



CORONAVIRUS

Epidemic outcomes following government responses to COVID-19: Insights from nearly 100,000 models

Eran Bendavid^{1,2*} and Chirag J. Patel³

Government responses to COVID-19 are among the most globally impactful events of the 21st century. The extent to which responses—such as school closures—were associated with changes in COVID-19 outcomes remains unsettled. Multiverse analyses offer a systematic approach to testing a large range of models. We used daily data on 16 government responses in 181 countries in 2020–2021, and 4 outcomes—cases, infections, COVID-19 deaths, and all-cause excess deaths—to construct 99,736 analytic models. Among those, 42% suggest outcomes improved following more stringent responses (“helpful”). No subanalysis (e.g. limited to cases as outcome) demonstrated a preponderance of helpful or unhelpful associations. Among the 14 associations with P values $< 1 \times 10^{-30}$, 5 were helpful and 9 unhelpful. In summary, we find no patterns in the overall set of models that suggests a clear relationship between COVID-19 government responses and outcomes. Strong claims about government responses’ impacts on COVID-19 may lack empirical support.

INTRODUCTION

COVID-19 was—and to a large extent remains—the most meaningful health event in recent global history (1). Unlike the 2003 Severe Acute Respiratory Syndrome (SARS) epidemic, it spread globally; unlike Zika, everyone is at risk of infection with COVID-19; and unlike recent swine flu pandemics, the disease severity and mortality from COVID-19 were so high it led to life expectancy reversals in many countries (2, 3).

If COVID-19 was a defining health event, the global responses to COVID-19 were a defining health policy experience (4, 5). The swiftness of global responses, their extensiveness, and direct implications for billions of people’s lives were historically unique: The responses to the 1918 influenza pandemic, in comparison, were largely localized, while the global response to the HIV pandemic was slower and smaller in extent than the response to COVID-19 (6, 7). Government responses to COVID-19 intended to limit the virus’ spread and disease burden, using encouragements or mandates on schools, travel, and masks, among others, as well as income support or debt relief to enable social distancing.

The rapid spread of the virus in early 2020 meant that many COVID-19 responses were implemented swiftly, based on partial information, often from simulation models, about transmission mechanisms and about anticipated benefits (8, 9). The swiftness of spread afforded effectively no time for careful studies of policy effects, and favored emergency measures implemented with relatively little information about the trade-offs of alternative policy options (10).

Many approaches are needed to understand the impacts of government responses to the pandemic. Qualitative approaches may help with understanding why different governments used different policy responses (for example, why Norway implemented shelter-in-place while Sweden did not). Observational epidemiologic studies may help characterize the relationships between different policy responses and COVID-19 outcomes, while meta-analyses and systematic reviews can summarize the observational evidence on government response impacts (11, 12). Experimental evidence is not available for

understanding the impact of policies: No government studied its responses directly with trials or experiments. As a result, one common thread is that the data available for studying policy responses are messy and complex, resulting in analyses that may also be complex (13). For example, most policy responses were implemented concurrently or in close sequence, posing challenges to identifying the unique impacts of individual policies (14). Existing studies of COVID-19 response impacts range from unrealistically positive to dismissively negative, further complicating balanced assessments (15–17).

Despite the complexity of the endeavor, its importance is undiminished. Definitive studies of government response impacts on the virus’ spread and disease burden would be enormously helpful for present decisions and future pandemic planning. The dearth of prospective and randomized studies means that, likely, no single study may settle this question.

In this analysis, we attempt to advance the science of government responses to COVID-19 by taking a multiverse approach to this topic (18–21). Multiverse analyses elevate epistemic humility by relaxing the number of subjective choices in the research design process. Multiverse approaches also prompt analysts to comprehensively probe the space of plausible models and results of assumptions. By expanding the number of analyses, they provide information about the stability of relationships’ magnitude and direction due to study design parameters and choice. We take this approach because (i) the data available for analysis are complex and rich, making possible a large number of plausible analyses and (ii) we aspire to limit the role of data and model choices in driving the results, or “researcher degree of freedom” (22). The emerging distributions of possible relationships can be considered an update to the strength of hypotheses about the effectiveness of COVID-19 responses that are in contrast with much of the highly cited literature. At the very least, the breadth of possible findings provides an understanding of what can or cannot be answered with limited observational data.

RESULTS

The daily analytic dataset includes 128,662 observations, an average of 711 observations in 181 countries in 2020 and 2021. The weekly and monthly datasets contain 18,795 and 4198 observations (an average of 103 and 23 observations per country, respectively). The average

Copyright © 2024 the
Authors, some rights
reserved; exclusive
licensee American
Association for the
Advancement of
Science. No claim to
original U.S.
Government Works.
Distributed under a
Creative Commons
Attribution License 4.0
(CC BY).

Downloaded from <https://www.science.org> on August 11, 2024

¹Department of Medicine, Stanford University, Stanford, CA, USA. ²Department of Health Policy, Stanford University, Stanford, CA, USA. ³Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA.

*Corresponding author. Email: ebd@stanford.edu

number of observations in each analysis was 18,930 and ranged for reasons such as period of aggregation (monthly, weekly, or daily), partial availability of outcome (especially for excess deaths), or outcome counts of zero, which prevent growth calculations. The earliest date with nonzero COVID-19 outcomes data is 22 January 2020.

Table 1 summarizes government response data. Across all countries, the containment measures that comprise the stringency index peaked, on average, in April or May 2020. Health system responses, especially access to testing and vaccine availability, peaked later and remained at their peak through the end of the study period in December 2021. The country with the highest stringency index for the entire study period was Honduras (average 73.8; China and the United States, by comparison, had average stringency of 69.7 and 56.3, respectively). Figure S1 provides an illustrative example of the time trends of four government responses in four countries: stringency index, school closure, workplace closure, and vaccine availability in the United States, South Korea, Sweden, and India.

Figure 1 provides a visual representation of the variations used to generate all 99,736 models. Of all models, 41,748 (42%) had a point estimate in the “helpful” direction, and 57,988 (58%) in the “unhelpful” direction (we use “helpful” as shorthand for a negative β_1 and “unhelpful” for a positive β_1 , without implying a causal link). The number of significant associations was similar between the helpful

and unhelpful models: 3692 (8.8%) of the helpful and 3811 (6.7%) of the unhelpful associations were statistically significant using a false discovery rate criteria (23).

Figure 2 illustrates the direction of the evidence among the analyses. Each panel demonstrates the portion of the models that are helpful and unhelpful, by subgroup. Figure 2A demonstrates that about half of all models suggest government responses were helpful, and half unhelpful when examining either of the three indices (stringency, government response, and economic support). A similar picture emerges when looking at subgroups by time period (about 50:50 for early 2020, all of 2020, or the entire time period), outcome measure (infections, cases, COVID-19 deaths, or excess deaths), dataset aggregation period (daily, weekly, or monthly), fixed effects (with or without), and covariates (with or without). The only subgroup level where the balance shifts away from 50:50 is the model type. Models 1, 3, and 5 range from 60 to 80% in the unhelpful direction, while models 2, 4, and 6 range from 55 to 70% in the helpful direction. When removing model 1, the simplest and least balanced (80% of results in the unhelpful direction) model, then, among the remaining 82,864 estimates, 46% have a point estimate in the helpful direction and 54% in the unhelpful direction. Figure 3 shows that the distribution of the standardized effect sizes (proxied using the *t*-statistic from each model) for the overall set of 99,736

Table 1. Descriptive features of COVID-19 government response data. The top eight responses are categorized by OxCGRT as “containment measures,” the next two as “economic measures,” and the next six as “health system measures.” The composite indices are shown at the bottom. The mean averages the response across all countries from 22 January 2020 to 21 December 2021. Max month indicates the month in which the indicator was highest across all countries. Max country refers to the country in which the indicator was highest over the entire period.

Response	N	Min	Max	Mean	Max month	Max country
School closure	126,946	0	3	1.6	4/20	Saudi Arabia
Workplace closure	123,319	0	3	1.4	4/20	Chile
Canceling public events	123,801	0	2	1.4	4/20	China
Limits on gathering	125,680	0	4	2.7	5/20	China
Closing public transport	127,661	0	2	0.6	4/20	Argentina
“Stay-at-home” orders	128,311	0	3	1.0	4/20	China
Restrict internal travel	127,029	0	2	0.8	4/20	Jamaica
Restrict internal travel	122,104	0	4	2.5	4/20	Kiribati
Income support	128,510	0	2	0.8	6/20	Belgium
Debt relief	128,510	0	2	0.9	6/20	Hong Kong
Public info campaigns	128,509	0	2	1.8	5/20	Uzbekistan
Access to testing	128,510	0	3	1.9	12/21	China
Contact tracing	128,510	0	2	1.3	11/20	Singapore
Masking policies	128,242	0	4	2.4	4/21	Singapore
Vaccine availability	128,829	0	5	1.7	12/21	Israel
Protecting elderly	127,773	0	3	1.3	4/20	Hong Kong
Stringency index	125,670	0	100	52.4	4/20	Honduras
Government response index	125,670	0	91.2	51.0	4/20	Italy
Econ support index	128,510	0	100	40.0	6/20	Cyprus

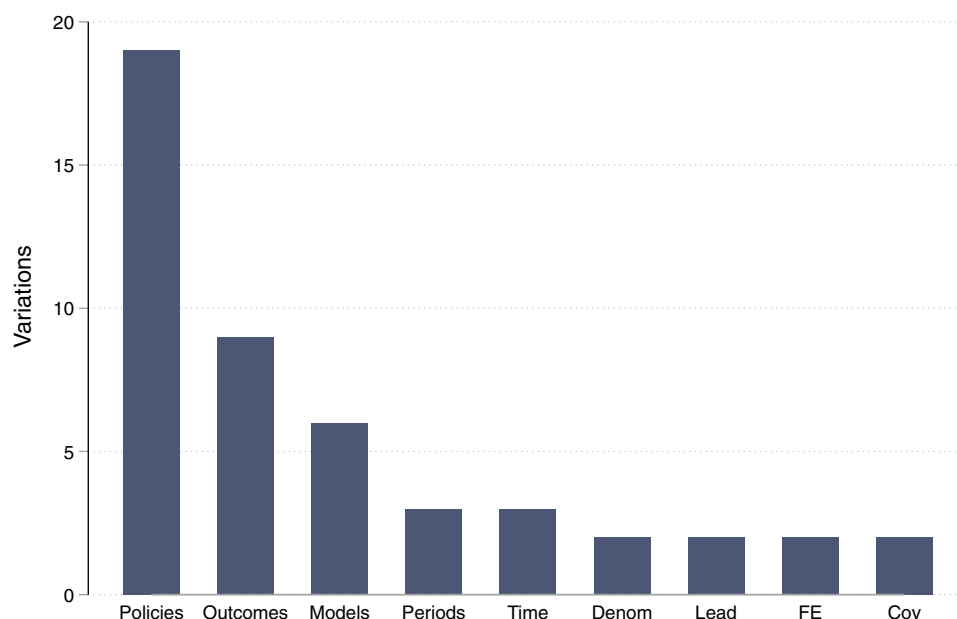


Fig. 1. Analysis and data variations. The product of all the variations (147,744)—19 response types (policies), 9 outcomes, 6 models, 3 time periods, 3 time aggregations, 2 denominators, 2 leads, 2 fixed effects, and 2 covariates—is greater than the number of models actually run (99,736). This is because some combinations were not feasible or reasonable. For example, only one lead (1 month) was used when the dataset was aggregated at the monthly level.

models and for five subsets are evenly and narrowly centered around zero (with small deviations from zero among the most skewed models in Fig. 2).

The five response-outcome pairs with the most consistent associations in the helpful and unhelpful directions are shown in Table 2. The number of infections makes up three of the five most consistently helpful associations, while excess all-cause deaths and COVID-19 deaths make up the outcome in the majority of unhelpful associations. Among the 14 most extremely strong associations with $P < 1 \times 10^{-30}$, 5 were in the helpful direction and 9 were in the unhelpful direction.

Last, a total of 12 models could be applied to the simulated measles dataset (six models, each with and without fixed effects). All results had effect sizes in the helpful direction, with all $P < 3.1 \times 10^{-6}$ and t -statistic ≤ -5.3 .

DISCUSSION

In this study, we perform a multiverse analysis of nearly 100,000 ways of probing the relationship between COVID-19 government responses and outcomes in 181 countries. The goal is to create a multiverse of plausible analyses and assess the sensitivity of the results to these choices. Exploring the multiverse for a question of high importance may be useful where there is no consensus. In this study, we found no clear pattern in the overall set of analyses or in any subset of analyses. We are left to conclude that strong claims about the impact of government responses on the COVID-19 burden lack empirical support.

Inferences from this analysis deserve careful consideration, including a clear understanding of what this study cannot illuminate. First, none of the models tested can tell the extent to which any government response could have improved COVID-19 outcomes. Perhaps with another virus, other implementation strategies, or different

populations, school closures could have extinguished transmission. Nor can we learn from this study what COVID-19 outcomes would have been like in the absence of these responses. Second, our analysis is global in scope and examines government responses and COVID-19 outcomes at the level of countries. This is suitable for inferring global patterns and trends but cannot exclude patterns at state, district, community, or even neighborhood levels.

Third, and perhaps most importantly, we cannot conclude that there is compelling evidence to support the notion that government responses improved COVID-19 burden, and we cannot conclude that there is compelling evidence to support the notion that government responses worsened the COVID-19 burden. The concentration of estimates around a zero effect weakly suggests that government responses did little to nothing to change the COVID-19 burden.

This conclusion departs meaningfully from many scientific studies of government responses. For example, a highly cited study on this topic notes that “Our results show that major non-pharmaceutical interventions—and lockdowns in particular—have had a large effect on reducing transmission” (9). Such conclusions are common in the scientific literature (table S1), but our analysis—extensive in scope and outcomes—suggests that such strong claims lack empirical justification.

The contribution of any study can be thought of as an update to the reader’s Bayesian prior. Most scientific studies aim to strengthen the reader’s posterior belief in a hypothesis, while this study explores the opposite: We argue that strong beliefs about the impact of COVID-19 government responses, as reflected in the studies in table S1, may deserve weakening.

We propose several ways to reduce such uncertainty in the design and evaluation of public health programs and policies. Perhaps the most important foundation for a better understanding of policy impacts is prospective, representative (or population-level), and

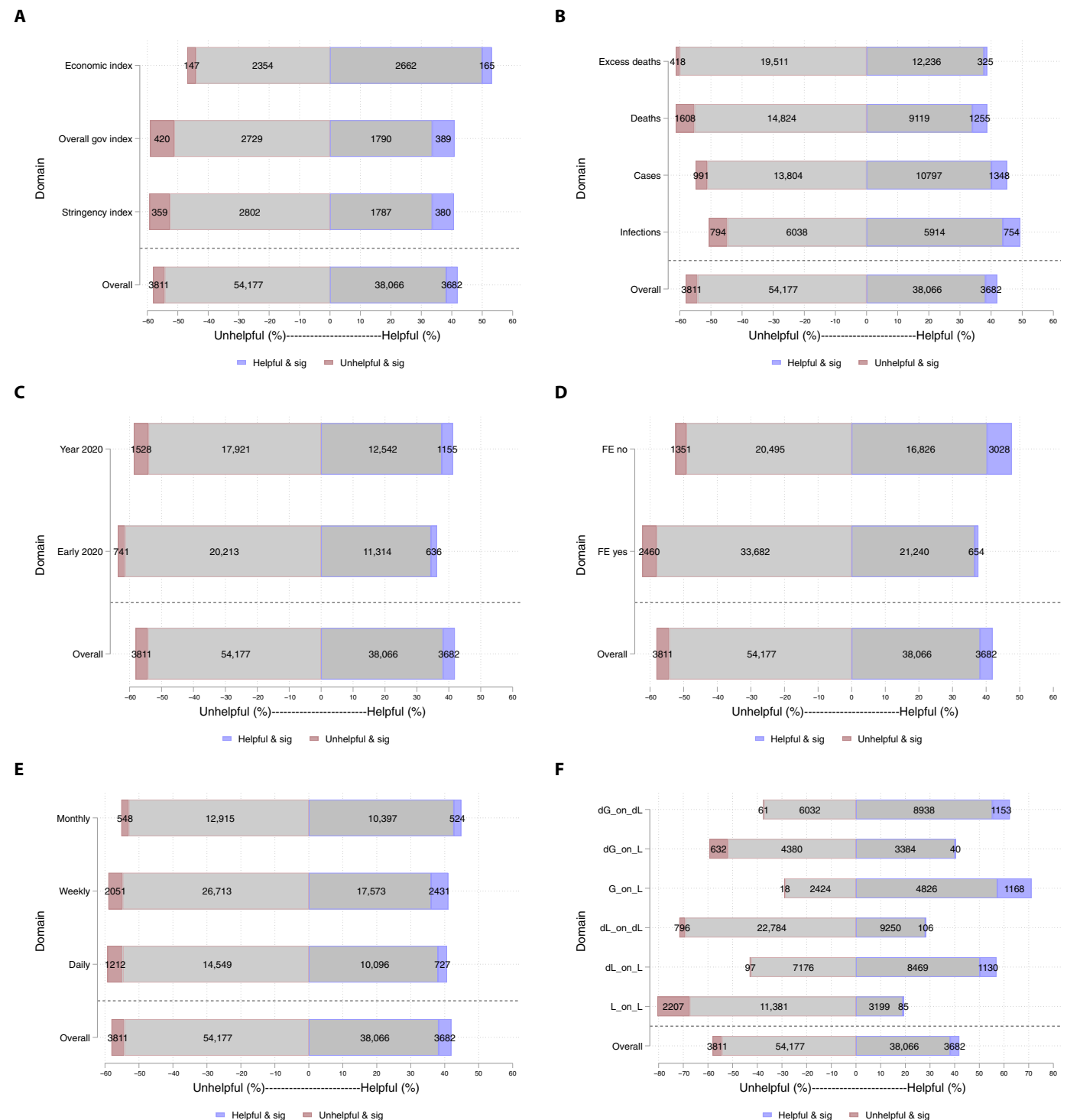


Fig. 2. Portion of models with point estimates in the helpful or unhelpful direction, by group. The portion of models with helpful and unhelpful associations, by domain. Each panel includes the overall distribution of effects (bottom bar), and the models that were significant (solid colors). Panel (A) shows the distribution for each of the three government response indices, panel (B) for the four outcome types (cases, infections, COVID-19 deaths, and all-cause excess deaths), panel (C) for the three study periods, panel (D) for models with and without fixed effects, panel (E) for the three time aggregations, and panel (F) for the six models (1 bottom, 6 top). The models are indexed as L (levels) or G (growth), with d indicating a difference.

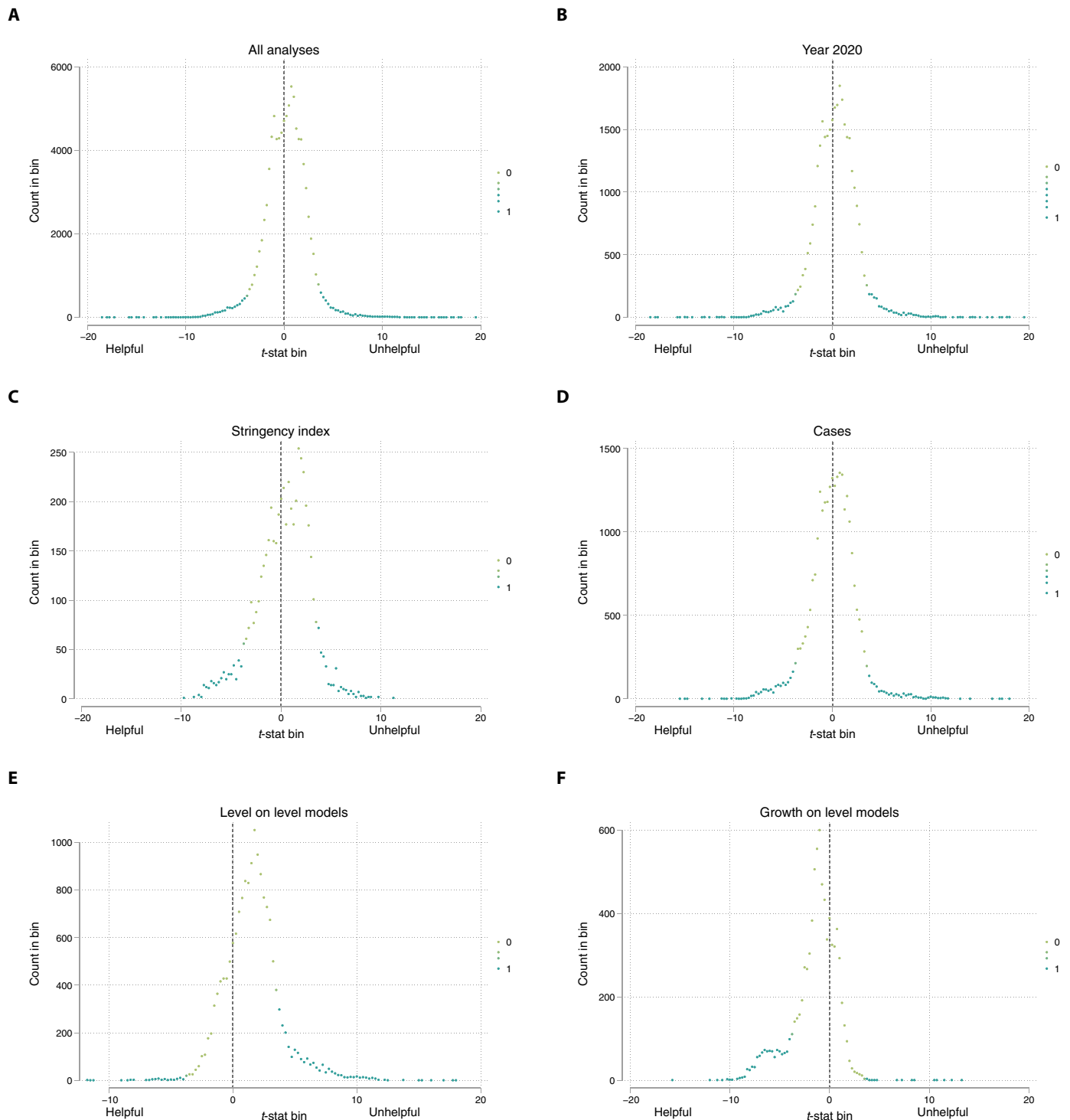


Fig. 3. Distribution of standardized effect sizes for all analyses and selected groups. The t -statistics are used as a proxy for the standardized effect size of different models. The count of models with t -statistics in each interval $[x, x + 0.25]$ for x ranging from -20 to $+20$ is shown in each panel (intervals with 0 models are not shown). The portion of the models that are significant using the false discovery rate is represented by the shade. In (A), for example, the t -statistic interval with the largest count of models is between 0.75 and 1.00. None of those models are significant. Panel (A) includes all models; (B) includes models for the year 2020; (C) includes models with the stringency index as the predictor; (D) includes models with cases as the outcome; (E) includes models designed to assess the relationship between the policy level and the outcome level (model 1); and panel (F) includes models designed to assess the relationship between the policy level and the change (growth) in the outcome (model 4).

Table 2. Five most consistent associations in the helpful and unhelpful directions. IHME, Institute for Health Metrics and Evaluation; NYT, the *New York Times*; Econ, *The Economist*; CSSE, Center for Systems Science and Engineering, the Johns Hopkins Covid-19 Dashboard.

Outcome	Government response	% Helpful/Unhelpful	% Significant
Most commonly helpful			
Infections	Income support	70.9	5.8
Infections	Econ support index	68.3	5.4
Excess deaths (NYT)	Debt relief	67.1	0.0
Infections	Debt relief	62.3	4.5
Excess deaths (Econ)	Econ support index	61.8	0.7
Most commonly helpful and significant			
Cases (IHME)	Government response index	46.8	12.4
Cases (IHME)	Stringency index	47.9	11.9
Infections	Government response index	52.6	11.0
Infections	School closure	49.6	10.8
Infections	Stringency index	49.9	10.3
Most commonly unhelpful			
Excess deaths (Econ)	Access to testing	81.2	3.7
Excess deaths (NYT)	School closure	78.5	0.2
Excess deaths (Econ)	Masking policies	76.2	3.5
Infections	Vaccine availability	75.7	12.3
Deaths (CSSE)	Vaccine availability	74.3	3.3
Most commonly unhelpful and significant			
Deaths (IHME)	Government response index	63.6	15.1
Deaths (IHME)	Stringency index	61.5	14.6
Deaths (CSSE)	Stringency index	63.1	13.6
Deaths (IHME)	Workplace closure	60.1	13.5
Deaths (CSSE)	Government response index	64.4	13.5

well-measured data collection platforms. The benefits of such platforms are demonstrated in the invaluable understanding of COVID-19 vaccine impacts from large registries in Israel and Qatar (24, 25). Large national prospective data platforms have been a long-lasting challenge in many countries, including the United States; local platforms, which may be easier to implement, could facilitate understanding at smaller scales. In the context of assessing government response, improved measurements of responses would be important. The OxCGRT is a critical resource, but a better understanding of implementation, enforcement, and compliance could further understand effect heterogeneity. Mitigating uncertainty due to flexible analytic design includes public registration of hypotheses in a public repository before analysis, like the process that precedes many randomized clinical trials (26).

The issue of subjecting government responses to experimentation is complicated. Trials of public health programs would yield extremely valuable information. Such trials may be thorny on practical or ethical grounds. This deserves further consideration, however, given the enormous stakes and inevitable trade-offs involved in responses such as mandatory school or business closures. A final important consideration that could improve the quality of evidence is keeping the issue at hand away from special and financial interests. The polarization that beset the scientific community has made asking and probing some questions difficult (27, 28). Keeping scientific

questions separate from public decisions could enable a spirit of greater collaboration around such issues.

The limitations of this study fall into four broad groups. First, it may be that a consistent signal—either of government response helpfulness or unhelpfulness—is contained in the data or models but not identified in this analysis. We encourage further probing of the results in our Shiny app. Second, the models are limited in their causal strength in the sense that a counterfactual to the policies implemented cannot be inferred. Two features temper this limitation: the use of leads such that the outcome is measured 2 weeks or a month following policy response and the use of fixed effects that assess “within country” associations and control for all time-invariant effects. While covariates with time-varying information on—say, health care capacity—may provide additional nuance, this information would be useful if it were available and comparable for all (or many) countries and time-varying at a daily or monthly level. We note that all our models examined short-term epidemic outcomes following policy responses (2 or 4 weeks), but that long-term outcomes remain an important but largely unexamined area of study. Third, country-level data hide more nuanced patterns that may be discernible in analyses of more granular data. Last, despite efforts to limit investigator choices, we made choices in the design of the study, and those may limit inferences. The data and models used in this analysis are open for other investigators to use, modify, or reassemble.

In sum, this comprehensive analysis of government responses and COVID-19 outcomes fails to yield clear inferences about government response impacts. This suggests that strong notions about the effectiveness or ineffectiveness of government responses are not backed by existing country-level data, and scientific modesty is warranted when learning from the responses to the COVID-19 pandemic.

METHODS

This analysis tests the relationships between government responses and COVID-19 outcomes. We test the extent to which COVID-19 outcomes improved or worsened following government responses. We recognize that the links between government responses and outcomes are mediated, for example, by the power of the government to enforce a response such as mask mandates. We also recognize that this approach cannot fully assess the counterfactual reality such as “what would outcomes have been had the government kept schools closed for longer?” Rather, we assess the observed relationship, implicitly assuming that if responses and outcomes go in opposite directions (for example, cases increase after easing mask requirements), this is generally consistent with the success of the government response. Conversely, if responses and outcomes go in the same direction (for example, cases increase after increasing mask requirements), this is evidence generally inconsistent with success.

The current paper presents the results of nearly 100,000 reasonable ways of assessing the relationship between government responses and COVID-19 outcomes. Government responses are represented as individual policies such as school closures, or as indices that aggregate the intensity and type of several policies. The rest of this section details the dimensions used in this analysis.

Government responses

The primary data source for government responses is the Oxford COVID-19 Government Response Tracker (OxCGRT). The OxCGRT recorded government responses daily in more than 180 countries using a standardized approach from publicly available sources such as news articles or government briefings. The OxCGRT recorded the official responses at the national level, not their implementation or enforcement. The complete details of the OxCGRT data-generating processes are publicly available (4, 5).

Government responses fall into three primary domains: containment and closure (such as school closures and restrictions on gatherings; eight ordinal variables), health system responses (such as contact tracing and mask mandates; six ordinal variables), and economic relief policies (income support and debt relief; two ordinal variables). Because government responses may work in concert or synergistically, the OxCGRT constructs composite indices that aggregate the individual responses. We use three composite indices: the “government response index” which pools all 16 government responses; the “stringency index” which pools the eight containment and closure variables and one health system response, and the “economic support index” which pools the two economic relief policies. We use a version of the indices that combines government responses for vaccinated and unvaccinated populations using a weighted average based on the portion of the population that is vaccinated. All 19 variables (16 individual variables and 3 indices) are available daily for the 181 countries for which we have outcomes data from 22 January 2020 to 31 December 2021.

COVID-19 outcomes

We use nine different outcome measures. We extracted two from the Johns Hopkins COVID-19 dashboard: daily confirmed COVID-19 cases and deaths (29). We extracted three outcomes from the Institute for Health Metrics and Evaluation (IHME): other estimates of daily COVID-19 cases and deaths, and daily estimates of infections (30, 31). The estimates of infections are modeled on the basis of age-specific infection fatality rates and age distribution of deaths. Last, we took weekly or monthly excess all-cause mortality from the *New York Times* (35 countries), the *Financial Times* (99 countries), the World Mortality Dataset (102 countries), and *The Economist* (181 countries) (32–35). A comparison of data sources for excess mortality is available elsewhere (36).

Statistical models

We use six statistical models. The models represent several patterns of relationships between government responses and COVID-19 outcomes. We chose models that broadly represent a stated expected impact of government response policies (such as, for example, models that assess a “flattening of the curve”), and models that capture historical patterns of public health efforts that succeeded in reducing infectious disease burden such as measles vaccination or polio elimination (8, 37, 38). (We test the models on a dataset of measles cases in the United States; see Plausibility Analysis below.) The formal models are presented below. Each model was estimated such that the coefficient on the *Policy* variable (β_1 below) would be negative if the government response was associated with reduced COVID-19 burden (we use “helpful” as shorthand for this relationship, without causal implication)

$$Y_{c(t+n)} = \beta_0 + \beta_1 Policy_{ct} + \mu^k X_{ct}^k + \delta_c + \lambda_t + \varepsilon_{ct} \quad (1)$$

$$Y_{c(t+n)} - Y_{ct} = \beta_0 + \beta_1 Policy_{ct} + \mu^k X_{ct}^k + \delta_c + \lambda_t + \varepsilon_{ct} \quad (2)$$

$$Y_{c(t+n)} - Y_{ct} = \beta_0 + \beta_1 [Policy_{ct} - Policy_{c(t-1)}] + \mu^k X_{ct}^k + \delta_c + \lambda_t + \varepsilon_{ct} \quad (3)$$

$$\frac{Y_{c(t+n)}}{Y_{c(t+n-1)}} = \beta_0 + \beta_1 Policy_{ct} + \mu^k X_{ct}^k + \delta_c + \lambda_t + \varepsilon_{ct} \quad (4)$$

$$\frac{Y_{c(t+n)}}{Y_{c(t+n-1)}} - \frac{Y_{ct}}{Y_{c(t-1)}} = \beta_0 + \beta_1 Policy_{ct} + \mu^k X_{ct}^k + \delta_c + \lambda_t + \varepsilon_{ct} \quad (5)$$

$$\frac{Y_{c(t+n)}}{Y_{c(t+n-1)}} - \frac{Y_{ct}}{Y_{c(t-1)}} = \beta_0 + \beta_1 [Policy_{ct} - Policy_{c(t-1)}] + \mu^k X_{ct}^k + \delta_c + \lambda_t + \varepsilon_{ct} \quad (6)$$

In each model, Y is the outcome of interest, c indexes a country, t indexes the observation time, and n represents the duration between the government response and outcome observation (2 or 4 weeks/1 month). $\mu^k X_{ct}^k$ represents a k -wide matrix of covariates, δ_c represents country fixed effects, and λ_t are time fixed effects. Country fixed effects remove all time-invariant differences between countries, such that the effects are estimated within the country. Time fixed effects control for temporal trends shared among all countries. All models were estimated using ordinary least squares.

To facilitate an intuitive understanding of the model, we use a concrete example with COVID-19 deaths as the outcome and stringency index as the policy. Model 1 tests the extent to which higher stringency is associated with fewer COVID-19 deaths 2 or 4 weeks later. Model 2 tests the extent to which higher stringency is associated with fewer COVID-19 deaths 2 or 4 weeks later, compared with the day of observation. Model 3 tests the extent to which increasing stringency is associated with fewer COVID-19 deaths 2 or 4 weeks later, compared with the day of the increase (or, conversely, decreasing stringency associated with more COVID-19 deaths). Model 4 tests the extent to which higher stringency is associated with a lower growth rate of COVID-19 deaths 2 or 4 weeks later. Model 5 tests the extent to which higher stringency is associated with a growth rate of COVID-19 deaths 2 or 4 weeks later that is lower than the day of observation. Model 6 tests the extent to which increasing stringency is associated with a growth rate of COVID-19 deaths 2 or 4 weeks later that is lower than the day of the increase.

Analytic and data variations

We analyzed a total of 99,736 models. The data were analyzed at the daily, weekly, or monthly level of aggregation, allowing for smoothing of idiosyncratic variation in daily data. Covariates included the number of borders (“island effect”), the portion of the population over age 60, the total fertility rate (to capture age structure), and the daily mobility (percent of baseline) from Google, obtained from IHME. Covariates were either included or excluded as a bloc. Country and time fixed effects are commonly used in econometric models to control for time-invariant between-country differences and shared time patterns. Including country fixed effects, in particular, yields a pooled within-country association. For example, changes in the U.S. stringency index are assessed in relation to COVID-19 growth rates in the United States. Models 1 to 3 are analyzed using totals or per-capita outcomes (per-capita outcomes are identical to totals with growth models 4 to 6). To prevent population differences from overwhelming estimates, models with total outcomes include fixed effects. Last, we analyze all models over three time periods: the early pandemic (January 2020 to June 2020); the first year (all of 2020); and the first 2 years (2020 to 2021).

Standard errors are clustered by country in all analyses. Statistical significance is assessed using a false discovery rate of 0.05 (23).

The analytic code is provided along with the analysis. In addition, the entire set of 99,736 model results can be explored using a Shiny app at <https://eranbendavid.shinyapps.io/CovidGovPolicies/>.

Plausibility analysis

With such a large number of models, we conducted a separate analysis to test whether the models would identify effects, should ones exist within the data. Specifically, we use our approach to study the introduction of measles vaccination in the United States, widely regarded as a public health success (39, 40). Following the licensing of the measles vaccine in 1963/1964, the number of reported measles cases dropped from approximately 400,000 annually in the 5 years before the licensing to 30,000 annually in the 5 years after (41, 42). We created a synthetic dataset with measles cases proportional to the state population between 1954 and 1990 and assigned a vaccination adoption year to each state between 1964 and 1967 with an effect size of around 93% (equivalent to the case rate decline in the entire United States). We thus construct a dataset with different units (states), an efficacious policy intervention (vaccination), and

different policy onset (1964 to 1967). We then applied our models to this dataset, with cases as the only outcome, scheduled vaccination onset as the main policy predictor, and fixed effects as with the main analysis.

Supplementary Materials

This PDF file includes:

Fig. S1
Table S1
Legend for data file S1
References

Other Supplementary Material for this manuscript includes the following:

Data file S1

REFERENCES AND NOTES

1. N. A. Christakis, *Apollo's Arrow: The Profound and Enduring Impact of Coronavirus on the Way We Live* (Little, Brown and Company, 2020).
2. A. A. Rabaan, S. H. Al-Ahmed, S. Haque, R. Sah, R. Tiwari, Y. S. Malik, K. Dhama, M. I. Yatoo, D. K. Bonilla-Aldana, A. J. Rodriguez-Morales, SARS-CoV-2, SARS-CoV, and MERS-CoV: A comparative overview. *Infect. Med.* **28**, 174–184 (2020).
3. United Nations Department of Economic and Social Affairs, Population Division. World Population Prospects 2022: Summary of Results (United Nations Department of Economic and Social Affairs, 2022); www.un.org/development/desa/pd/content/World-Population-Prospects-2022.
4. T. Hale, N. Angrist, R. Goldszmidt, B. Kira, A. Petherick, T. Phillips, S. Webster, E. Cameron-Blake, L. Hallas, S. Majumdar, H. Tatlow, A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nat. Hum. Behav.* **5**, 529–538 (2021).
5. T. Hale, N. Angrist, B. Kira, A. Petherick, T. Phillips, S. Webster, Variation in government responses to COVID-19, Blavatnik School of Government working paper (2023); <https://ora.ox.ac.uk/objects/uuid:0ab73a02-ca18-4e1f-a41b-cfea2d30e81>.
6. R. J. Hatchett, C. E. Mecher, M. Lipsitch, Public health interventions and epidemic intensity during the 1918 influenza pandemic. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7582–7587 (2007).
7. Political Declaration on HIV and AIDS: On the Fast Track to Accelerating the Fight against HIV and to Ending the AIDS Epidemic by 2030; www.unaids.org/en/resources/documents/2016/2016-political-declaration-HIV-AIDS.
8. Report 9 - Impact of Non-Pharmaceutical Interventions (NPIs) to Reduce COVID-19 Mortality and Healthcare Demand, Imperial College London; <https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/disease-areas/covid-19/report-9-impact-of-npis-on-covid-19/>.
9. S. Flaxman, S. Mishra, A. Gandy, H. J. T. Unwin, T. A. Mellan, H. Coupland, C. Whittaker, H. Zhu, T. Berah, J. W. Eaton, M. Monod, A. C. Ghani, C. A. Donnelly, S. Riley, M. A. C. Vollmer, N. M. Ferguson, L. C. Okell, S. Bhatt, Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* **584**, 257–261 (2020).
10. D. H. P. Balog-Way, K. A. McComas, COVID-19: Reflections on trust, tradeoffs, and preparedness. *J. Risk Res.* **23**, 838–848 (2020).
11. J. Herby, L. Jonung, S. H. Hanke, A literature review and meta-analysis of the effects of lockdowns on COVID-19 mortality, in *Studies in Applied Economics 200* (The Johns Hopkins Institute for Applied Economics, Global Health, and the Study of Business Enterprise).
12. S. Talic, S. Shah, H. Wild, D. Gasevic, A. Maharaj, Z. Ademi, X. Li, W. Xu, I. Mesa-Eguiaagaray, J. Rostron, E. Theodoratou, X. Zhang, A. Motee, D. Liew, D. Illic, Effectiveness of public health measures in reducing the incidence of covid-19, SARS-CoV-2 transmission, and covid-19 mortality: Systematic review and meta-analysis. *BMJ* **375**, e068302 (2021).
13. E. Bendavid, B. Mulaney, N. Sood, S. Shah, R. Bromley-Dulfano, C. Lai, Z. Weissberg, R. Saavedra-Walker, J. Tedrow, A. Bogan, T. Kupiec, D. Eichner, R. Gupta, J. P. A. Ioannidis, J. Bhattacharya, COVID-19 antibody seroprevalence in Santa Clara County, California. *Int. J. Epidemiol.* **50**, 410–419 (2021).
14. A. Atkeson, K. Kopecky, T. Zha, “Four stylized facts about COVID-19” (Working Paper 27719, National Bureau of Economic Research, 2020); <https://doi.org/10.3386/w27719>.
15. S. Hsiang, D. Allen, S. Annan-Phan, K. Bell, I. Bolliger, T. Chong, H. Druckenmiller, L. Y. Huang, A. Hultgren, E. Krasovich, P. Lau, J. Lee, E. Rolf, J. Tseng, T. Wu, The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature* **584**, 262–267 (2020).
16. P. Bagus, J. A. Peña-Ramos, A. Sánchez-Bayón, COVID-19 and the Political Economy of Mass Hysteria. *Int. J. Environ. Res. Public Health* **18**, 1376 (2021).

17. T. J. Bollyky, E. Castro, A. Y. Aravkin, K. Bhangdia, J. Dalos, E. N. Hullah, S. Kiernan, A. Lastuka, T. A. McHugh, S. M. Ostroff, P. Zheng, H. T. Chaudhry, E. Ruggiero, I. Turilli, C. Adolph, J. O. Amlag, B. Bang-Jensen, R. M. Barber, A. Carter, C. Chang, R. M. Cogen, J. K. Collins, X. Dai, W. J. Dangel, C. Dapper, A. Deen, A. Eastus, M. Erickson, T. Fedosseeva, A. D. Flaxman, N. Fullman, J. R. Giles, G. Guo, S. I. Hay, J. He, M. Helak, B. M. Huntley, V. C. Iannucci, K. E. Kinzel, K. E. LeGrand, B. Magistro, A. H. Mokdad, H. Nassereldine, Y. Oztun, M. Pasovic, D. M. Pigott, R. C. Reiner, G. Reinke, A. E. Schumacher, E. Serieux, E. E. Spurlock, C. E. Troeger, A. T. Vo, T. Vos, R. Walcott, S. Yazdani, C. J. L. Murray, J. L. Dieleman, Assessing COVID-19 pandemic policies and behaviours and their economic and educational trade-offs across US states from Jan 1, 2020, to July 31, 2022: An observational analysis. *Lancet* **401**, 1341–1360 (2023).
18. S. Steegen, F. Tuerlinckx, A. Gelman, W. Vanpaemel, Increasing Transparency Through a Multiverse Analysis. *Perspect Psychol Sci* **11**, 702–712 (2016).
19. J. A. Harder, The multiverse of methods: Extending the multiverse analysis to address data-collection decisions. *Perspect. Psychol. Sci.* **15**, 1158–1177 (2020).
20. X. Sala-i-Martin, I just ran two million regressions. *Am. Econ. Rev.* **87**, 178–183 (1997).
21. C. J. Patel, B. Burford, J. P. Ioannidis, Assessment of vibration of effects due to model specification can demonstrate the instability of observational associations. *J. Clin. Epidemiol.* **68**, 1046–1058 (2015).
22. T. Hale, A. J. Hale, B. Kira, A. Petherick, T. Phillips, D. Sridhar, R. N. Thompson, S. Webster, N. Angrist, Global assessment of the relationship between government response measures and COVID-19 deaths. medRxiv 2020.07.04.20145334 (2020); <https://doi.org/10.1101/2020.07.04.20145334>.
23. Y. Benjamini, D. Yekutieli, The control of the false discovery rate in multiple testing under dependency. *Ann. Stat.* **29**, 1165–1188 (2001).
24. H. N. Altarawneh, H. Chemaitelly, H. H. Ayoub, P. Tang, M. R. Hasan, H. M. Yassine, H. A. Al-Khatib, M. K. Smatti, P. Coyle, Z. Al-Kanaani, E. Al-Kuwari, A. Jeremijenko, A. H. Kaleeckal, A. N. Latif, R. M. Shaik, H. F. Abdul-Rahim, G. K. Nasrallah, M. G. Al-Kuwari, A. A. Butt, H. E. Al-Romaihi, M. H. Al-Thani, A. Al-Khal, R. Bertollini, L. J. Abu-Raddad, Effects of previous infection and vaccination on symptomatic Omicron infections. *N. Engl. J. Med.* **387**, 21–34 (2022).
25. Y. Goldberg, M. Mandel, Y. M. Bar-On, O. Bodenheimer, L. Freedman, E. J. Haas, R. Milo, S. Alroy-Preis, N. Ash, A. Huppert, Waning immunity after the BNT162b2 vaccine in Israel. *N. Engl. J. Med.* **385**, e85 (2021).
26. F. Naudet, C. J. Patel, N. J. DeVito, G. L. Goff, I. A. Cristea, A. Braillon, S. Hoffmann, Improving the transparency and reliability of observational studies through registration. *BMJ* **384**, e076123 (2024).
27. S. L. Myers, Is spreading medical misinformation a doctor's free speech right? *The New York Times*, 30 November 2022; www.nytimes.com/2022/11/30/technology/medical-misinformation-covid-free-speech.html.
28. P. S. Hart, S. Chinn, S. Soroka, Politicization and polarization in COVID-19 News coverage. *Sci. Commun.* **42**, 679–697 (2020).
29. E. Dong, H. Du, L. Gardner, An interactive web-based dashboard to track COVID-19 in real time. *Lancet Infect. Dis.* **20**, 533–534 (2020).
30. COVID-19 estimate downloads, *Institute for Health Metrics and Evaluation* (2020); www.healthdata.org/covid/data-downloads.
31. R. C. Reiner, R. M. Barber, J. K. Collins, P. Zheng, C. Adolph, J. Albright, C. M. Antony, A. Y. Aravkin, S. D. Bachmeier, B. Bang-Jensen, M. S. Bannick, S. Bloom, A. Carter, E. Castro, K. Causey, S. Chakrabarti, F. J. Charlson, R. M. Cogen, E. Combs, X. Dai, W. J. Dangel, L. Earl, S. B. Ewald, M. Ezalarab, A. J. Ferrari, A. Flaxman, J. J. Frostad, N. Fullman, E. Gakidou, J. Gallagher, S. D. Glenn, E. A. Goosmann, J. He, N. J. Henry, E. N. Hullah, B. Hurst, C. Johanns, P. J. Kendrick, A. Khemani, S. L. Larson, A. Lazzar-Atwood, K. E. LeGrand, H. Lescinsky, A. Lindstrom, E. Linebarger, R. Lozano, R. Ma, J. Månsson, B. Magistro, A. M. M. Herrera, L. B. Marczak, M. K. Miller-Petrie, A. H. Mokdad, J. D. Morgan, P. Naik, C. M. Odell, J. K. O'Halloran, A. E. Osgood-Zimmerman, S. M. Ostroff, M. Pasovic, L. Penberthy, G. Phipps, D. M. Pigott, I. Pollock, R. E. Ramshaw, S. B. Redford, G. Reinke, S. Rolfe, D. F. Santomauro, J. R. Shackleton, D. H. Shaw, B. S. Sheena, A. Sholokhov, R. J. D. Sorensen, G. Sparks, E. E. Spurlock, M. L. Subart, R. Syailendrawati, A. E. Torre, C. E. Troeger, T. Vos, A. Watson, S. Watson, K. E. Wiens, L. Woyczynski, L. Xu, J. Zhang, S. I. Hay, S. S. Lim, C. J. L. Murray, IHME COVID-19 FORECASTING TEAM, MODELING COVID-19 scenarios for the United States. *Nat. Med.* **27**, 94–105 (2021).
32. A. Karlinsky, D. Kobak, Tracking excess mortality across countries during the COVID-19 pandemic with the World Mortality Dataset. *eLife* **10**, e69336 (2021).
33. covid-19-excess-deaths nytimes github, *GitHub* (2023); <https://github.com/nytimes/covid-19-data>.
34. covid-19-excess-deaths Financial-Times github, *GitHub* (2022); <https://github.com/Financial-Times/coronavirus-excess-mortality-data>.
35. covid-19-excess-deaths The Economist github, *GitHub* (2023); <https://github.com/TheEconomist/covid-19-excess-deaths-tracker>.
36. M. Levitt, F. Zonta, J. P. A. Ioannidis, Comparison of pandemic excess mortality in 2020–2021 across different empirical calculations. *Environ. Res.* **213**, 113754 (2022).
37. J. A. Lewnard, N. C. Lo, Scientific and ethical basis for social-distancing interventions against COVID-19. *Lancet Infect. Dis.* **20**, 631–633 (2020).
38. M. E. Wharton, Measles elimination in the United States. *J. Infect Dis* **189**, S1–S3 (2004).
39. A. B. Bloch, W. A. Orenstein, H. C. Stetler, S. G. Wassilak, R. W. Amler, K. J. Bart, C. D. Kirby, A. R. Hinman, Health impact of measles vaccination in the United States. *Pediatrics* **76**, 524–532 (1985).
40. D. J. Sencer, H. B. Dull, A. D. Langmuir, Epidemiologic basis for eradication of measles in 1967. *Public Health Rep.* **82**, 253–256 (1967).
41. Reported cases and deaths of measles, *Our World in Data*; <https://ourworldindata.org/grapher/measles-cases-and-death>.
42. Centers for Disease Control and Prevention, MMWR Summary of Notifiable Diseases, United States (Centers for Disease Control and Prevention, 1993); www.cdc.gov/mmwr/preview/mmwrhtml/00035381.htm.
43. B. Krishnamachari, A. Morris, D. Zastrow, A. Dsida, B. Harper, A. J. Santella, The role of mask mandates, stay at home orders and school closure in curbing the COVID-19 pandemic prior to vaccination. *Am. J. Infect. Control* **49**, 1036–1042 (2021).
44. N. Dreher, Z. Spiera, F. M. McAuley, L. Kuohn, J. R. Durbin, N. F. Marayati, M. Ali, A. Y. Li, T. C. Hannah, A. Gometz, J. Kostman, T. F. Choudhri, Policy interventions, social distancing, and SARS-CoV-2 transmission in the United States: A retrospective state-level analysis. *Am. J. Med. Sci.* **361**, 575–584 (2021).
45. A. Tobias, Evaluation of the lockdowns for the SARS-CoV-2 epidemic in Italy and Spain after one month follow up. *Sci. Total Environ.* **725**, 138539 (2020).
46. J. Vlachos, E. Hertegård, H. B. Svaleryd, The effects of school closures on SARS-CoV-2 among parents and teachers. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2020834118 (2021).
47. J. Gibson, Government mandated lockdowns do not reduce Covid-19 deaths: Implications for evaluating the stringent New Zealand response. *New Zealand Economic Papers* **56**, 17–28 (2022).
48. M. J. Siedner, G. Harling, Z. Reynolds, R. F. Gilbert, S. Haneuse, A. S. Venkataramani, A. C. Tsai, Social distancing to slow the US COVID-19 epidemic: Longitudinal pretest-posttest comparison group study. *PLoS Med.* **17**, e1003244 (2020).
49. X. Liu, X. Xu, G. Li, X. Xu, Y. Sun, F. Wang, X. Shi, X. Li, G. Xie, L. Zhang, Differential impact of non-pharmaceutical public health interventions on COVID-19 epidemics in the United States. *BMC Public Health* **21**, 965 (2021).
50. V. Alfano, S. Ercolano, The efficacy of lockdown against COVID-19: A cross-country panel analysis. *Appl. Health Econ. Health Policy* **18**, 509–517 (2020).
51. W. M. Thayer, M. Z. Hasan, P. Sankhla, S. Gupta, An interrupted time series analysis of the lockdown policies in India: A national-level analysis of COVID-19 incidence. *Health Policy Plan.* **36**, 620–629 (2021).
52. G. Guzzetta, F. Riccardo, V. Marziano, P. Poletti, F. Trentini, A. Bella, X. Andrianou, M. Del Manso, M. Fabiani, S. Bellino, S. Boros, A. M. Urdiales, M. F. Vescio, A. Piccoli, COVID-19 Working Group 2, S. Brusaferro, G. Rezza, P. Pezzotti, M. Ajelli, S. Merler, Impact of a nationwide lockdown on SARS-CoV-2 transmissibility, Italy. *Emerg. Infect. Dis.* **27**, 267–270 (2021).
53. Y. Liu, C. Morgenstern, J. Kelly, R. Lowe, J. Munday, C. J. Villabona-Arenas, H. Gibbs, C. A. B. Pearson, C. Prem, Q. J. Leclerc, S. R. Meakin, W. J. Edmunds, C. I. Jarvis, A. Gimma, S. Funk, M. Quaife, T. W. Russell, J. C. Emory, S. Abbott, J. Hellewell, D. C. Tully, R. M. G. J. Houben, K. O'Reilly, G. R. Gore-Langton, A. J. Kucharski, M. Auzenberg, B. J. Quilty, T. Jombart, A. Rosello, O. Brady, K. E. Atkins, K. van Zandvoort, J. W. Rudge, A. Endo, K. Abbas, F. Y. Sun, S. R. Procter, S. Clifford, A. M. Foss, N. G. Davies, Y.-W. D. Chan, C. Diamond, R. C. Barnard, R. M. Eggo, A. K. Deol, E. S. Nightingale, D. Simons, K. Sherratt, G. Medley, S. Hué, G. M. Knight, S. Flasche, N. I. Bosse, P. Klepac, M. Jit, The impact of non-pharmaceutical interventions on SARS-CoV-2 transmission across 130 countries and territories. *BMC Med.* **19**, 40 (2021).

Acknowledgments

Funding: The authors acknowledge that they received no funding in support of this research.

Author contributions: E.B. and C.J.P. conceived the manuscript, conducted the analyses, interpreted the results, drafted the manuscript, and carried out all revisions.

Competing interests: The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. The analytic code is provided along with the analysis. In addition, the entire set of 99,736 model results can be explored using a Shiny app at <https://eranbendavid.shinyapps.io/CovidGovPolicies/>.

Submitted 20 November 2023

Accepted 3 May 2024

Published 5 June 2024

10.1126/sciadv.adn0671