Science Translation During the COVID-19 Pandemic: An Academic-Public Health Partnership to Assess Capacity Limits in California

Peter Maldonado, Angie Peng, BA, Derek Ouyang, MS, Jenny Suckale, PhD, and Daniel E. Ho, JD, PhD

On the basis of an extensive academic–public health partnership around COVID-19 response, we illustrate the challenge of science-policy translation by examining one of the most common nonpharmaceutical interventions: capacity limits. We study the implementation of a 20% capacity limit in retail facilities in the California Bay Area.

Through a difference-in-differences analysis, we show that the intervention caused no material reduction in visits, using the same large-scale mobile device data on human movements (mobility data) originally used in the academic literature to support such limits. We show that the lack of effectiveness stems from a mismatch between the academic metric of capacity relative to peak visits and the policy metric of capacity relative to building code.

The disconnect in metrics is amplified by mobility data losing predictive power after the early months of the pandemic, weakening the policy relevance of mobility-based interventions. Nonetheless, the data suggest that a better-grounded rationale for capacity limits is to reduce risk specifically during peak hours. To enhance the connection between science, policy, and public health in future times of crisis, we spell out 3 strategies: living models, coproduction, and shared metrics. (Am J Public Health. 2022;112(2):308–315. https://doi.org/10.2105/AJPH.2021.306576)

Public health responses to COVID-19 have faced serious challenges in light of rapid changes in the scientific understanding of both the virus and the effectiveness of policy responses. This article discusses lessons from an academic–public health partnership around COVID-19 response. We present findings based on a collaboration with the Public Health Department of Santa Clara County, California, one of the largest counties in the United States. In conjunction with 5 other Bay Area counties, Santa Clara was the first jurisdiction in the country to issue a shelter-in-place order in response to COVID-19.1 We illustrate challenges that can arise for evidence-based policy during times of crisis using a case study of a prominent nonpharmaceutical intervention—namely, the implementation of capacity limits on businesses (i.e., restricting businesses to some percent of capacity).

A main contribution of our work is to identify 3 tangible strategies for mutually enhancing science, policy, and public health, based on this partnership. We illustrate the gains to such a model in studying the implementation of a 20% capacity limit starting December 6, 2020, on the main affected sectors—namely, grocery stores, pharmacies, and general merchandise stores. (Indoor restaurant dining was already prohibited at this time.) Using data on human movements (mobility data) from mobile devices in a difference-in-differences framework,2,3 we show that the 20% capacity limit had no significant impact on decreasing the number of visits or peak hour visits, or the length
of visits to businesses in those sectors compared with prepandemic time periods. These are the same measures and data employed in the scientific literature to support capacity limits. The puzzle then is how to reconcile the existing scientific literature, which appears to support such limits, with an intervention that proved ineffective in practice.

To resolve this puzzle, we show that capacity limits were ineffective because of disparate definitions of maximum occupancy adopted by researchers as opposed to policymakers. Although scientists used measures available in retrospective data (e.g., 20% of peak capacity reported after a week from mobility data), policymakers require definitions that can be implemented and enforced on the ground in real time. The result was a limit that did not bind: most businesses were already below the enforced limit at baseline.

This disconnect highlights how profoundly human behavior had already shifted prior to the implementation of the capacity limit. Consistent with other evidence,4,5 we show that mobility loses predictive power of case spread as public health orders are put into place. Scientific studies that anchored capacity limits in associations between public health orders and COVID-19 through capacity limits hold policy relevance over time. The effort to reduce the spread of COVID-19 through capacity limits holds valuable lessons for future policy responses to crises. Through our collaboration with public-sector partners, we identified 3 specific strategies for improvement: ensuring that models used to inform policy are dynamic (living) rather than static, improving collaboration between scientists and policymakers through coproduction (not merely science translation), and shifting to more targeted and enforceable metrics in science.

This article assesses the impact of capacity limits and explains how the limits were mistranslated from academic literature, and then reflects on broader lessons for academic-public health collaborations to improve crisis response.

**IMPACT OF CAPACITY LIMITS**

Capacity limits were motivated by scientific studies showing that restricting visits could decrease the transmission rate of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the novel coronavirus that can cause COVID-19. One such study estimated the impact of reducing visits for 10 different metropolitan areas and found that, for instance, a reduction to 20% of maximum visits in Chicago, Illinois, could reduce new infections by more than 80% while cutting total visits by only 42%.6 The popular press framed this finding around 20% as a “magic number” for implementing capacity limits,7 without articulating what 20% of maximum visits refers to, leaving room for misinterpretation.

A majority of US states have maintained capacity limits in place on various types of businesses (see section A.1 of Appendix, available as a supplement to the online version of this article at http://www.ajph.org). Substantial litigation, up to the US Supreme Court, has involved capacity limits. There is hence urgency to rigorously ground policy in science. We note that our evidence here is limited to retail locations that were not consistently open throughout the entire 2019 and 2020 time period (Appendix, section A.2.1). Figure 2a compares visits for one of these sectors (general merchandise stores) before and after the limit was implemented in December. We defined our sector groups by excluding locations that were not consistently open throughout the entire 2019 and 2020 time period (Appendix, section A.2.1).

Figure 2b compares Santa Clara County (magenta) against San Mateo County (green), which did not implement capacity limits until mandated by the state 2 weeks later. We focused on San Mateo County because it lies just north of Santa Clara County, exhibits similar economic activity, and had comparable pretreatment visit time series, yet adopted a starkly different approach to capacity limits. San Mateo’s health...
officer, for instance, expressed “grave concerns about the unintended consequences of reducing [the county’s] grocery store capacity to 20%.”

To formally evaluate the impact, we created 2 comparisons. First, we compared the change in visits before and after the December 2020 order 10 to the same period 1 year earlier in 2019 for a difference-in-differences analysis. We used October 26 to December 6 as the preperiod when the capacity threshold was at 50% 12 and established a postperiod of December 7 to January 17. The same group of stores in 2019 formed the comparison group. (We note that the period between November 29 and December 6 was subject to an interim announcement of different capacity limits, which was quickly revised in favor of the 20% order, and our results are substantively identical when omitting this period.) Second, we compared the change in visits before and after the December 2020 order with the contemporaneous period in neighboring San Mateo County. Here, we used the same preperiod and an adjusted postperiod of December 7 to December 17, the day

**FIGURE 1** The Weekly Average Visits for Restaurants: Santa Clara County, CA, 2019–2020

*Note.* The green line shows the 2019 weekly average visit trend, and the magenta line shows the 2020 weekly average trend.

**FIGURE 2** Normalized Weekly Average Visits for General Merchandise Stores for (a) Santa Clara County Locations in 2019–2020 vs 2020–2021, and (b) San Mateo County vs Santa Clara County Locations: California

*Note.* Panel a shows the normalized weekly average visits by sector for Santa Clara County locations for general merchandise stores affected by the December 20% order. The green line shows the 2019–2020 weekly average visit trend, and the magenta line shows the 2020–2021 weekly average trend. The black dotted vertical line shows the implementation of the 20% capacity order on December 6. Panel b shows the normalized weekly average visits for general merchandise stores for San Mateo County (green line) and Santa Clara County (magenta line). The black dotted vertical line shows the implementation of the 20% capacity order on December 6 in Santa Clara, and the green dotted line shows when San Mateo County adopted the same order.
when all Bay Area counties became subject to the 20% capacity limit for these sectors (Appendix, section A.4). We note that the intent of the capacity limits was to have immediate effect, given the surge in cases.

If capacity limits curbed behavior, we would expect to see a drop in daily visits or daily peak hour visits following the implementation of the limit: the 20% capacity limit was originally supported by scientific literature utilizing the same SafeGraph mobility data set. However, we found no clear reduction in overall daily visits, the daily peak hourly visits, or the median number of minutes spent in store per visit, in either the 2020 versus 2019 comparison or the Santa Clara and San Mateo County comparison (Appendix, section A.4).

When comparing Santa Clara County visits in 2020 with the comparable time period in 2019, we did not see a significant decrease in daily visits, peak hour visits, or median visit time for pharmacies or groceries after the 20% capacity was implemented relative to the control group. We observed a slight decrease in daily visits and peak hour visits to general merchandise stores, but when we conducted a series of lead tests to test the parallel trends assumption, we saw that this effect was detected ahead of the December 6 order (Appendix, section A.6 for full analysis), suggesting that this comparative drop in visits between the 2 years predated the order.

When comparing neighboring Santa Clara County with San Mateo County under differing policies, we did not observe statistically significant decreases (at $\alpha = .05$) in daily visits, peak hour visits, or visit times. We observed 1 decrease in daily visits to groceries ($P = .06$). This effect was not corroborated by the first difference-in-differences design and may be an artifact from 9 tests conducted across 3 outcomes and 3 types of facilities. We also showed that there was no evidence of spillover effects (i.e., individuals visiting San Mateo because of the Santa Clara capacity limit; Appendix, section A.8). Comparing early versus late adopting counties across the Bay Area, we also observed no substantial decrease in visits upon the enactment of capacity limits (Appendix, section A.11).

**Explanation of Effects**

Why were the effects so negligible? We showed that the lack of effectiveness stemmed from differences in the definitions of occupancy and capacity between academic research and government. Government implementation focuses on the enforceable measures, such as 20% of the posted maximum occupancy by the fire code. The presence of the posted sign enabled inspectors to quickly check whether a facility violated the capacity limit (Appendix, section A.12). By contrast, academia may use convenience measures of maximum occupancy. One study, for instance, used the highest single number of hourly visits for each location. This definition is convenient for measuring in historical mobility data, but would be difficult to enforce. Store managers and inspectors would need to know the maximum number of visits to each location over specific periods of time.

To illustrate the significance of this deviation, Figure 3 compares the academic capacity limit on the x-axis against California’s enacted capacity limit on the y-axis for all stores (see Appendix, section A.2.4 for details). If definitions were comparable, the limits should line up on the dashed 45-degree line, but 77% of stores had a higher enacted capacity limit compared with the academic notion. The policy limit was, on average, at least 203% greater than the academic limit (see Appendix, section A.12 for additional comparisons). Put another way, although this was dependent on the baseline, a 5% capacity limit based on building square footage would have capped maximum occupancy at 20% of maximum mobility during the baseline time period used to compute the academic limit (Appendix, section A.2.4).

We then compared the enacted limit against baseline visits and showed that the vast majority of retail activity already complied with the 20% limit prior to the December restriction (both before and during the pandemic). We used estimated hourly occupancy from SafeGraph’s visits, dwell time, and square footage data for each location (Appendix, section A.2.5). We then compared the distribution of the average hourly occupancy of 744 grocery stores, pharmacies, and general merchandise stores before and after the Santa Clara County 20% capacity order went into effect with the same time periods during the prior year (Appendix, section A.14). When we compared the 6 weeks before and after December 6, 2020 (when the 20% capacity limit was in effect) to the same weeks in the previous year, locations were rarely above the capacity limits set by the Santa Clara County order, with or without the 20% order in place. There was also no notable decrease in occupancy when the capacity limit was in place. Even for locations that had experienced occupancy greater than 20% before the Santa Clara County order, there were only a few “peak hours” when occupancy was above 20% (Appendix, section A.15).
Across all 4 time periods, each sector displayed similar distributions of average hourly occupancy. Notably, the gap in number of nonzero occupancy hours across sectors was larger between 2019 and 2020 compared with the gap before and after the December 6 order in 2020. This finding suggests that there were already significant behavior changes before the December capacity limit, lessening its impact.

The analysis presented in this section illustrates the potential for mistranslating scientific findings into policy based on metric definitions and static models that fail to capture evolving human behavior. California’s implementation of the 20% capacity limit fell seriously short of what was warranted by the underlying science. The peak hour finding, however, does suggest an alternative rationale—distinct from prior accounts—for the capacity limits: reducing spread during the few (peak) hours of high risk, while minimizing disruption to business.

Across all 4 time periods, each sector displayed similar distributions of average hourly occupancy. Notably, the gap in number of nonzero occupancy hours across sectors was larger between 2019 and 2020 compared with the gap before and after the December 6 order in 2020. This finding suggests that there were already significant behavior changes before the December capacity limit, lessening its impact.

The analysis presented in this section illustrates the potential for mistranslating scientific findings into policy based on metric definitions and static models that fail to capture evolving human behavior. California’s implementation of the 20% capacity limit fell seriously short of what was warranted by the underlying science. The peak hour finding, however, does suggest an alternative rationale—distinct from prior accounts—for the capacity limits: reducing spread during the few (peak) hours of high risk, while minimizing disruption to business.

Across all 4 time periods, each sector displayed similar distributions of average hourly occupancy. Notably, the gap in number of nonzero occupancy hours across sectors was larger between 2019 and 2020 compared with the gap before and after the December 6 order in 2020. This finding suggests that there were already significant behavior changes before the December capacity limit, lessening its impact.

The analysis presented in this section illustrates the potential for mistranslating scientific findings into policy based on metric definitions and static models that fail to capture evolving human behavior. California’s implementation of the 20% capacity limit fell seriously short of what was warranted by the underlying science. The peak hour finding, however, does suggest an alternative rationale—distinct from prior accounts—for the capacity limits: reducing spread during the few (peak) hours of high risk, while minimizing disruption to business.

Across all 4 time periods, each sector displayed similar distributions of average hourly occupancy. Notably, the gap in number of nonzero occupancy hours across sectors was larger between 2019 and 2020 compared with the gap before and after the December 6 order in 2020. This finding suggests that there were already significant behavior changes before the December capacity limit, lessening its impact.

The analysis presented in this section illustrates the potential for mistranslating scientific findings into policy based on metric definitions and static models that fail to capture evolving human behavior. California’s implementation of the 20% capacity limit fell seriously short of what was warranted by the underlying science. The peak hour finding, however, does suggest an alternative rationale—distinct from prior accounts—for the capacity limits: reducing spread during the few (peak) hours of high risk, while minimizing disruption to business.

Across all 4 time periods, each sector displayed similar distributions of average hourly occupancy. Notably, the gap in number of nonzero occupancy hours across sectors was larger between 2019 and 2020 compared with the gap before and after the December 6 order in 2020. This finding suggests that there were already significant behavior changes before the December capacity limit, lessening its impact.

The analysis presented in this section illustrates the potential for mistranslating scientific findings into policy based on metric definitions and static models that fail to capture evolving human behavior. California’s implementation of the 20% capacity limit fell seriously short of what was warranted by the underlying science. The peak hour finding, however, does suggest an alternative rationale—distinct from prior accounts—for the capacity limits: reducing spread during the few (peak) hours of high risk, while minimizing disruption to business.

Across all 4 time periods, each sector displayed similar distributions of average hourly occupancy. Notably, the gap in number of nonzero occupancy hours across sectors was larger between 2019 and 2020 compared with the gap before and after the December 6 order in 2020. This finding suggests that there were already significant behavior changes before the December capacity limit, lessening its impact.

The analysis presented in this section illustrates the potential for mistranslating scientific findings into policy based on metric definitions and static models that fail to capture evolving human behavior. California’s implementation of the 20% capacity limit fell seriously short of what was warranted by the underlying science. The peak hour finding, however, does suggest an alternative rationale—distinct from prior accounts—for the capacity limits: reducing spread during the few (peak) hours of high risk, while minimizing disruption to business.

Across all 4 time periods, each sector displayed similar distributions of average hourly occupancy. Notably, the gap in number of nonzero occupancy hours across sectors was larger between 2019 and 2020 compared with the gap before and after the December 6 order in 2020. This finding suggests that there were already significant behavior changes before the December capacity limit, lessening its impact.

The analysis presented in this section illustrates the potential for mistranslating scientific findings into policy based on metric definitions and static models that fail to capture evolving human behavior. California’s implementation of the 20% capacity limit fell seriously short of what was warranted by the underlying science. The peak hour finding, however, does suggest an alternative rationale—distinct from prior accounts—for the capacity limits: reducing spread during the few (peak) hours of high risk, while minimizing disruption to business.

Across all 4 time periods, each sector displayed similar distributions of average hourly occupancy. Notably, the gap in number of nonzero occupancy hours across sectors was larger between 2019 and 2020 compared with the gap before and after the December 6 order in 2020. This finding suggests that there were already significant behavior changes before the December capacity limit, lessening its impact.

The analysis presented in this section illustrates the potential for mistranslating scientific findings into policy based on metric definitions and static models that fail to capture evolving human behavior. California’s implementation of the 20% capacity limit fell seriously short of what was warranted by the underlying science. The peak hour finding, however, does suggest an alternative rationale—distinct from prior accounts—for the capacity limits: reducing spread during the few (peak) hours of high risk, while minimizing disruption to business.

Across all 4 time periods, each sector displayed similar distributions of average hourly occupancy. Notably, the gap in number of nonzero occupancy hours across sectors was larger between 2019 and 2020 compared with the gap before and after the December 6 order in 2020. This finding suggests that there were already significant behavior changes before the December capacity limit, lessening its impact.

The analysis presented in this section illustrates the potential for mistranslating scientific findings into policy based on metric definitions and static models that fail to capture evolving human behavior. California’s implementation of the 20% capacity limit fell seriously short of what was warranted by the underlying science. The peak hour finding, however, does suggest an alternative rationale—distinct from prior accounts—for the capacity limits: reducing spread during the few (peak) hours of high risk, while minimizing disruption to business.

Across all 4 time periods, each sector displayed similar distributions of average hourly occupancy. Notably, the gap in number of nonzero occupancy hours across sectors was larger between 2019 and 2020 compared with the gap before and after the December 6 order in 2020. This finding suggests that there were already significant behavior changes before the December capacity limit, lessening its impact.

The analysis presented in this section illustrates the potential for mistranslating scientific findings into policy based on metric definitions and static models that fail to capture evolving human behavior. California’s implementation of the 20% capacity limit fell seriously short of what was warranted by the underlying science. The peak hour finding, however, does suggest an alternative rationale—distinct from prior accounts—for the capacity limits: reducing spread during the few (peak) hours of high risk, while minimizing disruption to business.

Across all 4 time periods, each sector displayed similar distributions of average hourly occupancy. Notably, the gap in number of nonzero occupancy hours across sectors was larger between 2019 and 2020 compared with the gap before and after the December 6 order in 2020. This finding suggests that there were already significant behavior changes before the December capacity limit, lessening its impact.

The analysis presented in this section illustrates the potential for mistranslating scientific findings into policy based on metric definitions and static models that fail to capture evolving human behavior. California’s implementation of the 20% capacity limit fell seriously short of what was warranted by the underlying science. The peak hour finding, however, does suggest an alternative rationale—distinct from prior accounts—for the capacity limits: reducing spread during the few (peak) hours of high risk, while minimizing disruption to business.

Across all 4 time periods, each sector displayed similar distributions of average hourly occupancy. Notably, the gap in number of nonzero occupancy hours across sectors was larger between 2019 and 2020 compared with the gap before and after the December 6 order in 2020. This finding suggests that there were already significant behavior changes before the December capacity limit, lessening its impact.

The analysis presented in this section illustrates the potential for mistranslating scientific findings into policy based on metric definitions and static models that fail to capture evolving human behavior. California’s implementation of the 20% capacity limit fell seriously short of what was warranted by the underlying science. The peak hour finding, however, does suggest an alternative rationale—distinct from prior accounts—for the capacity limits: reducing spread during the few (peak) hours of high risk, while minimizing disruption to business.

Across all 4 time periods, each sector displayed similar distributions of average hourly occupancy. Notably, the gap in number of nonzero occupancy hours across sectors was larger between 2019 and 2020 compared with the gap before and after the December 6 order in 2020. This finding suggests that there were already significant behavior changes before the December capacity limit, lessening its impact.

The analysis presented in this section illustrates the potential for mistranslating scientific findings into policy based on metric definitions and static models that fail to capture evolving human behavior. California’s implementation of the 20% capacity limit fell seriously short of what was warranted by the underlying science. The peak hour finding, however, does suggest an alternative rationale—distinct from prior accounts—for the capacity limits: reducing spread during the few (peak) hours of high risk, while minimizing disruption to business.

Across all 4 time periods, each sector displayed similar distributions of average hourly occupancy. Notably, the gap in number of nonzero occupancy hours across sectors was larger between 2019 and 2020 compared with the gap before and after the December 6 order in 2020. This finding suggests that there were already significant behavior changes before the December capacity limit, lessening its impact.

The analysis presented in this section illustrates the potential for mistranslating scientific findings into policy based on metric definitions and static models that fail to capture evolving human behavior. California’s implementation of the 20% capacity limit fell seriously short of what was warranted by the underlying science. The peak hour finding, however, does suggest an alternative rationale—distinct from prior accounts—for the capacity limits: reducing spread during the few (peak) hours of high risk, while minimizing disruption to business.

Across all 4 time periods, each sector displayed similar distributions of average hourly occupancy. Notably, the gap in number of nonzero occupancy hours across sectors was larger between 2019 and 2020 compared with the gap before and after the December 6 order in 2020. This finding suggests that there were already significant behavior changes before the December capacity limit, lessening its impact.

The analysis presented in this section illustrates the potential for mistranslating scientific findings into policy based on metric definitions and static models that fail to capture evolving human behavior. California’s implementation of the 20% capacity limit fell seriously short of what was warranted by the underlying science. The peak hour finding, however, does suggest an alternative rationale—distinct from prior accounts—for the capacity limits: reducing spread during the few (peak) hours of high risk, while minimizing disruption to business.
relationship, policy measures based on mobility patterns may impose hardship without affecting case growth.

The 2 largest deviations occurred in the fall and winter of 2020, when mobility–case growth correlation became positive and negative, respectively (Appendix, section A.16.3–4). These dynamics illustrate that mobility is not created equally: notwithstanding county nonpharmaceutical interventions, holiday travel, for instance, may be associated with greater risk of exposure than ordinary commuting, hence generating the positive association between mobility and case growth in winter 2020.

The pattern of strong-then-weakening correlation is significant because it suggests that mobility-based models developed at the beginning of the pandemic lose the ability to predict after the first few months. Because scientific findings around mobility and case growth were disproportionately based in the early months of the pandemic, they are less applicable for policy in later months, when the mobility–case growth relationship is weaker. Continuous (living) models that capture behavioral changes are critical for strengthening the evidence base in a rapidly evolving crisis.

Living models are also important when those performing data analysis are separate from data producers. During the research process, for instance, we identified real-time changes in SafeGraph’s data schema that, left unaddressed, could potentially confound intertemporal comparisons. When data are released (and modified) in real time, living models can more easily account for such changes.

Coproduction Instead of Science Translation

The traditional view of science translation is based on a linear model of knowledge production, which entails a unidirectional flow of information from researchers to policymakers. However, the rapidly evolving nature of the pandemic and human behavior during this unusually disruptive time means that policy priorities shift over time. We...
argue that the gap between scientific research and policy can be substantially shrunk through coproduction of solutions. Coproduction acknowledges that researchers and decision-makers hold “complementary and overlapping knowledge and skills that are essential for problem-solving.” In particular, policymakers often have insight into novel problems and constraints not yet considered by science. Thus, these types of partnerships are not merely translational, but rather reframe and redefine the nature of the questions posed. Our assessment of the effects of the public health order in Santa Clara County was only possible because of a direct partnership on implementation details. Such partnerships can enable real-time evaluations of the impact of policy implementation and also strengthen and solidify the feedback loop, especially for health policy.

Particularly, monitoring that includes up-to-date “best guess” estimates regarding the impact of ongoing policy interventions also allows for quicker diagnoses and adaptation of policy measures. Although conventional academic incentives are not well suited for this form of coproduction (e.g., publication timelines, negotiation of data sharing agreements), the ability of scientists and policymakers to coproduce strengthens both research and policy.

**Syncing Science and Policy Metrics**

Although the pandemic has transformed policy, it is a growth opportunity for impact-oriented science as well. Studies using convenience measures that are infeasible to implement are not useful for crisis response. Instead, scientists should work to incorporate such policy constraints into their models. Defining a capacity limit through aggregated mobility data reveals little about the spatial density of individuals within a store and does not necessarily equate to a direct reduction in physical or social contact. If we were to use hourly store visits as a proxy for social distancing, though, there are generally only a few hours of the day when managing capacity is most important (Appendix, section A.15). A more targeted approach to improving safety measures and enforcing capacity limits during these hours could be more effective than a blanket “magic number” capacity across all locations and hours. This also focuses the intervention on a more measurable and enforceable metric, namely, total number of visits during specific store hours.

During times of crisis, effective public health policy is rarely achieved by a one-size-fits-all approach, as human behavior evolves rapidly, informed by both health risk and economic hardship. Through the combination of living models, academic–public policy coproduction, and incorporation of policy constraints into science, there is a greater opportunity for policy interventions to be strengthened by research.

**ABOUT THE AUTHORS**

Peter Maldonado is with the Department of Computer Science and the Regulation, Evaluation, and Governance Lab (RegLab), Stanford University, Stanford, CA. Angie Peng is with the Department of Management Science and Engineering, School of Engineering, Stanford University. Derek Ouyang is with the Future Bay Initiative and the RegLab, Stanford University. Jenny Suckale is with the Department of Geophysics at Stanford Earth, School of Earth, Energy, and Environmental Sciences, and the Future Bay Initiative, Stanford University. Daniel E. Ho is with Stanford Law School, Department of Political Science, School of Humanities and Sciences, the Stanford Institute for Economic Policy Research, the Stanford Institute for Human-Centered Artificial Intelligence, and the RegLab, Stanford University.

**REFERENCES**