may continue to affect scientists, we conducted a survey of active academic researchers from June to August 2020. In this survey, we asked scientists about recent or ongoing disruptions to their work (e.g., changes in work time; a permanent disruption to an ongoing research project; a lost promotion), as well as their forecasts of how much their research might ultimately be disrupted in the future. Using five questions (see Figure 1 Panel A), we asked scientists to forecast changes in their research funding, publication output and impact, as well as the amount of time and funding it would take to return to a "normal" pace of work once the pandemic was over.

A major challenge to forecasting the long-term effects of the pandemic is that it will clearly depend on how long the pandemic lasts, but the ultimate length of the pandemic is very uncertain. Thus, if a scientist reported a relatively optimistic forecast, it might be because they were less affected by the pandemic compared to others, or

it might be that they expected the pandemic to end much sooner than others. To tackle this challenge, we incorporated a randomized design into our survey. When posing the five forecast questions, we asked the respondent to use a specific assumption about the eventual duration of the pandemic. We randomly presented each respondent with just one of these assumptions about the pandemic's horizon – between one and eight months into the future (e.g., "Assuming the pandemic ends in eight months…"). This random variation allows us to attribute differences in scientists' responses to how much the pandemic is affecting their work.

Using a simple standardized composite of their responses to the five forecast questions, we created a Research Disruption Index (RDI) where larger values indicated the scientist forecasted relatively more disruptions compared to others. The RDI is based on standardizations of responses, so that a difference of one unit indicates that the respondent forecasts disruptions that are one standard deviation larger than the sample average. For reference, a one standard deviation change in the RDI is roughly equivalent to either a 20% decline in research funding or a 16-22% decline in publication output or impact in the coming two years, or needing two months longer or 35% more funding to resume a normal pace of work once the pandemic is over.

In addition to the current and forecasted research disruption questions, we asked scientists to rate the overall quality of their own institution's response to the pandemic, and we also included a series of professional and socio-demographic questions. The Supplementary Information contains further details on the survey design, sampling, response rates, and how we construct the RDI.

Here, we report the results from the responses from 3,444 respondents who self-identified as either research faculty or principal investigators. We compared how scientists' forecasts differed along five dimensions: (1) own institution's pandemic response rating; (2) U.S. versus non-U.S. (mostly Canada and Europe); (3) broad field of research; (4) age of dependents; and (5) gender identity. For each dimension, we used a simple linear regression model to estimate how scientists' disruption forecasts varied with the (randomly assigned) pandemic horizon and allowed this relationship to vary by group.

An important limitation of this approach is that when soliciting forecasts, some of the variation may reflect behavioral biases, an unwillingness to use the assumption we provided, or other incentives that lead respondents to not put effort into their response or to not report honestly. To address this concern, we encouraged respondents to report honestly by using incentives that stem from the Bayesian Truth Serum mechanism.[vi] In short, each respondent is rewarded in a way that (theoretically) induces all respondents to report honestly; see the Supplementary Information for more.

Are institutions helping?

Grouping scientists based on how they rate their institution's response to the pandemic reveals that the overall level of disruption is clearly lower for scientists who report their institution's response to the pandemic as better (see Figure 1 Panel B). However, our approach does not necessarily imply that this relationship is due to institutions' policies, per se. For instance, scientists less disrupted by the pandemic likely demanded less support from their institutions and may thus report both a high institutional rating and a low disruption (even if their institution did not implement truly helpful policies).

To investigate this further, the Supplementary Information includes results from regression analyses where we find that, even after controlling for the degree to which scientists have already been disrupted (per changes in their work time and their responses to questions about effects they have already experienced) as well as many factors that might also generate a spurious correlation (e.g., field, age, gender, country, dependents, funding levels, institutional ranking), the relationship between disruptions and institutional ratings remains very similar. In other words, even when comparing scientists who have similar socio-demographics, who have access to similar amounts of research resources, and who have already experienced similar changes in their work due to the pandemic, we still find that researchers with better ratings of their institutions have more optimistic forecasts about the years to

come. This suggests that, despite the grand scale of the pandemic, there is a role for institutional-level policies to help researchers cope with their research disruptions.

Notably, we do not find any important differences between scientists inside or outside of the U.S. (see Figure 1 Panel C). This international similarity is consistent with a hypothesis that there is much more variation in institutional responses within countries than across. This similarity is also interesting when compared to other metrics of the pandemic's impact, such as case rate or mortality rate, which show marked differences between the U.S. and other countries at the time of this survey.[vii] This finding is consistent with evidence that in the U.S. the burden of the pandemic has fallen largely on socio-economically disadvantaged individuals and communities,[viii] which likely does not include most of the academic faculty and principal investigators in our sample.[ix]

Stabilizing versus deteriorating effects

When comparing broad scientific disciplines, we find that biologists and chemists clearly are among the most sensitive to the duration of the pandemic, forecasting significantly more research disruptions with each additional month of the pandemic (see Figure 1 Panel D). Physicists, engineers, and environmental scientists are the only other set of disciplines that appear to forecast significant deterioration as the pandemic continues. Statistically speaking, we can reject null hypotheses that these "bench" sciences are insensitive to the pandemic's horizon and that this sensitivity is equal to that amongst the social and health sciences. This concentration of sensitivity amongst the bench sciences does not imply that only biologists and chemists should receive support to help cope with their disruptions. Rather, it suggests that researchers whose work resembles that of an average bench scientist today (e.g., work is done in a laboratory by large teams using expensive equipment and specialized or perishable inputs) are most likely to fare worse as the pandemic continues.

Grouping scientists based on the presence and age of their dependents reveals that those reporting young dependents are more sensitive to the pandemic overall, although the rate of deterioration across these groupings are all similar (see Figure 1 Panel C). When comparing male and female scientists we estimate that, overall, female scientists forecast larger disruptions in their research due to the pandemic (i.e., the difference in the y-intercepts for male and female scientists in Figure 1 Panel D is 0.26 standard deviations). This is consistent with prior work.[iii,iv]

Is there hope for those who expect the worst?

Does the relationship between scientists' ratings of their institutions and their forecasts persist within the groups of scientists with the most pessimistic forecasts? Yes. In the Supplementary Information, we show that the association between institutional ratings and forecasts – where better ratings predict more optimistic forecasts – is persistent within all groupings of scientists, whether they appear to have stabilized or forecast some further deterioration. For instance, this relationship is very similar for both those in the bench sciences and those in the health and social sciences. We find this same result whether we compare those with or without young dependents, those in or outside of the U.S., as well as male or female scientists. It appears that individuals experiencing larger disruptions may just as easily be helped by their institutions as those who have been more immune to disruptions.

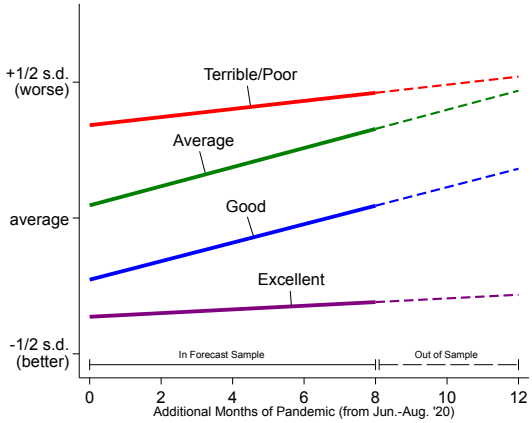
While still far from conclusive, these patterns suggest that, despite the dramatic scale of the pandemic and its heterogeneous impacts, research institutions can play an important in mitigating the effects of the pandemic and, eventually, rebuilding the scientific enterprise. We have found that there are some scientists who have stabilized and there are some that, without further support, may continue to deteriorate as the pandemic continues. Those in the latter group should be prioritized as responses continue to develop. These findings will also eventually prove useful to those who will come to study the longer-term effects of the pandemic on the scientific enterprise.

Figure 1. Research disruption forecasts over different COVID-19 pandemic horizons

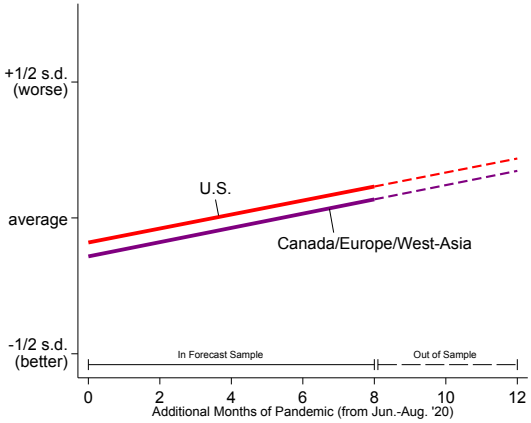
A. Research disruption questions

Preface: Assuming the pandemic ends in __ months…
Funding: …how much more/less research funding do you expect in 2020-2021 compared to 2018-2019?
Publication quantity: …how do you expect your pub. quantity in 2020-2021 to compare to 2018-2019?
Publication impact: …how do you expect your pub. impact in 2020-2021 to compare to 2018-2019?
Recovery time: …how much time do you expect it to take to resume your normal pace of work?
Recovery funding: …how much funding do you expect it will take to resume your normal pace of work?

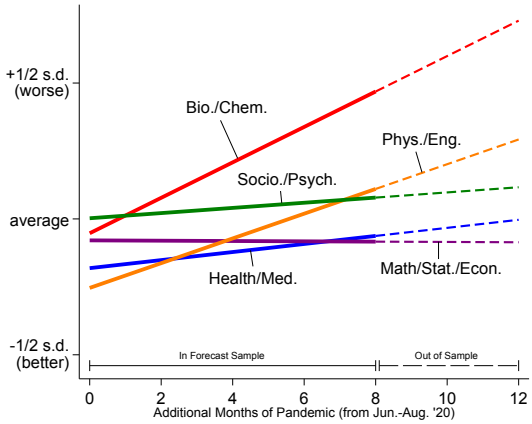
B. By Institutional rating



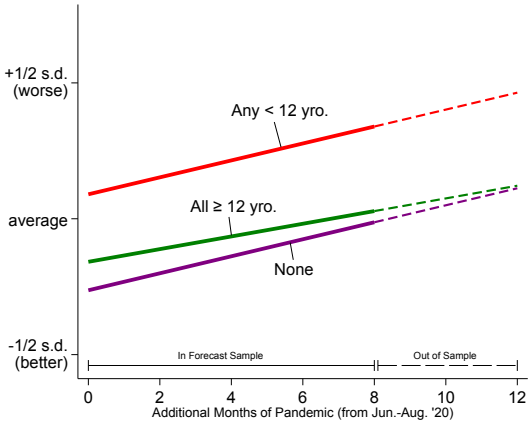
C. By region



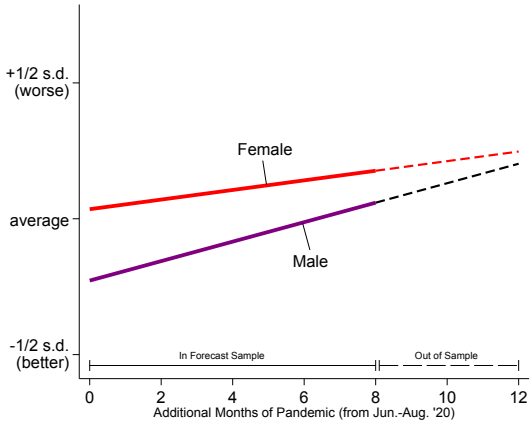
D. By field



E. By dependent status



F. By gender identity



Note: Panel A presents abbreviated versions of the questions used to construct the Research Disruption Index (RDI) along with the preface where we randomize the to-be-assumed pandemic horizon. Panels B-F plot the predicted RDI (y-axis; larger indicates more disruption) as a function of how long scientists are randomly asked to assume the duration of the pandemic will be (x-axis). For reference, a 1 s.d. increase in the RDI is roughly equivalent to either a 20% decline in research funding or a 22% decline in publication output in the coming two years, or needing two months longer or 35% more funding to resume a normal pace of work once the pandemic is over. Each panel is based on a separate OLS regression that controls for the date of the survey and includes group-specific intercepts and linear horizon coefficients. Only horizons of 1, 2, 4, 6, and 8 months or less were tested in the survey; the linear projections after 8 months are out of sample.

References

[i] Vincent-Lamarre, P., Sugimoto, C. R., & Larivière, V. (2020, May 19). The decline of women's research production during the coronavirus pandemic. Nature Index. https://www.natureindex.com/news-blog/decline-women-scientist-research-publishing-production-coronavirus-pandemic

[ii] Myers, K. R., Tham, W. Y., Yin, Y., Cohodes, N., Thursby, J. G., Thursby, M. C., Schiffer, P., Walsh, J. T., Lakhani, K. R., & Wang, D. (2020). Unequal effects of the COVID-19 pandemic on scientists. Nature Human Behaviour, 4(9), 880–883. https://doi.org/10.1038/s41562-020-0921-y

[iii] Lauer, M. (2020, July 28). An Early Look at Applications Submitted During the Pandemic. National Institutes of Health. https://nexus.od.nih.gov/all/2020/07/28/an-early-look-at-applications-submitted-during-the-pandemic/

[iv] Council On Government Relations. (2020). Institutional and Agency Responses to COVID-19 and Additional Resources. https://www.cogr.edu/institutional-and-agency-responses-covid-19-and-additional-resources

[v] The National Academies of Science, Engineering, and Medicine. (2020). COVID-19 Responses & Resources. https://www.nationalacademies.org/topics/covid-19-resources

[vi] Cvitanić, J., Prelec, D., Riley, B., & Tereick, B. (2019). Honesty via Choice-Matching. American Economic Review: Insights, 1(2), 179–192. https://doi.org/10.1257/aeri.20180227

[vii] Johns Hopkins University. (2020). Coronavirus Resource Center. https://coronavirus.jhu.edu.

[viii] Tackle coronavirus in vulnerable communities. (2020). Nature, 581(7808), 239–240. https://doi.org/10.1038/d41586-020-01440-3

[ix] American Association of University Professors. (2017). Visualizing change: The annual report on the economic status of the profession, 2016-2017. https://www.aaup.org/file/FCS_2016-17.pdf

Supplementary Information for Towards recovery: Scientists with better ratings of their institution's response to the COVID-19 pandemic have more optimistic forecasts about their future research

Kyle R. Myers*[1,2], Karim R. Lakhani[1,2,3], Dashun Wang[4,5,6,7]

*Corresponding author: kmyers@hbs.edu
[1]Harvard Business School, Harvard University, Boston, MA, USA
[2]Laboratory for Innovation Science at Harvard, Harvard University, Boston, MA, USA
[3]Institute for Quantitative Social Science, Harvard University, Boston, MA, USA
[4]Center for Science of Science and Innovation, Northwestern University, Evanston, IL, USA
[5]Northwestern Institute on Complex Systems, Northwestern University, Evanston, IL, USA
[6]McCormick School of Engineering, Northwestern University, Evanston, IL, USA
[7]Kellogg School of Management, Northwestern University, Evanston, IL, USA

Human research participants

The study protocol has been approved by the Institutional Review Board (IRB) from Harvard University and Northwestern University. Informed consent was obtained from all participants.

Acknowledgements

We are extremely thankful to Nina Cohodes, Alexandra Kesick, Wei Yang Tham, Jerry Thursby, Marie Thursby, Ruihan Wang, and Yian Yin for assistance with the creation, distribution, and management of the survey and responses. Peter Schiffer and Jay Walsh shared helpful thoughts on the research disruption questions. Drazen Prelec provided important guidance on the adaptation of the Honesty-Via-Choice-Matching incentive mechanism to this survey. This work was financially supported by the Air Force Office of Scientific Research under award number FA9550-19-1-0354, National Science Foundation SBE 1829344, and the Alfred P. Sloan Foundation G-2019-12485 and G-2020-13873, and the Harvard Busines School.

Author contributions

KRM, KRL, and DW conceived the project and designed the survey; KRM collected data, performed empirical analyses, and wrote the manuscript; KRM, KRL, and DW edited the manuscript.

Competing interest declaration

The authors declare no competing interests.

Data availability

Because of the sensitive nature of some of the variables collected, the IRB-approved protocol does not permit individual-level data to be made unrestricted and publicly available. Researchers interested in obtaining restricted, anonymized versions of this individual-level data should contact the authors to inquire about obtaining an IRB-approved institutional data sharing agreement.

Code availability

Code necessary to reproduce all plots and statistical analyses will be made freely available.

Table of Contents

S1 Survey sampling and recruitment

S1.1 Web of Science corresponding authors

To compile a large, plausibly random list of active scientists, we leverage the Web of Science (WoS) publication database. The WoS database is useful for two reasons: (1) it is one of the most authoritative citation corpuses available[i]; (2) among other large-scale publication datasets, WoS is the only one, to our knowledge, with systematic coverage of corresponding author email addresses.

By mid-April 2020, the cumulative number of deaths due to COVID-19 had reached approximately 115,000 with nearly 1,800 deaths per day in the U.S. and 3,000 deaths per day in Europe.[ii] Throughout the U.S. and Europe, schools and workplaces were typically required to be closed and restrictions on gatherings of more than 10 people were in place in most countries.[iii] For scientists, not only did this drastically change their daily lives, it severely limited the possibilities of using traditional workspaces.[iv]

Therefore, in this paper we are primarily interested in active scientists residing in the U.S. and Europe. We start from 21 million WoS papers published in the last decade (2010-2019). In an attempt to focus on scientists likely to still be active and in a more stable research position, we link the data to journal impact factor information (WoS Journal Citation Reports), and exclude papers published in journals in the bottom 25% of the impact factor distribution for its WoS-designated category. We use the journal impact factor calculated for the year of publication, and for papers published in 2019, we use the latest version (2018). We then extract all author email addresses associated with papers. For each email address in this list, we consider it as a potential participant if: (1) it is associated with at least two papers in the ten-year period, and (2) the most recent country of residence, defined by the first affiliation of the most recent paper, is in the U.S. or Europe.

We have approximately 2.5 million unique email addresses after filtering, with about 521,000 in the U.S. and 938,000 in Europe. We then randomly shuffled the two lists separately and sampled roughly 280,000 email addresses from the U.S. and 200,000 from Europe for a total of 480,000. We oversampled the U.S. as a part of a broader outreach strategy underlying this and other research projects.

S1.2 Participant recruitment

We recruited participants by sending them email invitations with the following text:

> *We need your help to shed light on how the coronavirus pandemic is affecting scientists like you. Please take a brief moment to complete this short five-minute survey as part of a research study. Your responses will help scientists and policymakers understand and respond to this rapidly evolving situation.*
>
> *Follow this link to the Survey:*
> *[link]*
> *Or, copy and paste the URL below into your internet browser:*
> *[link]*
>
> *Upon completion, you can choose charities to receive a donation on your behalf, and you may have the chance of winning a $100 gift card. Please feel free to forward this survey to any other scientists you know (e.g., professors, post-doctoral researchers, graduate*

students). We need everyone's input to fully understand the breadth of how science is currently changing.

Thank you for your time,

Kyle Myers, Ph.D. & Karim Lakhani, Ph.D.
Laboratory for Innovation Science at Harvard

Dashun Wang, Ph.D.
Northwestern Kellogg Center for Science of Science & Innovation (CSSI)

S2 Survey instrument

S2.1 Survey questions

The survey includes questions on professional information (position type, institution type, fields of study, type of research, tenure status), demographic information (age, gender, cohabitation, dependents, citizenship), time allocation (time spent on different activities before and after the pandemic outbreak) and predicted changes in future publication and funding. Below are the excerpted questions underlying the variables used in our analyses. We did not require respondents to answer any of the demographic questions regarding gender, age, dependents, cohabitation, or citizenship.

Q. Which of the following best describes your current position?
Faculty or Principal Investigator | Post-doctoral researcher | Graduate student in a doctoral program | Retired scientist no longer engaged in research | Other

Q. Which of the following best describes your field of study?
General: [list of 20 fields]

Q. Gender:
Male | Female | Other | Prefer not to say

Q. Number of dependents of any age you care for:
0 | 1 | 2 | 3 or more | Prefer not to say

Q. In what age group(s) are your dependents? Note. You may select multiple
0-2 years old | 3-5 years old | 6-11 years old | 12-18 years old | 18-65 years old | Over 65 years old

Q. How would you rate your institution's response to the pandemic?
Terrible | Poor | Average | Good | Excellent

Q. Previously in 2018 and 2019, approximately how much research funding did you oversee or manage per year? Note: Ignore overhead or indirect costs; focusing only on funds used directly for research regardless of source (e.g., home institution, federal grant). Approximately:

<10,000 $/year | 10,000-20,000 $/year | 20,000-50,000 $/year | 50,000-100,000 $/year | 100,000-200,000 $/year | 200,000-500,000 $/year | 500,000-1,000,000 $/year | 1,000,000-2,000,000 $/year | >2,000,000 $/year

For the following, please consider your "research publications" as all of your publications that focus on a research question. (e.g., journal articles, conference proceedings, patents, books. Ignore commentary, editorials, etc.)

Q. Assuming this pandemic lasts for another [1, 2, 3, 4, 6, 8] *months (until* [May, June, July, August, October, December]*), how do you think the quantity and impact of your research publications will change in 2020 and 2021 compared to 2018 and 2019? (i.e., what will be the percent change?)*
Quantity (i.e., number): [slider scale from -100% to +50% in 10% increments]
Impact (i.e., quality influence): [slider scale from -100% to +50% in 10% increments]

Q. Assuming this pandemic lasts for another [1, 2, 3, 4, 6, 8] *months (until* [May, June, July, August, October, December]*), how much time do you expect it to take to resume your normal pace of work?*
Approximately:
No time needed | About a week | 2 to 3 weeks | 3 to 5 weeks | 1 to 2 months | 3 to 5 months | 6 months or more

Q. Assuming this pandemic lasts for another [1, 2, 3, 4, 6, 8] *months (until* [May, June, July, August, October, December]*), how much funding do you expect it will take to resume your normal pace of work?*
Approximately:
No funds needed | Less than $1,000 | $1,000-$5,000 | $5,000-$10,000 | $10,000-$50,000 | $50,000-$100,000 | $100,000-$500,000 | More than $500,000

Q. Besides what you have just reported, have any of the following happened to you and your work because of the coronavirus pandemic, social distancing, or related events?
I lost important opportunities to disseminate research | A research project has been disrupted to the point that it will likely never be completed as intended | My income will be lower than expected | I am concerned about the security of my position

To fix quantities, scientists reporting values above the highest option were set along the following rules: > $2M became $3M, >6 mo. became 9 mo., and > $0.5M became $0.75M. These values are approximately what the next "largest" choice would have been if we had continued to use the same scale in response options and included an additional (bounded) choice.

While we allowed for twenty broad fields for respondents to choose from, the analyses that compare fields are based on aggregating those twenty fields into five broader categories (that contain roughly equivalent response numbers): Biology or Chemistry or Related, Health or Medicine or Related, Mathematics or Statistics or Economics or Related, Physics or Engineering or Environmental Sciences or Related, Sociology or Psychology or Other Social Sciences.

S2.2 Truth-telling incentive

A notorious challenge of collecting data via self-reported surveys is assessing the validity of respondents' responses. While secondary data sources can be used to validate responses in some settings, our solicitation of forecasts obviously prevents any ex-ante assessment of validity. However, a number of mechanisms have been developed that encourage respondents to report truthfully (or honestly, or to exert effort in their responses) by using incentive schemes where truthful reporting by all respondents is the equilibrium outcome that maximizes each respondent's payoff (e.g., probability they receive a prize). We choose to implement the Honesty Via Choice Matching (HVCM) mechanism, which stems from the "Bayesian Truth Serum" line of work.[v]

This was communicated to respondents with the following text in the consent script:

> *Compensation - This survey rewards your accuracy. After some questions, we ask you to predict the responses of scientists similar to you. If your predictions score in the top 10% of accuracy, you will be entered into a lottery to win a $100 gift card.*

And it was also stated clearly in the introduction page of the survey (immediately succeeding the consent):

> *This survey rewards your accuracy. After some questions, we ask you to predict the responses of other scientists—specifically, other scientists who are in your field and are about your age. Your input will have a greater impact the more accurate it is. If your predictions score in the top 10% of accuracy, you will be entered into a lottery to win a $100 gift card.*

Like many of these sorts of mechanisms, HVCM requires respondents to answer an auxiliary question after a focal question that, essentially, asks the respondent to predict how all other respondents will respond. The payment rule associated with HVCM (and many other mechanisms) incorporates respondents' accuracy in these predictions, rewarding those who are better able to predict others' responses. In the HVCM mechanism, respondents who answer the focal question in the same way actually influence each other's payoffs, which theoretically leads to (or, practically, at least nudges respondents closer to) an equilibrium where all respondents report truthfully.

We use this auxiliary question on all five of the research disruption questions. Based on pilot tests of this mechanism, our use of the auxiliary predication question asks respondents to predict where they lie within the distribution only of individuals who are in the same broad age group and the same broad field of study. We make one slight modification of the standard HVCM mechanism to accommodate the fact that our five research disruption questions are all continuous in nature. Instead of asking respondents to predict the share of (other) respondents that will choose all of the possible answers to the focal question (which is large because we have discretized a continuous variable in each case), we only ask the respondent to predict where they lie in the distribution of all responses (i.e., "what share of your peers will answer with a number lower than the number you chose?"). Under the assumption that respondents with larger valued responses to the focal question will systematically predict that there is also a larger share of other respondents below them in the distribution, this approach preserves the one-to-one mapping of "types" in the sample and their choices in their response.

The HVCM payment rule is somewhat complicated to the point that the resulting scores are not very intuitive so below we simply report the "prediction error" from the auxiliary questions. This error is simply

a scientist's prediction about what percentile their response is at in the distribution of responses from their peers minus their actual location. As shown below, scientists tend to slightly overestimate their location.

For payments based on respondents ultimate HVCM score, we offered $100 Amazon gift cards to a randomly chosen subset of 75 respondents who scored in the top 10% of the full survey sample (which was communicated to participants at the beginning of the survey).

S2.3 Participation incentive

To encourage participation in the survey, the final question of the survey allowed the respondents to choose amongst six approved charities (all with active COVID-19 initiatives) to receive a donation on their behalf. We distributed $15,000 amongst the charities in accordance with the share of votes each charity received.

S3 The sample

S3.1 Distributions and responses

The recruitment for the survey occurred in two broad waves. The first occurred via a single mass e-mailing in mid-April 2020 to all 480,000 addresses. The second occurred via a series of weekly reminder e-mailings to a subset of those 480,000 addresses over ten weeks from early June to mid-August 2020. For reasons described below, all of the responses to the first wave were used as the basis for an earlier study,[vi] and all of the responses to the second wave form the basis for this study.

In the first mailing, out of a total of 480,000 emails sent, approximately 64,000 emails were directly bounced either due to incorrect spelling in the WoS data or the termination of the email account. Overall 9,968 individuals entered the survey and 8,447 continued past the consent stage. Of those that did not, 412 were not an active scientist, post-doc, or graduate student and thus not within our population of interest, 81 did not consent, and 1,028 did not make any consent choice. When a respondent continued past the consent stage, we asked them to report the type of role they were in. Out of the 8,447 consenting (though not necessarily complete) responses, there were 5,728 responses from faculty or principal investigators (PIs), 1,023 responses from post-doctoral researchers, 701 from graduate students in a doctoral program, and 52 from retired scientists. 551 of the remaining respondents were some other type of position and another 392 did not report their position. Thus, the response rate to this single, first-wave email was roughly 2.0%.

We could not use any of this first-wave data in the analyses we conducted here for two reasons. First, the initial-wave survey instrument used did not include all of the research disruption forecast questions; it lacked the two questions regarding "startup" timing and costs, and we received input from respondents and colleagues that these two metrics would be important to capture. Second, the initial-wave instrument did not contain a theoretically correct implementation of the truth-telling incentives (described in the previous section). With the help of an author who developed these incentives (Drazen Prelec), we were able to identify the flaw in our initial implementation and correct it in the instrument used as the basis of this data.

In the second mailing – which generated the data for this study – a total of 108,345 reminder emails were sent (to only addresses that did not participate in the first wave), of which 20,621 emails were undelivered (e.g., because they bounced). Overall 7,263 individuals entered the survey and 6,653 continued past the consent stage. Of those that did not, 513 were not an active scientist, post-doc, or graduate student and thus not within our population of interest, 97 did not consent, and 0 did not make any consent choice. When a respondent continued past the consent stage, we asked them to report the type of role they were in. Out of the 6,653 consenting responses, there were 4,187 responses from faculty or principal investigators (PIs), 528 responses from post-doctoral researchers, 294 from graduate students in a doctoral program, 710 from research staff, 336 from retired but active scientists, and 11 from retired and inactive scientists. The response rate to this second-wave of emails was 7.6% for those that reported their position.

Our response rate may reflect the disruptive nature of the pandemic, but it also raises concerns for generalizability of our results. However, after we received feedback from the initial distribution that many individuals had received the email in their "junk" folder, we became concerned with our distribution being automatically flagged as spam. Based on spot-checking of five individuals that we ex-post identified as being randomly selected by our sample, and who we had professional relationships with, found that in four of the five cases the recruitment email had been flagged as spam. We know of no systematic way of estimating the true spam-flagging rate (nor how to avoid these spam filters when using email distributions at this scale) without using high-end, commercial-grade products.

For our analyses here, we focus entirely on the 4,187 responses from the sample of faculty or PIs. Our goal is to focus only on "academic" scientists – hence our reliance on the peer-reviewed publication record for sampling. To exclude industry scientists, we drop twelve respondents who reported working for any kind of private company or firm based on the name of their institution, flagging institutions as "industry" that include prefixes, suffixes, or any abbreviations that are clearly related to a for-profit entity (e.g., "company", "L.L.C.", "Corp", etc.)

We then drop observations that did not answer any of the research disruption questions or did not report a major field of study (731), so the final sample consists of 3,444 responses.

S3.2 Summary statistics

The broad field breakdown is as follows: Biology or Chemistry or Related (20.5%), Health or Medicine or Related (17.8%), Mathematics or Statistics or Economics or Related (15.1%), Physics or Engineering or Environmental Sciences or Related (24.1%), Sociology or Psychology or Other Social Sciences (22.5%). Of the 3,324 scientists in either the U.S., Canada, Europe, or Western Asia, 73.2% responded from the U.S. Of the 3,014 scientists who reported their gender identity as either female or male, 42.5% reported female. 32.2% of scientists reported having any dependents under 12 years old, 32.1% reported having dependents all of which were 12 years old or over, and the remainder 35.7% reported no dependents. The institutional rating breakdown is as follows: 10.7% reported Terrible or Poor, 30.4% reported Average, 39.3% reported Good, and 19.6% reported Excellent.

As evidence that our allocation of horizon assumptions were randomly presented to respondents, even after selecting this sample, the distribution of assumed future months of the pandemic are as follows: 1 month (15.7%), 2 months (17.2%), 3 months (16.6%), 4 months (16.7%), 6 months (17.0%) 8 months (16.8%).

For each of the five research disruption measures, the sample averages (s.d.) for the measures are as follows: 17.8% decrease in funding (s.d.=36.3); 11.4% decrease in publication quantity (s.d.=31.1); 6.3% decrease in publication impact (s.d.=24.0); 80.2 days to resume a normal pace (s.d.=93.6); 38% of current funding levels to resume a normal pace (s.d.=65.4). For these same measures, the averages of the prediction error from the HVCM predictions are as follows (in the same order): 13.8 p.p., 1.7 p.p., 8.2 p.p., 28.6 p.p., and 37.3 p.p., respectively. These numbers reflect the percentage point difference between the scientist's prediction of where they lie within the distribution of responses from scientists who receive the same horizon assumption, are approximately the same age, and are in the same broad field of study. The positive errors for all metrics indicate that scientists tend to overestimate their location in the distribution.

S3.3 Comparison to The Survey of Doctorate Recipients

As with any opt-in survey, there may be correlations between which scientist choose to opt-in and their experiences about which they want to report. For example, scientists who felt strongly about sharing their situation, whether they experienced large positive or negative changes, may be more likely to respond, which would increase the heterogeneity of the sample. Furthermore, there may also be non-negligible gender differences that arise not due to actual differences in outcomes but due to differences in reporting known to occur across genders.[vii,viii,ix,x]

To estimate the generalizability of our respondent sample, we use the public microdata from The Survey of Doctorate Recipients (SDR) as the best sample estimates of the population of principal investigators in the U.S. The SDR is conducted by the National Center for Science and Engineering Statistics within the National Science Foundation, sampling from individuals who have earned a science, engineering, or health doctorate degree from a U.S. academic institution and are less than 76 years of age. The survey is conducted every two years, and we use the latest data available (2017 cycle). For this comparison, we focus only on university faculty in the SDR and only U.S. respondents in our survey.

First, we compared the distribution of fields of study. Due to variations in field definitions, for the purposes of this comparison we combined the Bio./Chem. fields with the Health/Medicine fields since the SDR field categories do not include many medical-specific fields. For each of the four major groups of disciplines the sample averages and differences are as follows: Biology or Chemistry or Health or Medicine or Related is 40% of our sample and 46% of SDR, Mathematics or Statistics or Economics or Related is 11% of our sample and 13% of SDR, Physics or Engineering or Environmental Sciences or Related is 21% of our sample and 19% of SDR, and Sociology or Psychology or Other Social Sciences is 28% of our sample and 23% of SDR. All of the differences are within roughly two to six p.p., giving us confidence that our sample is not dramatically skewed away from or towards certain broad fields.

When we compare some of the common demographics in the two samples, we find no substantial differences in average age (47.5 in SDR versus 48.0 in our survey), whether the scientist has a partner at home (84% in SDR versus 85% in our survey), or the region of the U.S. where the scientist resides (Midwest: 22% in SDR versus 23% in our survey; West: 22% in SDR versus 27% in our survey; South: 33% in SDR versus 24% in our survey; Northeast: 22% in SDR versus 25% in our survey).

However, we do have a substantially larger number of female respondents, who account for only about 36% of the SDR sample but 50% of our sample, as well as a larger number of respondents with at least one child under 5 or one child between 6 and 11 years old, who account for 10-12% of the SDR sample

but 13-20% of our sample. Since we estimate female scientists and those with these young children report larger disruptions on average, this difference in samples suggests that the overall level of disruptions in our sample may be larger than what the population is experiencing.

S4 Research Disruption Index and regression approach

S4.1 Research Disruption Index

For simplicity, we construct a single index of "Research Disruption" based on scientists' responses to the five disruption questions. To do so, we first ensure the units are relatively comparable by standardizing the responses to each of 5 questions, sorting the values so that large values indicate what would generally be agreed as "worse" outcomes (e.g., less research funding, fewer publications, lower impact publications, more time to restart, more funding to restart). We then sum these five standardized variables and then again standardize that sum so that the magnitudes of the resulting Research Disruption Index are interpretable in terms of standard deviations (i.e., a value of one indicates that a scientist's responses altogether are 1 standard deviation worse than average). This implicitly weights relative differences in each of the five metrics equally.

For reference, a 1 s.d. change (a unit increase in the Index) is roughly equivalent to either a 20% decline in research funding, a 16 or 22% decline in publication quantity or impact, or needing two months longer (or 35% more funding) to recover from the pandemic once it is over.

S4.2 Regression approach

For each respondent i that completes the survey in week w and who belongs to group g (e.g., field, region), we model their Research Disruption Index using the following linear model:

$$ResearchDisruptionIndex_i = \alpha_g + \sum_g \beta_g \times Horizon_i + \omega_w + \varepsilon_i \,,$$

and recover estimates of the vector of group-specific intercepts (α_g), the group-specific horizon coefficients (β_g), and the survey wave intercepts (ω_w) using a separate OLS regression for each of the five groupings reported in Figure 1 Panels B-F of the main text. These estimates are then used to project the group-specific in-sample averages for the dependent variable over the zero- to eight-month horizons as well as the group-specific out-of-sample averages for the nine- to twelve-month horizons as shown in Figure 1 Panels B-F. Across each of the five regressions, all horizon coefficients (β_g) lie within the range of 0 to 0.3, and all standard errors, which are clustered at the level of the week of response, lie within the range of 0.01 to 0.03.

S5 Additional results

S5.1 Role of institutional ratings overall

We further investigate the role of institutions by performing a series of regressions of the following general form:

$$ResearchDisruptionIndex_i = \delta \times InstitutionRating_i + \boldsymbol{\gamma} \times \mathbf{X}_i + \beta \times Horizon_i + \omega_w + \varepsilon_i \,,$$

where i and w index respondents and weeks, respectively, $\mathbf{X}_i$ is a (possibly empty) vector of control variables, $\beta \times Horizon_i$ conditions out average variation due to the horizon assumptions, ω_w are survey

wave intercepts, and $InstitutionRating_i$ takes on the integers one through four corresponding to the four groupings of institutional ratings: Terrible/Poor, Average, Good, and Excellent. δ is the focal coefficient to be estimated, describing the conditional average relationship between scientists' ratings of their institutions and their Research Disruption Index.

With no control variables (i.e., leaving $\mathbf{X}_i$ empty), we estimate δ to equal -0.25 (s.e.=0.017) indicating that each incremental improvement in the rating (e.g., moving from "Good" to "Excellent") is associated with a 0.25 standard deviation decline in the Research Disruption Index. As noted in the main text, we cannot be sure the variation in ratings is due to institutional policies, per se, so this relationship may instead reflect confounding factors.

In our best initial attempt to remove variation in the data that may be due to any confounding factors (and not the policies underling these ratings) we estimate a series of additional regressions with additional controls. As we grow the vector of controls ($\mathbf{X}_i$) to include various combinations of scientists' immediate disruptions due to the pandemic (e.g., per changes in their time allocated to work and their responses to questions about effects they have already experienced), institutional rankings (based on the U.S. News Top 500 Global Universities), the amount of funding the scientist reported, interactions of those variables, as well as controls for field of study, gender, an dependents, we persistently estimate δ to be in the range of -0.15 and -0.25 with standard errors all within 0.017 and 0.019. While these more saturated models are still far from isolating variation in institutional policies that could clearly assign this relationship to the causal effect of institutional policies, they do provide suggestive evidence that these polices are influencing scientists' perceptions of how the pandemic may come to disrupt their research.

S5.2 Role of institutional ratings within groups

We further investigate how the role of institutions may vary within certain groups by performing a series of regressions of the following general form:

$$ResearchDisruptionIndex_i = \sum_g \delta_g \times InstRating_i + \boldsymbol{\gamma} \times \mathbf{X}_i + \beta \times Horizon_i + \omega_w + \varepsilon_i ,$$

which mirrors the prior regression, although we now allow the coefficient on institutional ratings to vary by certain groups of scientists (δ_g).

In all of these regressions, we include the most saturated version of $\mathbf{X}_i$. For reference, in the full sample (when we estimate a single δ with the most saturated control set), we estimate δ to equal -0.15 (s.e.=0.019) indicating that a one-point increase in the institutional rating (on the five-point scale) is associated with a 0.15 standard deviation decline in the forecasted Research Disruption Index.

We find the similar associations between ratings and forecasts when we allow for variations across the various groups: U.S. versus non-U.S., scientists in the hard sciences versus all others, individuals with young dependents versus those without, and male versus female scientists. The coefficients across all groups and models all lie within -0.11 and -0.17 with standard errors all within 0.020 and 0.052. This indicates that, even amongst scientists who tend to have the most pessimistic forecasts in our sample, we find relatively more optimistic forecasts among those who are most pleased with their institution's response to the pandemic. As noted in the main text, this association still may reflect omitted factors that influenced both scientists' forecasts (i.e., being relatively less disrupted even conditional on the

observables) and their ratings (i.e., wanting less support from their institutions because of their relatively lower disruptions). But we take this to be some of the first systematic evidence that some institutions are developing policies and initiatives that can in fact help scientists handle and recover from the disruptions caused by the COVID-19 pandemic.

S6 References for Supplementary Information

[i] Birkle, C., Pendlebury, D. A., Schnell, J. & Adams, J. Web of Science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies* 1, 363–376 (2020).

[ii] Roser, M., Ritchie, H., Ortiz-Ospina, E. & Hasell, J. Coronavirus (COVID-19) deaths. *Our World in Data* https://ourworldindata.org/covid-deaths.

[iii] Ritchie, H., Roser, M., Ortiz-Ospina, E. & Hasell, J. Policy responses to the coronavirus pandemic. *Our World in Data* https://ourworldindata.org/policy-responses-covid.

[iv] Powell, K. Science-ing from home. *Nature* 580, 419–421 (2020).

[v] Cvitanić, J., Prelec, D., Riley, B., & Tereick, B. "Honesty via choice-matching." American Economic Review: Insights, 1.2 (2019): 179-192.

[vi] Myers, K.R., et al. "Unequal effects of the COVID-19 pandemic on scientists." Nature Human Behaviour, 4.9 (2020): 880-883.

[vii] Lundeberg, M. A., Fox, P. W. & Punćcohaŕ, J. Highly confident but wrong: gender differences and similarities in confidence judgments. *Journal of Educational Psychology* 86, 114–121 (1994).

[viii] Bollinger, C. R., Hirsch, B. T., Hokayem, C. M. & Ziliak, J. P. Trouble in the tails? What we know about earnings nonresponse 30 years after Lillard, Smith, and Welch. *Journal of Political Economy* 127, 2143–2185 (2019).

[ix] Bound, J., Brown, C., & Mathiowetz, N. Measurement error in survey data. In *Handbook of econometrics* (Vol. 5, pp. 3705-3843). Elsevier (2001).

[x] Kim, C. & Tamborini, C. R. Response error in earnings: an analysis of the survey of income and program participation matched with administrative data. *Sociological Methods & Research* 43, 39–72 (2014).