On the Manipulation of Information by Governments*

Ariel Karlinsky† and Moses Shayo‡

June 2, 2024

Abstract

Governmental information manipulation has been hard to measure and study systematically. We hand-collect data from official and unofficial sources in 134 countries to estimate misreporting of Covid mortality during 2020-21. Deception is remarkably widespread as we find that between 45%–55% of governments misreported the number of deaths. The lion's share of misreporting cannot be attributed to a country's capacity to accurately diagnose and report deaths. Contrary to some theoretical expectations, there is little evidence of governments exaggerating the severity of the pandemic but many have significantly downplayed it. Misreporting is higher where governments face few social and institutional constraints, in countries holding elections, and in countries with a communist legacy.

*Data last updated March 2024. We thank Ekatrina Borisova, Yehonatan Givati, Orit Kedar, Dmitry Kobak, Ro'ee Levy, Yotam Margalit, Lior Sheffer, Alois Stutzer, Assaf Zimring and many seminar and conference participants for comments and suggestions.

†The Hebrew University of Jerusalem. ariel.karlinsky@mail.huji.ac.il

‡The Hebrew University of Jerusalem and King's College London. mshayo@huji.ac.il

1 Introduction

When facing crises, governments may try to manipulate public information. Qualitatively, this can take two different directions. The first is to exaggerate the gravity of the situation in order to induce a "flight to safety" (Quattrone and Tversky, 1988; Bisbee and Honig, 2022); to rally the population around the flag by presenting the crisis as a national (preferably external) threat (Mueller, 1970); to instill fear in order to suppress dissent (Young, 2019); or to create an atmosphere of emergency that legitimizes repressive measures and restrictions on civil liberties (Sturm 1949; Davis and Silver 2004; Alsan et al. 2023; but see Lowande and Rogowski 2021). In countries that rely on foreign aid, exaggerating hardship might also be used in an attempt to attract disaster relief (Becerra et al., 2014). The second form of manipulation is to downplay, conceal, or sugarcoat the hardship in an attempt to project competence (Edmond, 2013; Guriev and Treisman, 2015; Gehlbach et al., 2016; Rozenas and Stukal, 2019); curtail public distrust in the government (Honryo and Yano 2021); reduce public opposition to corrupt or undemocratic behavior (De Vries and Solaz, 2017; Frederiksen, 2022), and enhance the leader's image both domestically and globally (Guriev and Treisman, 2022). Governments whose response to the crisis has been inadequate might also seek to downplay the gravity of the situation in order to justify their inaction. Beyond the direction of information manipulation, its magnitude might also vary across governments, either due to differences in incentives and motivations for obfuscation, or different social and institutional constraints on their ability to do so.

While it is easy to motivate why we need to understand which governments manipulate information, to what extent, and in what direction, answering these questions is very hard. Journalists, scholars and NGOs sometimes expose idiosyncratic cases and techniques of information manipulation. However, a systematic evaluation of governmental information manipulation requires (a) a quantifiable gap between governmental reports and the truth; and (b) governments confronting comparable issues. One approach, proposed in political science, is to examine *how much* information is released by governments (Hollyer et al., 2011). Another approach, proposed in economics, uses night time light as a less manipulable measure of economic activity that can be compared to GDP figures (Martinez, 2022). In this paper we develop an alternative approach to directly evaluate the reliability of official data, while still relying on observable and comparable governmental

actions. The onset of COVID-19 (henceforth "Covid") in 2020 presented governments around the world with a similar crisis at roughly the same time. Studying information disseminated during the first two years of the pandemic, we can assess the extent and direction of information manipulation by governments across the globe.

To achieve this, we construct a novel dataset of all-cause-mortality before and during the Covid pandemic. For every country and territory in the world, we conduct an exhaustive search of the relevant official institutions: national statistical offices, population registries, ministries of health, etc. Where we could not locate the required information, we queried the relevant institutions (using email, contact forms, social media, or right-to-information requests). In some countries, we were able to obtain data from local journalists also employing right-to-information requests from national or sub-national institutions. In total, we managed to compile the required data for 134 countries. Our sample represents 93% of countries with at least 75% Death Registration Completeness (a measure of the country's capacity to count and report deaths, explained in section 4.2). In addition, we have partial data for Bangladesh, China and Indonesia.

For each country we compute two measures: (1) the number of officially-reported Covid deaths; and (2) the excess number of deaths in the country during the pandemic. Excess deaths—defined as the number of deaths above and beyond expectation—is a well-established method for assessing mortality during epidemics (Farr, 1885; Langmuir, 1976; Viboud et al., 2016), natural disasters (Milken Institute, 2018) and human-made disasters such as famine, conflict and genocide (de Walque and Verwimp, 2010; Grada, 2013; Ekamper et al., 2017; Naumenko, 2021). There is wide consensus that excess deaths paint the most accurate and objective picture of the death toll from Covid, as it transcends differences in testing and cause of death certification between countries; and spillovers from the pandemic to other causes of death are small.[1] We note that, given the importance of these data for health and public policy, we have made the data publicly available in real time, and they formed the basis for analysis and policy by the WHO and others (Msemburi et al., 2023; The Economist, 2021).

With our estimates of excess deaths at hand, we can then use officially-reported Covid deaths to reveal the existence, direction, and magnitude of information manipulation. The logic underly-

[1]For medicine and public health see Beaney et al. (2020); Riffe et al. (2021); Whittaker et al. (2021) and in the social sciences see Boka and Wainer (2020); Aron and Muellbauer (2020); Alsan et al. (2021).

ing our approach is simple but, we believe, compelling. Following the outbreak of the pandemic, officially-reported Covid deaths (measure (1) above) attracted most of the attention of journalists, decision-makers, researchers, and the public at large. However, scattered-but-persistent indications suggest that certain reported numbers of Covid cases and deaths were manipulated and, furthermore, that politicians in both democratic and autocratic regimes have urged state institutions to report distorted figures. In contrast, all-cause mortality data—which we use to estimate excess deaths (measure (2))—are published under a very different set of incentives and constraints. Mortality records have been collected and used by the relevant state authorities for a long time, regardless of Covid. The bureaucratic apparatus routinely collects and reports these data, and has continued to do so during the pandemic. Moreover, all-cause mortality is far easier to ascertain than Covid deaths. When someone dies, many people know about it (relatives, neighbors, colleagues, medical professionals, priests, funeral directors, undertakers). All-cause mortality data is essentially a simple count of these instances. In contrast, a Covid death involves an attribution of cause of death. This is far harder to gauge and easier to manipulate. In India, for example, crematorium workers were reportedly told to record the cause of many Covid deaths as "Sickness". This means a death has been recorded, but the cause of death was not properly registered (Gettleman et al., 2021).

Our analysis yields five main findings:

1. About 45% of the countries in our sample have accurately reported Covid deaths. Between 45%-55% of countries have misreported the number of deaths.

2. There is very little evidence of governments exaggerating the severity of the pandemic as suggested by the first set of mechanisms outlined above. We find no evidence for death overreporting in countries experiencing more demonstrations or challenges to state legitimacy in the year preceding the pandemic. This is consistent with the difficulty to either portray the pandemic as an outgroup-related threat that can be harnessed to rally the population around the flag, or to use Covid deaths to signal the government's repressive capacity. We similarly find no evidence for overreporting in countries that are more reliant on foreign aid, or in countries with an exceptionally severe initial response to the pandemic.

3. We find robust and widespread evidence of governments downplaying the impact of the pan-

demic, beyond what can be explained by state capacity to count, diagnose, and register deaths.

4. The extent of underreporting can best be explained by constraints on the government's ability to effectively manipulate information. We find less evidence relating it to cultural factors and to macroeconomic incentives to underreport. The most important set of constraints appear to be institutional constraints on the executive. Additionally, the severity of under-reporting drops significantly the more sophisticated is the domestic audience, as measured by education levels and internet access. There is also evidence that misreporting is higher in countries holding elections during the pandemic.

5. We find a strikingly higher level of misreporting in countries with a communist legacy, even when holding constant current measures of institutional and social constraints on the executive.

Our measure of information manipulation has three valuable features: (i) it is comparable across countries; (ii) it is based on observable governmental actions regarding a central policy issue; and (iii) it covers a large and diverse set of countries. Our approach is similar in spirit to Olken (2007), who uses the gap between reported and actual expenditure on infrastructure to measure corruption, and Martinez (2022), who studies the relationship between published economic growth (GDP) figures and the economic growth implied by the increase in Night Time Light coverage. Consistent with our results, Martinez finds that autocracies overstate their economic growth by more than a third. With regards to Covid, previous studies documented anomalies in reported figures such as under-dispersion (Kobak, 2022) and deviations from Benford's law (Kilani, 2021). Consistent with our results, such anomalies were especially common in authoritarian regimes.

We begin in Section 2, with qualitative evidence of Covid-related information manipulation. Sections 3–4 outline our data collection strategy and report our estimates of misreporting across countries. We also analyze the relation of misreporting to state capacity to diagnose and report mortality. Section 5 examines the potential of contemporary social, economic, cultural and institutional factors to explain variation in misreporting, as well as prominent historical factors. Section 6 offers concluding remarks.

2 Qualitative Evidence

We begin by documenting prominent reported cases of governments manipulating Covid-related information. These reports complement our quantitative finding, that inaccurate Covid reporting cannot be attributed solely to limited state capacity to diagnose and count Covid deaths.

The BBC (2020a) reported data leaked from Iran, showing that the regime maintained two separate data series for the number of daily Covid deaths from January to July 20 2020. The data released to the public showed total Covid deaths at approximately 14,400, while deaths in the unpublished internal (leaked) data was more than 42,000.

Heifez and Liapeika (2021) obtained data from medical facilities in Minsk (Belarus) on detected Covid cases from June to November 2020. This information was highly detailed, included patient level data and was verified by sources inside the medical facilities. The number of cases reported by medical facilities in Minsk alone were higher than those reported by the Belarus government for the entire country: a total of almost 102,000 cases in Minsk City compared to fewer than 90,000 officially-reported cases.

Romero (2020) reported leaked data from the Nicaraguan Ministry of Health. While at the time health facilities reported to the ministry more than 9,600 confirmed Covid cases in the country, the ministry publicly reported less than 3,500 Covid cases.

The above cases seem to indicate both the will and the ability to manipulate Covid information. Importantly, however, there are also reports of less successful attempts to manipulate information. Ventura and Rocha (2021) compile speeches, actions and statutes by President Jair Bolsonaro and other officials in the Brazilian executive branch. Two days after Bolsonaro declared that the number of reported Covid deaths was inflated ("if you die of anything, [they say] it's covid"), his newly-appointed Secretary of Science, Technology and Strategic Input indicated an impending downward revision to Covid deaths. On the same day, the official Covid dashboard went offline and returned to show only the last 24 hours of data, removing all the cumulative data.[2] However, this was subsequently rectified following pressure from the judiciary, civil society and the media. The proposed downward revision to Covid deaths did not materialize either.

In the United States, Jankowicz (2020) and Mendez (2021) reported that the Trump admin-

[2]See also reports from Biller and Savarese (2020); Brum (2020); La Presna Latina (2021).

istration pressed the Centers for Disease Control (CDC) to revise the reported number of Covid cases and deaths downward. In a recording from March 2020 released by Bob Woodward, President Trump said "I wanted to always play it [Covid] down... I still like playing it down, because I don't want to create a panic" (Costa and Rucker, 2020). Notwithstanding such motivation, however, our data below shows far less evidence of misreporting in Brazil and in the United States than in Iran, Belarus and Nicaragua.

3 Data Sources

We hand-collect all-cause mortality data from 134 countries and territories (henceforth: countries) during 2015-2021 at weekly, monthly or annual frequency.[3] Our main sources are National Statistics Offices, Population Registries, Ministries of Health and Ministries of Public Health. We have already made much of the data available as the World Mortality Dataset (WMD, see Karlinsky and Kobak 2021 for a detailed account of the collection strategy). See Appendix A for details on data and estimates not included in WMD. Some countries, mostly in Africa, lack the infrastructure to record and publish vital registration data (Mikkelsen et al., 2015; UNSD, 2021; Karlinsky, 2021) and thus we could not include them.[4]

Officially-reported Covid deaths are obtained from the World Health Organization (2020), the Johns Hopkins University database, or directly from the governments of countries not in these databases. Measures of the state's mortality-counting capacity—in particular the completeness of death registration and the share of well-certified death registrations—are from Karlinsky (2021) and Vos et al. (2020). See Appendix Table D1 for additional data sources.

[3]For some countries, data begins later: Algeria (data from 2018), Bhutan (2019), Chile (2016), Iraq (2016), Palestine (2018), Peru (2017), Saint Kitts and Nevis (2016), Transnistria (2016), United Arab Emirates (2018) and Vietnam (2018).

[4]In many countries in Africa, available mortality data are irregular survey-based-estimates, which are not suited to study excess mortality which requires vital registration data for both pre-pandemic and during-pandemic period. In our meeting with a United Nations Statistical Division official in 2021, it was estimated that even in a decade we will probably not know what happened in Africa in 2020-2021.

4 Misreporting Covid

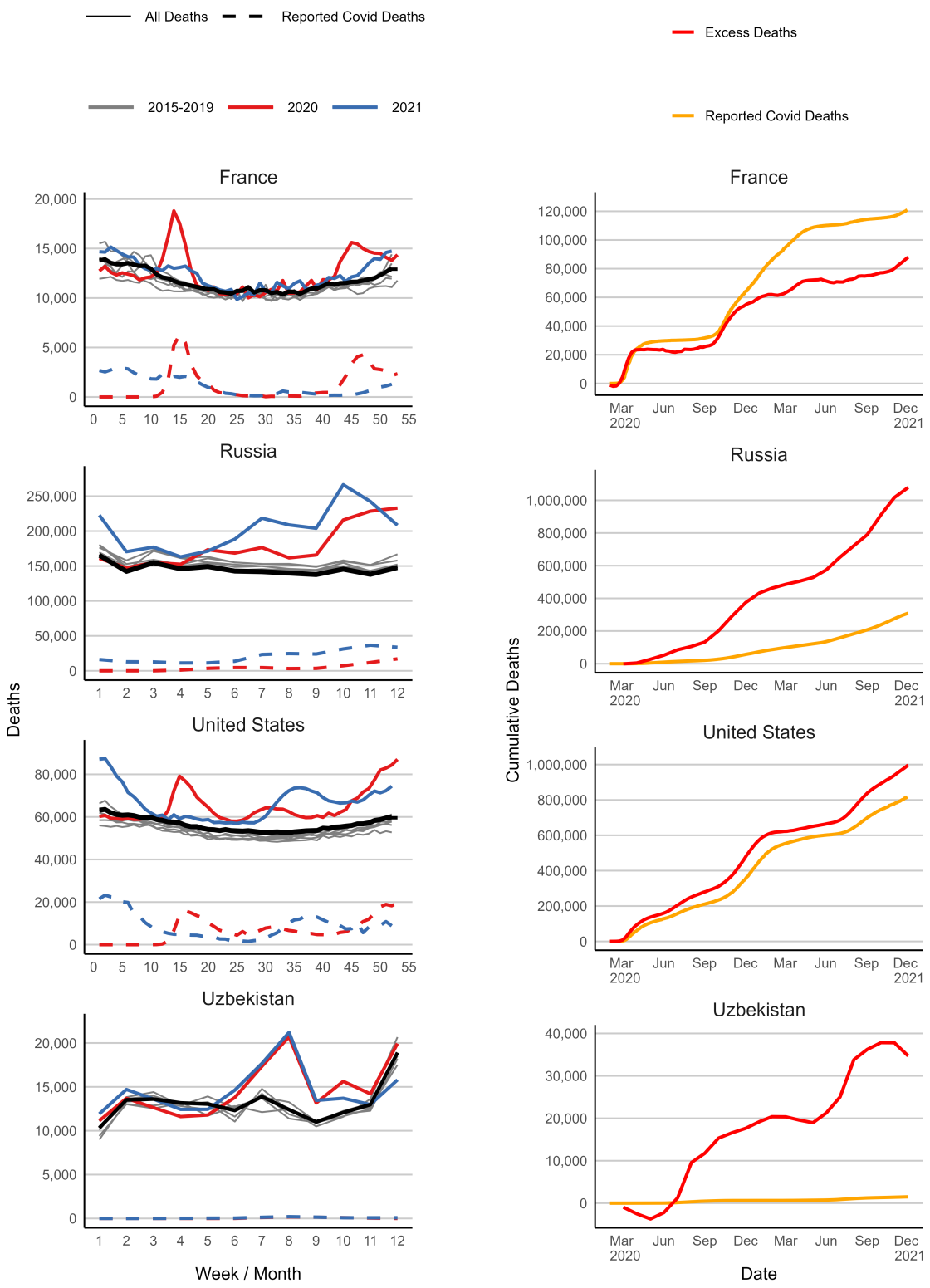
Our main goal is to estimate, for each country i, its Misreporting Rate (MRR), which we define as:

$$Misreporting\,Rate_i = MRR_i = \frac{Estimated\ Covid\ Deaths_i - Official\ Covid\ Deaths_i}{Expected\ Total\ Deaths_i} \quad (1)$$

The left panels in Figure 1 illustrate the calculation of *Estimated Covid Deaths*$_i$. Appendix B provides details. Thin gray lines show mortality in pre-covid years. The thick black lines show our estimated expected mortality. As we show in Appendix Figure B1, our method for calculating expected mortality performs extremely well in predicting pre-covid mortality ($R^2 = .997$). *Estimated Covid Deaths*$_i$ is then the excess deaths in country i, calculated as the cumulative difference between observed mortality in 2020 (red solid line) and 2021 (blue), and the expected mortality in these years. Next, *Official Covid Deaths*$_i$, shown by the dashed lines, is the number of Covid deaths reported by i's government. These reported numbers attracted most of the public's attention, and were used by governments to showcase their performance. For example, Russian officials used these numbers as testament to their health care system's effectiveness (Chance et al., 2020). Finally, the right panels show the evolution of the cumulative *Estimated Covid Deaths* and *Official Covid Deaths.* Thus, by the end of 2021 France has reported more Covid deaths than our estimates, whereas Russia and Uzbekistan have reported far less.

The MRR then normalizes the difference between estimated and official mortality by the expected number of deaths in country i based on pre-Covid mortality patterns, to account for population size, age-structure, comorbidities and other factors which influence routine mortality (Aron and Muellbauer, 2020). Thus, if a government reports 30,000 Covid deaths, while the estimated actual Covid deaths is 50,000, and the total expected number of deaths for the same country in the same period is 100,000, then misreporting is equivalent to 20% of the expected deaths in the country in this period. Results using alternative normalizations are reported in Appendix Figure D3 (country rankings) and in Table 3 (correlates of misinformation). Results are similar. Notice that the MRR measures the extent to which a government is undercounting (and possibly concealing) Covid deaths. A negative MRR indicate Covid over-counting (e.g. France).

Figure 1: Deriving Covid Misreporting

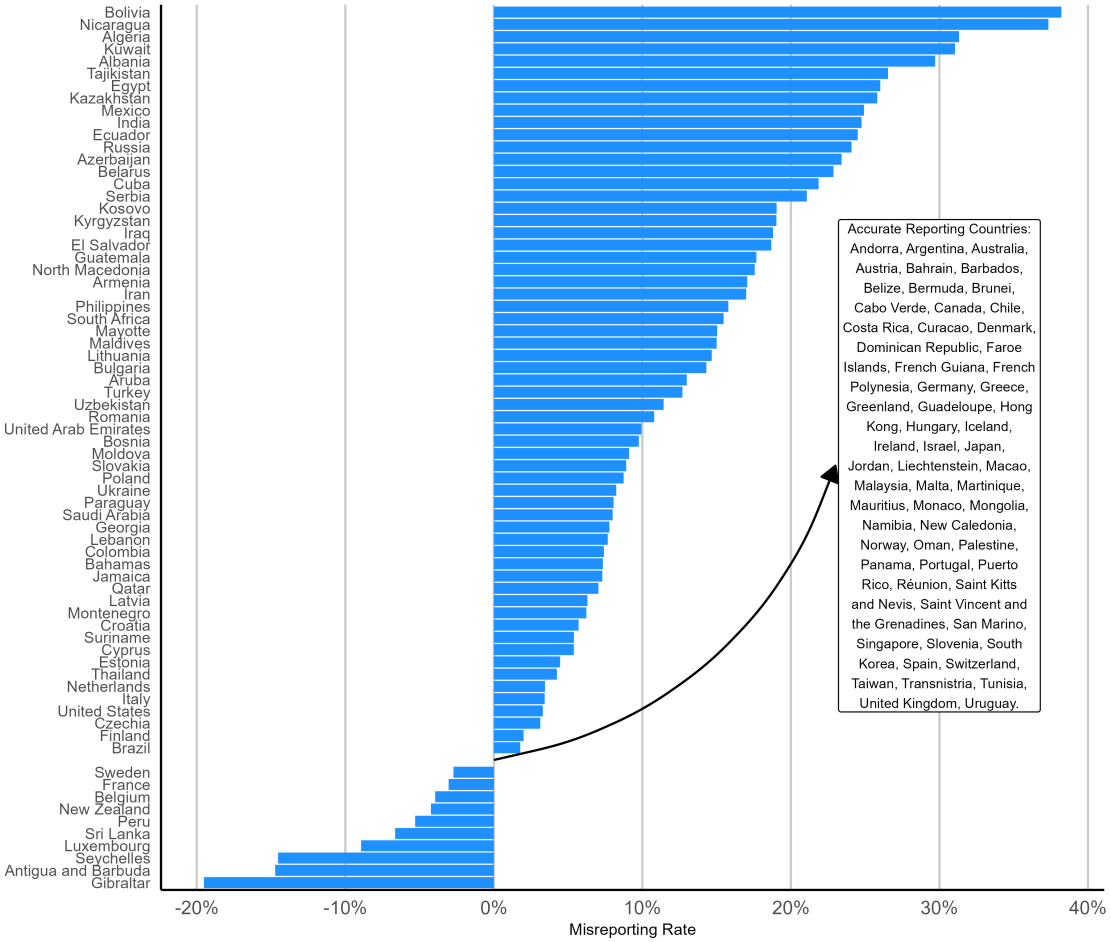


Notes: The left panel shows periodical (weekly or monthly) mortality and expected mortality from 2015 to 2021 in solid lines, and reported Covid mortality in dashed lines. The right panel shows the tracking between cumulative excess mortality (difference between actual to expected mortality) and reported Covid deaths in 2020-21.

4.1 Results

In total, the countries in our sample reported 5.07 million Covid deaths during 2020-2021. Our estimated actual Covid deaths is 12.39 million, and total expected death is 57.35. Thus the "global" Misreporting Rate in our sample is 12.8%. Figure 2 presents Misreporting Rates by country, ranked from highest to lowest. MRRs range from over 37% of expected deaths in Bolivia and Nicaragua to −19% in Gibraltar. For the countries listed as accurately reporting (see text box), the officially-reported number of Covid deaths is within the 95% confidence interval of the Estimated Covid deaths (see Appendix B).

Figure 2: Misreporting Rate (MRR) by Country



Notes: The figure shows the Misreporting Rate (MRR) for each country in our data, ranked from highest (estimated deaths higher than reported) to lowest. Countries where the estimated MRR is insignificantly different from 0 (accurate reporting countries) are shown in the inset. The point estimates of the MRR in the accurate reporting countries ranges from −9% to 15%. Bhutan, Fiji and Vietnam do not have sufficient data to estimate a CI, and are excluded from this figure. Their MRRs are −3.1%, 9.7% and 9.8%, respectively.

As discussed in the introduction, in theory the direction of information manipulation when

facing a crisis can go either way. In the case of Covid, however, the empirical finding is very sharp: regimes were much more likely to conceal than exaggerate mortality. Out of 131 countries, 60 exhibit accurate reporting and 61 reported significantly less deaths than their excess mortality. Only 10 countries reported more Covid deaths than excess deaths. Furthermore, there is little evidence for systematic overreporting in countries facing domestic unrest or that depend on foreign aid. The stringency of the initial response to Covid is also not significantly associated with either over- or under-reporting (see Appendix C).

These results can shed light on the mechanisms that may lead governments to exaggerate threats. Our evidence indicates a limited tendency of governments to purposefully exaggerate the gravity of a crisis in order to induce a "flight to safety" (Quattrone and Tversky, 1988; Bisbee and Honig, 2022), or to create an atmosphere of emergency that can legitimize repressive measures and induce citizens to accept restrictions on civil liberties and democratic procedures (Sturm, 1949; Davis and Silver, 2004; Alsan et al., 2023). With respect to the "rally around the flag" mechanism, however, our results support a somewhat nuanced view. Not every national threat or disaster generates a rally effect. Rather, we would argue that the threat needs to be outgroup-related to be effectively harnessed to rally the population. While some politicians have tried this route (Notably president Trump and the "Chinese virus"), by and large this was not the case with Covid. Similarly, our results do not rule out the possibility of governments using misinformation to instill fear in order to suppress dissent (Young, 2019), as Covid deaths do not signal the government's repressive capacity. A final observation is that the extent of misreporting is hardly a corner solution. While some governments appear to manipulate information to a shocking extent—hiding more than a third of the number of *total* deaths in the country—the MRR varies quite smoothly over the misreporting spectrum.

Note that the MRR may not capture the full extent of information manipulation if, in addition to Covid mortality data, governments manipulate the all-cause-mortality data on which we base our excess mortality estimates. Indeed, countries might manipulate all-cause-mortality data in response to studies such as this one. In practice, however, this effect is probably very modest in magnitude. First, as mentioned in the introduction, all-cause mortality data have routinely, and for a long time, been collected and reported by the bureaucratic apparatus, independently of Covid. Second, all-cause mortality is far easier to ascertain than Covid deaths. Third, during the first two years of the

pandemic, officially-reported Covid deaths and all-cause mortality data attracted vastly different levels of public attention, with Covid figures being far more visible, making their manipulation more urgent and influential. Finally, in instances where governments did seem to realize the potentially-revealing nature of excess mortality, their response has not been to manipulate information on all-cause-mortality, but rather to delay its publication (perhaps indefinitely).[5]

4.2 The Role of State Capacity

The MRR estimates provide a systematic database on the extent to which governments release inaccurate and misleading information regarding a central policy issue. However, some misreporting may be unavoidable given limited capacity to accurately diagnose, count and report Covid deaths. To understand manipulation, comparing misreporting across countries with similar capacity should provide a better gauge.

While conceptually straightforward, implementing this idea requires some care. Employing broad proxies for state capacity risks introducing confounds. For example, a common proxy for state capacity is GDP per capita. However, systematic information manipulation may well lead to lower GDP, and GDP figures may themselves be manipulated (Martinez, 2022).[6] Conditioning on specific Covid-related variables, such as the number of tests administered, is also unsuitable, as these variables may themselves reflect efforts to reveal or obfuscate Covid mortality. If a government wishes to obscure information on Covid deaths, one way is to simply administer fewer Covid tests (see World Bank 2021a on testing policy in Egypt).

Therefore, our baseline approach employs two measures that are both specific to a country's capacity to diagnose, count and report deaths, and yet can be computed independently of the response to Covid, using pre-pandemic data. This has the added benefit of reducing measurement errors compounded by the pandemic, when vital registration systems reported disruptions to their ability to collect data (Centre of Excellence for CRVS Systems, 2021). As with all-cause mortality, the inputs used to compute these measures have in general received very little public attention.

[5]For example, Belarus has officially released mortality data only up to June 2020, which showed significant undercounting. This was widely published by the regime's dissidents. In response, Belarus' statistical office has removed deaths reporting in following publications where this information was included in the past. For this study, we use additional monthly mortality information from Belarus up to March 2021, obtained in a data leak by Soshnikov (2021). The estimated MRR is almost identical using only the official data.

[6]In practice, controlling for GDP per capita does not make a big difference. See Table 3.

Table 1: Misreporting and State Capacity
Dependent Variable: Misreporting Rate

	(1)	(2)	(3)	(4)	(5)	(6)
Death Registration Completeness	-0.200**				-0.104	-0.0345
	(0.101)				(0.110)	(0.118)
Well Certified Death Registrations		-0.263***			-0.208*	-0.155
		(0.0999)			(0.120)	(0.134)
Statistical Performance Indicator			-0.195**			-0.0592
			(0.0855)			(0.102)
Tax Revenue (% GDP)				-0.259***		-0.166
				(0.0882)		(0.123)
R^2	0.040	0.095	0.079	0.106	0.104	0.152

Notes: OLS. N=134. All models include dummies for missing data. Death Registration Completeness is from 2019. Percent Well Certified Death Registrations is from latest year available (mostly 2017). Statistical Performance Indicator is from 2019. Tax Revenue as share of GDP in 2016. Coefficients are standardized. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

The measures are: (1) ***Death Registration Completeness* (DRC)**: the share of deaths captured by a country's vital registration system each year. DRC is considered the most basic but important feature of a well-functioning vital registration system. We use an array of sources to update, expand, and resolve inconsistencies in previously-compiled data, see Karlinsky (2021). (2) ***Percent of Well-Certified Death Registrations* (PWC):** the share of death certificates with proper cause of death attribution.[7]

In Table 1 and Appendix Tables D, we further include two broader measures of statistical and state capacity: (3) ***Statistical Performance Indicator* (SPI)** (World Bank, Dang et al. 2021) measures the performance of national statistical systems. SPI strongly correlates with GDP per capita ($\rho = 0.47$), and one should be careful interpreting results using it as control. (4) ***Tax Revenue*** *(% GDP, 2016,* Besley et al. 2021) proxies the general government's command and (fiscal) extraction capacity. While suffering from some of the limitations of GDP per capita, the correlation between them is low ($\rho = 0.17$).

Table 1 regresses the MRR on each of these measures. To the extent that misreporting is due to capacity constrains, we expect a negative relationship. This is confirmed for all measures. The R^2 ranges from 0.04 to 0.106. Column 5 includes both DRC and PWC, which will serve as our main state capacity set. Column 6 includes all four measures.

[7]While our DRC is mostly up-to-date (2019), Currently available PWC data include many outdated estimates. For example, Bolivia's latest PWC is from 2003.

Overall, while low capacity is associated with underreporting, it does not account for much of the cross-country variation. For example, Chile and Russia both have complete DRC and almost identically high PWC, suggesting high mortality counting capacity. However, Chile exhibits accurate reporting while Russia exhibits significant undercounting, with an MRR of 24%.

5 Correlates of Misinformation

Which governments are more likely to manipulate information? We begin by exploring four broad domains: macroeconomic incentives, culture, audience sophistication, and institutions. We also examine whether manipulation is associated with the electoral cycle, and probe potential historical roots. Our analysis is descriptive, employing simple regressions of the form:

$$MRR_i = \alpha + \beta Factor_i + Capacity_i'\gamma + \epsilon_i \tag{2}$$

where $Factor_i$ is a potential determinant of Covid misreporting, and $Capacity_i$ is a vector of measures of country i's capacity for diagnosing, counting and reporting mortality. Variables are standardized, such that β measures the change in standard deviations of the MRR associated with a one standard deviation change in each $Factor$. Appendix Table D1 reports descriptive statistics and data sources for each factor (all measured pre-pandemic). Table 2 examines each domain separately, and Table 3 combines them to gauge their relative importance. Appendix Figures D1 and D2 show the raw relationships.

A natural starting point is that countries might have economic and financial reasons to underreport Covid deaths. We explore three plausible factors. First, a country that relies heavily on tourism might have more to gain from painting a rosy picture in order to attract (or bring back) tourists. Second, countries facing recessions or high unemployment might wish to boost (or avoid loss of) confidence in the country's economic performance over the business cycle. Finally, a country that relies on foreign investments might try to persuade international investors that the spread and impact of Covid is contained, and the economy can return to normal.

Table 2, Columns 1-3, present the results. The bottom row reports Bonferroni adjusted p-values for the 16 multiple comparisons. None of these economic factors is significantly associated with misreporting Covid, possibly consistent with some level of sophistication on the part of international

Table 2: Correlates of Misinformation

Dependent Variable: Misreporting Rate

	Macroeconomic Incentives				Culture			
	International Tourism	Unemployment	Net FDI	1st PC	Interpersonal Trust	Individualism	Religiosity	1st PC
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Std. β	-0.123	-0.00973	-0.00477	-0.0312	-0.317***	-0.282***	0.382***	-0.341***
	(0.0968)	(0.0732)	(0.0413)	(0.0517)	(0.0685)	(0.0676)	(0.0721)	(0.0645)
Controls:								
DRC	✓	✓	✓	✓	✓	✓	✓	✓
PWC	✓	✓	✓	✓	✓	✓	✓	✓
R^2	0.138	0.115	0.111	0.127	0.240	0.181	0.275	0.217
R^2 Partial	0.0120	0.000227	0.000390	0.00121	0.109	0.0894	0.145	0.122
Adj. p_{val}	1	1	1	1	< 0.001	< 0.001	< 0.001	< 0.001
	Audience Sophistication				Institutions			
	Tertiary Education	HDI Education Index	Internet Access	1st PC	Clean Elections	Executive Constraints	Press Freedom	1st PC
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Std. β	-0.147*	-0.284***	-0.308***	-0.358***	-0.474***	-0.559***	-0.418***	-0.539***
	(0.0796)	(0.0890)	(0.0955)	(0.0779)	(0.110)	(0.0836)	(0.0871)	(0.0871)
Controls:								
DRC	✓	✓	✓	✓	✓	✓	✓	✓
PWC	✓	✓	✓	✓	✓	✓	✓	✓
R^2	0.130	0.167	0.194	0.228	0.286	0.343	0.238	0.333
R^2 Partial	0.0215	0.0835	0.106	0.130	0.187	0.239	0.162	0.223
Adj. p_{val}	1	.0288	.0259	< 0.001	< 0.001	< 0.001	< 0.001	< 0.001

Notes: OLS. N=134. DRC = Death Registration Completeness. PWC = Percent of Well-Certified Death Registrations. All regressions include dummies for missing data. Coefficients are standardized. R^2 Partial is the R^2 attributable to the shown explanatory variables (Huettner and Sunder, 2012). Adj. p_{val} is the Bonferroni adjusted p-value for 16 multiple comparisons. International Tourism is share of GDP in 2018 (latest avilable year). FDI Net is difference between inbound to outbound foreign direct investment, as share of GDP in 2019. Unemployment is rate in 2019. Interpersonal Trust is share of population who agree that "Most people can be trusted". Individualism is "The degree of interdependence a society maintains among its members". Religiosity is share of people who answer that religion in their life is very or rather important. Tertiary Education is share of population aged 25-64 with completed tertiary education. HDI Education Index is average of mean years of schooling and expected years of schooling in the population. Internet Access is share of population with internet access. Clean Elections measures "To what extent are elections free and fair". Exectuive Constraints is average of "Judicial Constraints" and "Legislative Constraints" on the executive indices. Freedom of the Press measures the degree to which journalists and reporters are able to operate and report freely. 1st PC accounts for 43%, 76%, 74% and 88% of the common variation in domains Macroeconomic Incentives, Culture, Audience Sophistication and Institutions, respectively. Variables sources and descriptive statistics are in Appendix Table D1. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

economic actors. Variation in these factors explains very little of the variation in misreporting relative to the overall R^2, suggesting they do not play a major role in explaining misreporting.

A second set of potential factors relates to the country's culture. Societies with high interpersonal trust also display more civic honesty as measured by return rates in wallet drop experiments (Cohn et al. 2019; Bjørnskov 2021). In individualistic societies, truth is highly valued, even when honesty is costly (Hofstede et al. 2010). Religiosity has also been associated with honesty and trust, although the evidence is mixed (Alesina and La Ferrara, 2002; Barro and McCleary, 2003; Guiso et al., 2006; Iyer, 2016). Columns 5-7 seem to suggest that these cultural factors might indeed play some role. Governments in countries with higher trust show significantly less underreporting. This is also true for more individualistic countries. In contrast, governments in more

religious countries underreport more. The explanatory power of these cultural factors is markedly higher than the above economic factors. As we show below, however, the association of cultural factors with misreporting can be attributed to other factors.

Turning to restraining factors, governments that seek to manipulate information may face two types of constraints. The first arises from the nature of the target audience: the more sophisticated the audience, the less likely it is to be misled. In equilibrium, if deception is costly—due, e.g., to fabrication costs, legal penalties, or moral and cultural constraints—governments may take this into account and only deceive in the presence of unsophisticated, naïve audiences (see Kartik et al., 2007). We proxy for (domestic) audience sophistication in three ways. First, we use the share of the adult population with tertiary education, which represents the relative size of the educated elite in the population (Guriev and Treisman, 2019). Second, we use the Human Development Education Index (UNDP, 2020) as a measure of the sophistication of the general population. Finally, we use the share of the population with internet access, which allows exposure to more sources of information, and has been shown to reduce trust in government, especially when traditional media are censored (Guriev et al., 2020).

Columns 9-11 in Table 2 report the results. Indeed, after adjusting for capacity, the size of the educated elite is negatively associated with underreporting. However, both the general educational attainment of the population and widespread access to the internet are more strongly and negatively associated with governmental manipulation of information. The Partial R^2s of these variables are between 8-10%, similar in magnitude to the cultural factors.

Beyond the nature of the audience, a second and widely-discussed set of constraints arises from institutions, both formal and informal. Free and fair elections can serve as a check on executive deception if voters punish politicians for lying. The executive can further be constrained by other branches of the government such as the legislature, government agencies (e.g., the CDC in the US), and the judiciary. We measure these factors using the "Clean Elections" and "Executive Constraints" indices from Varieties of Democracy Institute (2021). Finally, a free press can serve as an additional constraint by uncovering deception and manipulation, which we measure using the "Freedom of the Press" index (Reporters Without Borders, 2019).

Our data are consistent with these expectations (Columns 13-15). Institutional constraints are significantly and negatively related to Covid underreporting (all measures are transformed such

that higher values indicate stronger constraints). The magnitudes are sizable. For each of the three factors, a one standard deviation stronger institutional constraints is associated with about 0.4 to 0.5 standard deviations lower MRR. The explanatory power of these institutional constraints is markedly larger than that of any of the factors explored in the preceding domains.

The factors within each domain are, of course, correlated. To help disentangle the different forces, we start by conducting a principal component analysis in each domain to recover the common variation. In columns 4,8,12,16 we use each domain's first principal component as a single measure of the variation in it.[8] Both the coefficients and the Partial R^2 are highest for institutional constraints, followed by audience sophistication and culture.

In Table 3 we use these principal components to gauge the relative importance of the different domains. We further include two other factors of interest. First, incentives to manipulate information may depend on the political cycle (see Wallace 2016). Indeed, there is evidence that the political cycle affected the stringency of anti-Covid measures (Pulejo and Querubin, 2021). We therefore distinguish between the 76 countries where elections took place during 2020-21, and the rest. Second, so far we studied the relationship between misreporting and contemporaneous country characteristics. However, countries may carry different historical legacies. In particular, information manipulation was rife in communist countries (Harrison, 2017). We classify countries as having a communist legacy if they had a communist or socialist regime for at least 10 years (34 countries).[9]

Column 1 in Table 3, reports the results [the R^2 attributable to each factor is reported in square brackets]. As expected, some coefficients are attenuated compared to Table 2, as there is significant correlation between former communist countries and weaker constraints, for example. Most notably, cultural factors are not significantly associated with misreporting once we condition on other factors. The strongest predictor of misreporting is weak institutional constraints on the executive, but Audience Sophistication also remains substantial in size. Holding elections and having a Communist Legacy are also significant predictors of misreporting, both statistically and

[8] For Culture, Audience Sophistication and Institutions, the first principal component accounts for a significant share (73%-88%) of the total variation. In the case of Macroeconomic Incentives, the first principal component accounts for 43% of the total variation.

[9] Information manipulation has also been used by some colonial powers (Cohn, 1996). We explore this potential legacy in Appendix Table D2, finding little evidence that formerly colonized countries are systematically different from other countries in misreporting.

Table 3: Correlates of Misinformation (Continued)

	$MRR = \frac{\text{Excess - Reported}}{\text{Expected}}$				$\frac{\text{Excess - Reported}}{\text{Population}}$	$\log(\frac{\text{Excess}}{\text{Reported}})$
	(1)	(2)	(3)	(4)	(5)	(6)
Macroeconomic Incentives (1st PC)	-0.0154	-0.0210	-0.0102	-0.0249	-0.0271	0.0387
	(0.0544) [0.0026]	(0.0556)	(0.0593)	(0.0573)	(0.0402)	(0.0491)
Culture (1st PC)	-0.0257	-0.0450	-0.0219	-0.151	-0.0904	0.163
	(0.105) [0.0537]	(0.105)	(0.107)	(0.118)	(0.100)	(0.109)
Audience Sophistication (1st PC)	-0.278**	-0.315***	-0.238*	-0.156	-0.113	-0.337**
	(0.106) [0.0761]	(0.104)	(0.137)	(0.115)	(0.105)	(0.133)
Institutions (1st PC)	-0.354***	-0.383***	-0.354***	-0.319***	-0.311***	-0.464***
	(0.0823) [0.1349]	(0.0898)	(0.0827)	(0.0986)	(0.106)	(0.0874)
Elections during 2020-21	0.410***	0.430***	0.411***	0.472***	0.217	0.287*
	(0.138) [0.0547]	(0.137)	(0.142)	(0.147)	(0.137)	(0.150)
Communist Legacy	0.479***	0.399*	0.447**	0.444**	0.956***	0.385**
	(0.180) [0.0791]	(0.204)	(0.179)	(0.170)	(0.204)	(0.184)
Population≥65 (%)		0.127				
		(0.132)				
GDP Per Capita (log)			-0.0618			
			(0.135)			
Controls:						
DRC	✓	✓	✓	✓	✓	✓
PWC	✓	✓	✓	✓	✓	✓
Sample	Main	Main	Main	Extended	Main	Main
Observations	134	134	134	137	134	134
R^2	0.507	0.513	0.508	0.463	0.473	0.444

Notes: OLS. DRC = Death Registration Completeness. PWC = Percent of Well-Certified Death Registrations. All regressions include dummies for missing data. Continuous coefficients are standardized and discrete coefficients are y-standardized. 1st PC accounts for 43%, 76%, 74% and 88% of the common variation in domains Macroeconomic Incentives, Culture, Audience Sophistication and Institutions, respectively. Elections during 2020-21 equals 1 for countries that had national parliamentary or presidential elections in 2020-21. Communist Legacy equals 1 for countries that had a communist or socialist regime (lowest duration is Nicaragua 1979-1989 (Bjørnskov and Rode, 2020)). Square brackets in column 1 show the R^2 Partial attributable to each variable (Huettner and Sunder, 2012). Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

economically. In total, this model accounts for more than 50% of the variation in MRR.

Columns 2-6 examine the robustness of these conclusions to several plausible concerns . First, Covid mainly affects the elderly (O'Driscoll et al., 2021) and countries with good governance tend to have older populations. In Column 2, we thus control for the share of the population aged 65 or older. This does not meaningfully affect the results. Second, as discussed in section 4.2, adjusting for GDP per capita carries the risk of eliminating some of the very variation we wish to analyze. Nonetheless, one worries that the observed association between (for example) institutions and misreporting might be driven by the relationship between institutions and economic development. To check this, we add log GDP per-capita in Column 3. The estimated coefficients of Institutions, Elections and Communism are largely unaffected.

Next, we collected data from Bangladesh, China and Indonesia. As we explain in Appendix section A, these data should be taken with a grain of salt. Adding these countries to our analysis

(Column 4) yields very similar results to our main specification.

Finally, we analyze alternative normalizations to the misreporting gap. In Column 5, we normalize the gap between Excess and Reported Covid deaths by population. In column 6, we use the log of the ratio. The coefficients on Audience Sophistication and Elections are attenuated and lose statistical significance, but the explanatory power of institutions is very robust across all specifications.

6 Conclusions

Governments, whether democratically-elected or not, might wish to manipulate information in order to appear more competent than their performance indicates. Using the global Covid pandemic as a manipulation-revealing mechanism, we find evidence for systematic discrepancies in Covid reporting which cannot be attributed to state capacity to collect, diagnose, and report mortality information. The evidence shows that manipulation is remarkably widespread and large in magnitude. Within the 61 countries in our data that underreport Covid mortality, the average under-reporting is equivalent to 14.5% of the expected deaths in the country in the relevant period. However, the manipulation of information varies significantly across countries. Its extent is best explained by the constraints governments face: both institutional and those deriving from the sophistication of their audience. In countries where formal and informal checks and balances are weak, executives tend to act as "Informational Autocrats" (Guriev and Treisman, 2022).

Three implications of our findings are worth noting. First, consistent with recent evidence on GDP reporting (Martinez, 2022), we document extensive information manipulation by authoritarian regimes. Unconstrained regimes seem to manipulate not just economic information, but also information on arguably the most important public policy issue during this period. The implications of this manipulation extended beyond domestic policy: for example, travel restrictions during the pandemic heavily relied on *reported* Covid stats by countries. This underscores questions regarding which and how much of the official information provided by such governments should be trusted. It also complicates the implementation and evaluation of international initiatives such as foreign aid, climate action, and vaccination campaigns. Second, our estimates provide a measure of institutional quality based on actual governmental action that is comparable across countries. They

can therefore contribute to efforts to understand the relationship between institutional quality and conflict, electoral outcomes, economic performance, and trust. Finally, our analysis identifies high misreporting rates in some countries that present respectable scores on executive constrains (e.g. Albania, El Salvador, India, Serbia). These MRRs might be leading indicators for the erosion of democratic institutions in these countries – erosion that has yet to filter thorough to some of the currently-available measures.

A final methodological point. Deriving our misreporting rate requires historical and current data on all-cause mortality. Thus, we are unable to score and analyze information manipulation for important and large countries such as Pakistan, Nigeria, Ethiopia, and Venezuela. This lack of public data may itself be informative: data may not be publicly available precisely because authorities do not wish to reveal inconvenient information. But while the authorities in Pakistan and Venezuela have the capacity to collect the required data, we have little indication that this is the case in Ethiopia and Nigeria (Karlinsky, 2021). In our assessment, lack of published data from countries with functioning vital registration systems is probably indicative of some (unknown) degree of information manipulation. This is an important area for further research.

References

Abramovich, Felix and Ritov, Ya'acov. *Statistical Theory: A Concise Introduction*. Chapman and Hall/CRC, Boca Raton, 1st edition edition, April 2013.

Alahmad, Barrak, AlMekhled, Dawoud, Odeh, Ayah, Albloushi, Dalia, and Gasana, Janvier. Disparities in excess deaths from the COVID-19 pandemic among migrant workers in Kuwait. *BMC Public Health*, 21(1):1668, September 2021.

Alesina, Alberto and La Ferrara, Eliana. Who trusts others? *Journal of Public Economics*, 85(2): 207–234, 2002.

Alissa, D.A., Aburas, W., Almasuood, R., Almudaiheem, H.Y., Al Aseri, Z., Alrabiah, F., Ezzat, H., Moulana, A.A., Alawi, M.M., Al-mutairy, E., Alaama, T., Alamri, M.S., Bamousa, M.S., Alshehri, A.A., Alosaimi, M.H., Alali, A., Nori, R., Devol, E.B., Mohamed, G., and Al-jedai, A.H. Prevalence and epidemiological trends in mortality due to covid-19 in saudi arabia. *Public Health*, 215:31–38, 2023.

Alsan, Marcella, Chandra, Amitabh, and Simon, Kosali. The Great Unequalizer: Initial Health Effects of COVID-19 in the United States. *Journal of Economic Perspectives*, 35(3):25–46, August 2021.

Alsan, Marcella, Braghieri, Luca, Eichmeyer, Sarah, Kim, Minjeong Joyce, Stantcheva, Stefanie, and Yang, David Y. Civil liberties in times of crisis. *American Economic Journal: Applied Economics*, 15(4):389–421, October 2023.

Anand, Abhishek, Sandefur, Justin, and Arvind, Subramanian. Three New Estimates of Indias All-Cause Excess Mortality during the COVID-19 Pandemic. *Center For Global Development*, 2021.

Aron, Janine and Muellbauer, John. A pandemic primer on excess mortality statistics and their comparability across countries. Technical report, Our World in Data, 2020.

Banaji, Murad and Gupta, Aashish. Estimates of pandemic excess mortality in india based on civil registration data. *PLOS Global Public Health*, 2022.

Bangladesh Bureau of Statistics. Bangladesh sample vital statistics 2021. 2023.

Barro, Robert J. and Lee, Jong Wha. A new data set of educational attainment in the world, 1950–2010. *Journal of Development Economics*, 104:184–198, 2013.

Barro, Robert J. and McCleary, Rachel M. Religion and economic growth across countries. *American Sociological Review*, 68(5):760–781, 2003.

BBC. Coronavirus: Iran cover-up of deaths revealed by data leak, August 2020a, *BBC News*. https://www.bbc.com/news/world-middle-east-53598965.

BBC. Coronavirus: Israel passes law to ban mass protests during lockdown, September 2020b, *BBC News*. https://www.bbc.com/news/world-middle-east-53598965.

Beaney, Thomas, Clarke, Jonathan M, Jain, Vageesh, Golestaneh, Amelia Kataria, Lyons, Gemma, Salman, David, and Majeed, Azeem. Excess mortality: the gold standard in measuring the impact of COVID-19 worldwide? *Journal of the Royal Society of Medicine*, 113(9):329–334, 2020.

Becerra, Oscar, Cavallo, Eduardo, and Noy, Ilan. Foreign aid in the aftermath of large natural disasters. *Review of Development Economics*, 18(3):445–460, 2014.

Becker, Bastian. Introducing coldat: The colonial dates dataset. 2019.

Besley, Timothy J., Dann, Chris, and Persson, Torsten. Pillars of Prosperity: A Ten-Year Update. (16256), June 2021.

Biller, David and Savarese, Mauricio. Bolsonaro's concealment of brazilian coronavirus data a survival move: experts, 2020, *Global News*. https://globalnews.ca/news/7042838/coronavirus-brazil-bolsonaro-data-concealment/.

Bisbee, James and Honig, Dan. Flight to safety: Covid-induced changes in the intensity of status quo preference and voting behavior. *American Political Science Review*, 116(1):70–86, 2022.

Bjørnskov, Christian. Civic honesty and cultures of trust. *Journal of Behavioral and Experimental Economics*, 92:101693, June 2021.

Bjørnskov, Christian and Rode, Martin. Regime types and regime change: A new dataset on democracy, coups, and political institutions. *The Review of International Organizations*, 15(2): 531–551, April 2020.

Boka, Debra M and Wainer, Howard. How can we estimate the death toll from COVID-19? *CHANCE*, 33(3):67–72, 2020.

Brum, Eliane. Bolsonaro government imposes data blackout on covid-19 in brazil amid spike in deaths. *El Pais*, 2020.

Centre of Excellence for CRVS Systems. Synthesis Brief: The impact of the COVID-19 pandemic on Africa: How civil registration and vital statistics systems supported an emergency response, 2021.

Chance, Matthew, Ullah, Zahra, and Hodge, Nathan. Exclusive: Did Russia pass the coronavirus test? Kremlin spokesman Dmitry Peskov responds, 2020, *CNN*. https://www.cnn.com/2020/06/09/europe/russia-peskov-interview-intl/index.html.

Cohn, Alain, Maréchal, Michel André, Tannenbaum, David, and Zünd, Christian Lukas. Civic honesty around the globe. *Science*, 365(6448):70–73, 2019.

Cohn, Bernard S. *Colonialism and its forms of knowledge: The British in India*. Princeton University Press, 1996.

Costa, Robert and Rucker, Philip. Woodward book: Trump says he knew coronavirus was deadly and worse than the flu while intentionally misleading Americans. *Washington Post*, 2020.

Dang, Hai-Anh H., Dinc, Mustafa, Dias, Juderica Zilla Josephine, Maeda, Hiroko, Pullinger, John James, Serajuddin, Umar, Stacy, Brian William, and Wolde, Dereje Ketema. Measuring the Statistical Performance of Countries : An Overview of Updates to the World Bank Statistical Capacity Index. *World Bank*, 2021.

Davis, Darren W and Silver, Brian D. Civil liberties vs. security: Public opinion in the context of the terrorist attacks on america. *American journal of political science*, 48(1):28–46, 2004.

De Vries, Catherine E and Solaz, Hector. The electoral consequences of corruption. *Annual Review of Political Science*, 20:391–408, 2017.

de Walque, D. and Verwimp, P. The Demographic and Socio-economic Distribution of Excess Mortality during the 1994 Genocide in Rwanda. *Journal of African Economies*, 19(2):141–162, 01 2010.

Edmond, Chris. Information manipulation, coordination, and regime change. *Review of Economic studies*, 80(4):1422–1458, 2013.

Ekamper, Peter, Bijwaard, Govert, van Poppel, Frans, and Lumey, L. H. War-related excess mortality in the netherlands, 1944–45: New estimates of famine- and non-famine-related deaths from national death records. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 50(2):113–128, 2017.

Farr, William. *Vital statistics: a memorial volume of selections from the reports and writings of William Farr*. Offices of the Sanitary Institute, London, 1885.

Farrington, CP, Andrews, Nick J, Beale, AD, and Catchpole, MA. A statistical algorithm for the early detection of outbreaks of infectious disease. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 159(3):547–563, 1996.

Frederiksen, Kristian Vrede Skaaning. Does competence make citizens tolerate undemocratic behavior? *American Political Science Review*, pages 1–7, 2022.

Fund for Peace. Fragile states index and cast framework methodology, 2017. https://fragilestatesindex.org/.

Gehlbach, Scott, Sonin, Konstantin, and Svolik, Milan W. Formal models of nondemocratic politics. *Annual Review of Political Science*, 19:565–584, 2016.

Gettleman, Jeffrey, Yasir, Sameer, Kumar, Hari, and Raj, Suhasini. As covid-19 devastates india, deaths go undercounted, 2021, *New York Times*. https://www.nytimes.com/2021/04/24/world/asia/india-coronavirus-deaths.html.

Grada, Cormac O. Great leap, great famine: A review essay. *Population and Development Review*, 39(2):333–346, 2013.

Guiso, Luigi, Sapienza, Paola, and Zingales, Luigi. Does culture affect economic outcomes? *Journal of Economic Perspectives*, 20(2):23–48, June 2006.

Guriev, Sergei and Treisman, Daniel. How modern dictators survive: An informational theory of the new authoritarianism. Technical report, National Bureau of Economic Research, 2015.

Guriev, Sergei and Treisman, Daniel. Informational Autocrats. *Journal of Economic Perspectives*, 33(4):100–127, November 2019.

Guriev, Sergei and Treisman, Daniel. *Spin Dictators*. Princeton University Press, 2022.

Guriev, Sergei, Melnikov, Nikita, and Zhuravskaya, Ekaterina. 3G Internet and Confidence in Government*. *The Quarterly Journal of Economics*, 136(4):2533–2613, 12 2020.

Haerpfer, C., Inglehart, R., Moreno, A., Welzel, C., Kizilova, K., J., Diez-Medrano, Lagos, M., Norris, P., Ponarin, E., and Puranen, B. WVS Database, 2020. https://www.worldvaluessurvey.org/WVSDocumentationWV7.jsp.

Hale, Thomas, Angrist, Noam, Goldszmidt, Rafael, Kira, Beatriz, Petherick, Anna, Phillips, Toby, Webster, Samuel, Cameron-Blake, Emily, Hallas, Laura, Majumdar, Saptarshi, and Tatlow, Helen. A global panel database of pandemic policies (Oxford COVID-19 Government Response Tracker). *Nature Human Behaviour*, 5(4):529–538, April 2021.

Harrison, Mark. Secrecy and state capacity: A look behind the iron curtain. *Warwick Economics Research Papers*, 2017.

Heifez, Kyrill and Liapeika, Hleb. Minsk in times of coronavirus. official covid-19 death data shows more belarusians deceased in minsk than reported for the entire country, april 2021, *Mediazona*. https://mediazona.by/translate/2021/04/23/covid.

Hofstede, Geert, Hofstede, Geert Jan, and Minkov, Michael. *Cultures and organizations: software of the mind: intercultural cooperation and its importance for survival*. McGraw-Hill, 2010.

Hollyer, James R, Rosendorff, B Peter, and Vreeland, James Raymond. Democracy and transparency. *The Journal of Politics*, 73(4):1191–1205, 2011.

Honryo, Takakazu and Yano, Makoto. Idiosyncratic information and vague communication. *American Political Science Review*, 115(1):165–178, 2021.

Huettner, Frank and Sunder, Marco. Axiomatic arguments for decomposing goodness of fit according to Shapley and Owen values. *Electronic Journal of Statistics*, 6(none):1239 – 1250, 2012.

IFES. Electionguide, 2023. https://www.electionguide.org/.

Islam, Nazrul, Shkolnikov, Vladimir M, Acosta, Rolando J, Klimkin, Ilya, Kawachi, Ichiro, Irizarry, Rafael A, Alicandro, Gianfranco, Khunti, Kamlesh, Yates, Tom, Jdanov, Dmitri A, White, Martin, Lewington, Sarah, and Lacey, Ben. Excess deaths associated with covid-19 pandemic in 2020: age and sex disaggregated time series analysis in 29 high income countries. *BMJ*, 373: n1137, 2021.

Iyer, Sriya. The new economics of religion. *Journal of Economic Literature*, 54(2):395–441, 2016.

Jankowicz, Mia. Trump is pressuring the CDC to change its death-toll methodology and produce a lower figure, report says, 2020, *Business Insider*. https://www.businessinsider.com/trump-pressuring-cdc-change-coronavirus-toll-method-daily-beast-2020-5.

Jha, Prabhat, Deshmukh, Yashwant, Tumbe, Chinmay, Suraweera, Wilson, Bhowmick, Aditi, Sharma, Sankalp, Novosad, Paul, Fu, Sze Hang, Newcombe, Leslie, Gelband, Hellen, and Brown, Patrick. Covid mortality in india: National survey data and health facility deaths. *Science*, 2022.

Karlinsky, Ariel. International Completeness of Death Registration 2015-2019. *medRxiv*, page 2021.08.12.21261978, August 2021.

Karlinsky, Ariel and Kobak, Dmitry. Tracking excess mortality across countries during the COVID-19 pandemic with the World Mortality Dataset. *eLife*, 10:e69336, June 2021.

Kartik, Navin, Ottaviani, Marco, and Squintani, Francesco. Credulity, lies, and costly talk. *Journal of Economic Theory*, 134(1):93–116, 2007.

Kilani, Ahmad. Authoritarian regimes' propensity to manipulate covid-19 data: a statistical analysis using benford's law. *Commonwealth & Comparative Politics*, 59(3):319–333, 2021.

Kobak, Dmitry. Excess mortality reveals Covid's true toll in Russia. *Significance*, 18(1):16, 2021.

Kobak, Dmitry. Underdispersion: A statistical anomaly in reported covid data. *Significance*, 19 (2):10–13, 2022.

La Presna Latina. Brazil senate: Bolsonaro used false data to say covid deaths were inflated, 2021, *La Presna Latina*. https://www.laprensalatina.com/brazil-senate-bolsonaro-used-false-data-to-say-covid-deaths-were-inflated/.

Langmuir, Alexander D. William Farr: founder of modern concepts of surveillance. *International Journal of Epidemiology*, 5(1):13–18, 1976.

LAPOP Lab. The americasbarometer, 2019. https://www.vanderbilt.edu/lapop.

Latinobarometro. Public opinion latin america, 2018. https://www.latinobarometro.org/.

Leffler, Christopher T., V, Joseph D. Lykins, Das, Saurav, Yang, Edward, and Konda, Sneha. Preliminary analysis of excess mortality in india during the covid-19 pandemic. *The American Journal of Tropical Medicine and Hygiene*, 2022.

Lowande, Kenneth and Rogowski, Jon C. Executive power in crisis. *American Political Science Review*, 115(4):1406–1423, 2021.

Malani, Anup and Ramachandran, Sabareesh. Using household rosters from survey data to estimate all-cause excess death rates during the covid pandemic in india. *Journal of Development Economics*, 159:102988, 2022.

Martinez, Luis R. How much should we trust the dictator's gdp growth estimates? *Journal of Political Economy*, 130(10):2731–2769, 2022.

Mendez, Rich. Trump officials bragged about pressuring CDC to alter Covid reports, emails reveal, April 2021, *CNBC*. https://www.cnbc.com/2021/04/09/trump-officials-bragged-about-pressuring-cdc-to-alter-covid-reports-emails-reveal-.html.

Mikkelsen, Lene, Phillips, David E., AbouZahr, Carla, Setel, Philip W., Savigny, Don de, Lozano, Rafael, and Lopez, Alan D. A global assessment of civil registration and vital statistics systems: monitoring data quality and progress. *The Lancet*, 386(10001):1395–1406, October 2015.

Milken Institute. Ascertainment of the estimated excess mortality from hurricane Maria in Puerto Rico. Technical report, Milken Institute School of Public Health, 2018.

Msemburi, William, Karlinsky, Ariel, Knutson, Victoria, Aleshin-Guendel, Serge, Chatterji, Somnath, and Wakefield, Jon. The WHO estimates of excess mortality associated with the COVID-19 pandemic. *Nature*, 613(7942):130–137, 2023.

Mueller, John E. Presidential popularity from truman to johnson. *American political science review*, 64(1):18–34, 1970.

Naumenko, Natalya. The political economy of famine: The ukrainian famine of 1933. *The Journal of Economic History*, 81(1):156–197, 2021.

Noufaily, Angela, Enki, Doyo G, Farrington, Paddy, Garthwaite, Paul, Andrews, Nick, and Charlett, Andre. An improved algorithm for outbreak detection in multiple surveillance systems. *Statistics in medicine*, 32(7):1206–1222, 2013.

Olken, Benjamin A. Monitoring corruption: evidence from a field experiment in indonesia. *Journal of political Economy*, 115(2):200–249, 2007.

O'Driscoll, Megan, Ribeiro Dos Santos, Gabriel, Wang, Lin, Cummings, Derek AT, Azman, Andrew S, Paireau, Juliette, Fontanet, Arnaud, Cauchemez, Simon, and Salje, Henrik. Age-specific mortality and immunity patterns of sars-cov-2. *Nature*, 590(7844):140–145, 2021.

Pulejo, Massimo and Querubin, Pablo. Electoral concerns reduce restrictive measures during the covid-19 pandemic. *Journal of Public Economics*, 198:104387, 2021.

Quattrone, George A and Tversky, Amos. Contrasting rational and psychological analyses of political choice. *American Political Science Review*, 82(3):719–736, 1988.

Reporters Without Borders. 2019 World Press Freedom Index – A cycle of fear, 2019. https://rsf.org/en/2019-world-press-freedom-index-cycle-fear.

Riffe, Tim, Acosta, Enrique, and COVerAGE-DB team. Data resource profile: COVerAGE-DB: a global demographic database of COVID-19 cases and deaths. *International Journal of Epidemiology*, 50(2):390–390f, 2021.

Romero, Keyling T. Data leaked by "anonymous" reveal that the ministry of health hid more than 6200 cases of covid-19, aug 2020, *Confidencial*. https://confidencial.com.ni/nacion/datos-filtrados-por-anonymous-revelan-que-el-minsa-oculto-mas-de-6200-casos-de-covid-19/.

Rozenas, Arturas and Stukal, Denis. How autocrats manipulate economic news: Evidence from russia's state-controlled television. *The Journal of Politics*, 81(3):982–996, 2019.

Soshnikov, Andrey. Excess mortality - 32 thousand people. Belarusian authorities repeatedly underestimate statistics during the coronavirus epidemic – leaked data, 2021, *Current Time*. https://www.currenttime.tv/a/smertnost-v-belarusi/31401342.html.

Sturm, Albert L. Emergencies and the presidency. *The Journal of Politics*, 11(1):121–144, 1949.

The Economist. There have been 7m–13m excess deaths worldwide during the pandemic, 2021. https://www.economist.com/briefing/2021/05/15/there-have-been-7m-13m-excess-deaths-worldwide-during-the-pandemic.

Times of Israel. Opposition MK calls for recount of coronavirus deaths, claims figure inflated, 2020, *Times of Israel*. https://www.timesofisrael.com/opposition-mk-calls-for-recount-of-coronavirus-deaths-claims-figure-inflated/.

UNDP. Human Development Report 2020: The Next Frontier: Human Development and the Anthropocene. 2020.

UNSD. Demographic yearbook 2021: Coverage of birth and death registration. Technical report, United Nations, Statistics Division, 2021.

Varieties of Democracy Institute. V-Dem Dataset 2021, 2021. doi: 10.23696/VDEMDS21. Type: dataset.

Ventura, Deisy de Freitas Lima and Rocha, Reis Rossana. An unprecedented attack on human

rights in brazil: the timeline of the federal governments strategy to spread covid-19. *Conectas*, 2021.

Viboud, Cécile, Simonsen, Lone, Fuentes, Rodrigo, Flores, Jose, Miller, Mark A, and Chowell, Gerardo. Global mortality impact of the 1957–1959 influenza pandemic. *The Journal of Infectious Diseases*, 213(5):738–745, 2016.

Vos, Theo et al. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. *The Lancet*, 396(10258): 1204–1222, 2020.

Wallace, Jeremy L. Juking the stats? authoritarian information problems in china. *British Journal of Political Science*, 46(1):11–29, 2016.

Whittaker, Charles, Walker, Patrick G T, Alhaffar, Mervat, Hamlet, Arran, Djaafara, Bimandra A, Ghani, Azra, Ferguson, Neil, Dahab, Maysoon, Checchi, Francesco, and Watson, Oliver J. Under-reporting of deaths limits our understanding of true burden of covid-19. *BMJ*, 375, 2021.

World Bank. Mena crisis tracker – december 6th 2021. 2021a.

World Bank. World bank open data, 2021b. https://data.worldbank.org/.

World Health Organization. Covid-19 dashboard, 2020. https://covid19.who.int/. Last accessed: 2022-01-29.

World Travel & Tourism Council. Travel and tourism total contribution to gdp, 2021. https://tcdata360.worldbank.org/indicators/tnt.tot.contrib.gdp.

Young, Lauren E. The psychology of state repression: Fear and dissent decisions in zimbabwe. *American Political Science Review*, 113(1):140–155, 2019.

On the Manipulation of Information by Governments

Ariel Karlinsky and Moses Shayo

Online Appendices

Contents

A Additional Mortality Data **2**

B Estimating Excess Deaths **4**

C Inflating vs. Concealing Mortality **7**

D Appendix Tables & Figures **11**

A Additional Mortality Data

We supplement the data we have collected in WMD by adding additional data for Bahrain, Curacao, India, Iraq, Saudi Arabia, Sri Lanka and Vietnam. We detail the sources for each below:

- *Bahrain* published annual mortality counts for 2015 to 2020. Expected deaths for 2020 were obtained using a trend line across 2015-2019.

- *Curacao* published annual mortality counts for 2015 to 2021. Expected deaths for 2020-2021 were obtained using a trend line across 2015-2019.

- *India* has several estimates of excess mortality, based on large household surveys, the national rural health facility system or state and municipal level vital registration data (Leffler et al., 2022; Anand et al., 2021; Malani and Ramachandran, 2022; Banaji and Gupta, 2022; Jha et al., 2022). These studies differ both on data sources and in methods such as projections and the time frame analyzed. However, two key points emerge: first, that India's 2021 death toll was much higher than 2020 and that excess mortality is much higher than reported Covid mortality in both years. Our estimate for the total excess deaths in India is from Msemburi et al. (2023) as it is the most recent and comprehensive one.

- *Iraq* published annual number of deaths from 2016 to 2021. Expected deaths for 2020-2021 were obtained using a trend line across 2016-2019.

- *Saudi Arabia*'s national statistics office has replied to our request for mortality data with annual numbers of Saudi-nationals deaths from 2015 to 2021. Like other gulf countries, Saudi-Arabia has a large non-citizen resident population and in 2018 Saudi-nationals deaths were 70% of total deaths. Saudi Arabia's reported Covid deaths are for both Saudi-nationals and non-citizens. Thus, we estimated the Misreporting Rate for Saudi-Arabia on Saudi-nationals, which accounted for half of the reported Covid deaths (Alissa et al., 2023). Note that the excess mortality and misreporting for non-citizens in Saudi Arabia is likely higher than for citizens, as has been observed in other gulf countries (Alahmad et al., 2021).

- *Sri Lanka* published annual mortality counts for 2015 to 2021. Expected deaths for 2020-2021 were obtained using a trend line across 2015-2019.

- *Vietnam* published annual mortality data for 2018 to 2021. Expected deaths for 2020-2021 were obtained using a trend line across 2018-2019.

we additionally add excess mortality estimates from Bangladesh, China and Indonesia (Table 3 column 4). For reasons we detail below, we have more limited confidence in these data and thus do not include it in the main analysis.

- *Bangladesh* have released annual reports of their Sample Vital Registration System (SVRS, 2023). The SVRS is a representative sample of more than 1 million respondents in 300 thousand households from which Bangladesh's statistical office estimates annual mortality and fertility indicators. In both 2020 and 2021 – but mostly in 2021 – the SVRS showed a significant increase in mortality. Our confidence in this data is more limited since it is national civil vital registration. Based on these data, we estimate $MRR = 10.1\%$.

- *China* has reported annual number of deaths to the UN. However, these are "...estimated on the basis of the annual National Sample Survey on Population Changes" and not true civil vital registration counts. Based on these data, we estimate $MRR = -0.03\%$ during 2020-21.

- *Indonesia* has provided us with monthly national vital registration data for 2015 to 2021. However, the completeness of this data is very low – about 31% in 2015 – and unclear if remained similar or changed during the pandemic. In addition to this, our limited confidence is also due to discrepancies with other information from Indonesia. The national data shows significant excess mortality in 2021 yet little to none in 2020 in contrast to sub-national vital registration data from Jakarta — which is likely complete — that showed high excess mortality in both years. Based on these data, we estimate $MRR = 3.5\%$.

B Estimating Excess Deaths

To estimate excess mortality, we start by estimating, for each country, the expected counter-factual number of deaths that would have been observed in the absence of Covid (Boka and Wainer, 2020). This forecast is obtained by estimating the following model, separately for each country, using the historical data (2015-2019):

$$Deaths_{ity} = \alpha_{it} + \beta_i Year_y + \epsilon_{ity}. \tag{3}$$

Where $Deaths_{ity}$ is the number of deaths in country i at period t (either week or month, depending on country-specific data availability) in year y; α_{it} are period-fixed-effects (to capture seasonality); and β_i is a (country specific) annual trend in mortality that captures changes in the number of deaths due to population aging, better health-care, improvements in registration, etc. Having estimated the model separately for each country for the years 2015-2019, we then predict the expected mortality for each country in each time period ($\widehat{Deaths}_{i,t}$ for 2020 and 2021).[1]

As a validation test, we use the same procedure to predict 2019 periodical mortality, using only data from 2015 to 2018. Figure B1 show a scatter of the predicted number of deaths against the actual number of deaths. The model performs extremely well, with an $R^2 = 0.997$.[2]

Figure 1 left panel (main text) illustrates the predicted 2020 and 2021 periodical mortality for 4 countries. The thin gray lines show the weekly (France, USA) or monthly (Russia, Uzbekistan) mortality during 2015-2019, with one line per year. The thick black lines shows the expected mortality in 2020-21,[3] based on 2015-2019 data and equation 3. Note that in both the US and France, the expected number of deaths is above the mean of 2015-2019, indicating a positive annual trend. In contrast, in Russia, the expected number of deaths is below the mean, due to decreasing mortality and rising life expectancy in recent years (Kobak, 2021). In Uzbekistan, the estimated trend is small and thus expectation is similar to the mean.

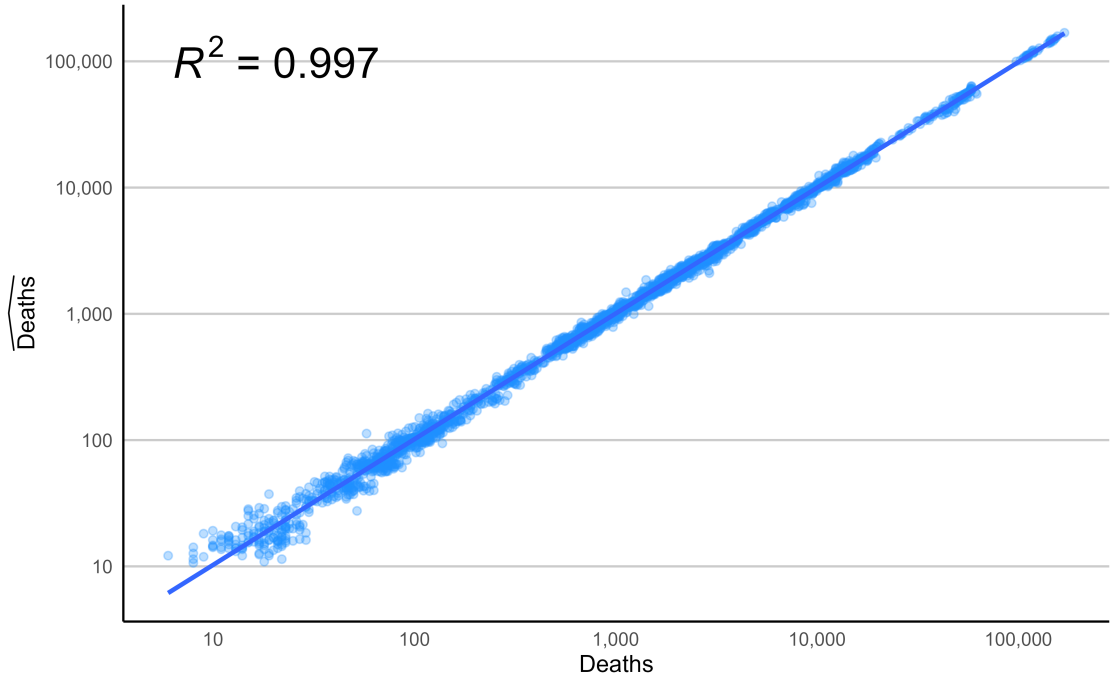
With these expected morality numbers at hand, we define Excess Deaths for country i in period

[1] The expected number of deaths in 2020-2021 is virtually identical when using the Farrington model where applicable (Farrington et al., 1996; Noufaily et al., 2013). The Farrington model is often used in the epidemiological literature but is harder to interpret and requires data at weekly frequency at least.

[2] For example, based on 2015-2018 data, the model predicts 62,554 deaths in the US in the first week of 2019; and the observed deaths in that week was 58,463.

[3] 2021 expected deaths are equal to the 2020 expectation plus the annual trend, and thus hard to distinguish the two in a single plot.

Figure B1: Predicted versus actual mortality in 2019



Notes: The figure shows the WMD model Predicted Deaths in 2019 on the y-axis vs. Actual Deaths in 2019 on the x-axis. Each point is country-period. Axis are log-scaled. Blue line is linear fit. R^2 is from a regression of $\widehat{Deaths}$ on $Deaths$.

t in year y as:

$$Excess\,Deaths_{ity} = Observed\,Deaths_{ity} - \widehat{Deaths}_{ity} \tag{4}$$

where $Observed\,Deaths_{ity}$ is the total number of deaths observed in the country in that period. The observed deaths for 2020 and 2021 are illustrated in by the thick red and blue lines, respectively. The vertical difference between the red/blue to the black lines is thus the Excess Deaths for 2020-2021. Notice that Excess Deaths can be negative (deficit deaths). Reported Covid deaths are shown in dashed red/blue lines.

To compute total Excess Deaths during the Covid pandemic in 2020-21, we sum over the period-level excess deaths from date t_0 to the end of 2021 (T) or latest data-point available, where t_0 is the 10th week of 2020 (March 8th to 14th) for countries with weekly data, and March 2020 for countries with monthly data. That is:

$$Excess\,Deaths_i = \sum_{t_0}^{T} \text{Excess Deaths}_{i,t} \tag{5}$$

Notice that this method also allows for (short-term) "Mortality Displacement" (Islam et al., 2021).

Assume for example that excess deaths of 100 persons were observed in a given week. If in the following week, excess deaths are negative 100, the sum of excess deaths will cancel out. This method also handles countries in which deaths are reported by date of registration rather than by date of occurrence (e.g, UK and Panama), and hence some periods might lack reporting which is back-added in the following periods.

As the sum (equation 5) is a linear combination of the equation 3 coefficients, it is straightforward to also produce an estimate of its variance which combines the uncertainty in the coefficients and that of the error (ϵ_{ity}). This is also known as "Forecast Variance", in contrast to "Prediction Variance" (Abramovich and Ritov, 2013).[4] Out of 134 countries, 112 have a positive and statistically significant excess mortality in 2020-2021.

Figure 1 right panel illustrates the time-pattern of both the cumulative sum of the reported Official Covid Deaths and Excess Deaths for the same four countries as in the figure's left panel. In both France and the US, the curves clearly track each other, while in Russia and Uzbekistan they diverge substantially, resulting in high Misreporting Rates.

[4]See Karlinsky and Kobak (2021) for an explicit derivation.

C Inflating vs. Concealing Mortality

Conditional on capacity, information manipulation can take two distinct forms. Overreporting of Covid deaths might be used to curtail political opposition, attract foreign aid, or justify the government's harsh response to the pandemic. For example, some Israeli opposition leaders argued that reported Covid deaths were inflated, and that the associated restrictions on gatherings were put in place to suppress anti-corruption demonstrations (see reports from Times of Israel (2020); BBC (2020b)). Underreporting of Covid deaths might be used to project competence, both domestically and internationally, enhance trust in the government, or to justify the government's lax response to the pandemic.

The analyses reported in the main text suggest that underreporting is by far the more common strategy. There are almost four times as many countries that undercount Covid deaths as there are overcounters. Table C1 estimates multionmial logit models where the dependent variable is the government's reporting policy—inflate, conceal, or report accurately—and we control for capacity. More precisely, the dependent variable is categorical and takes three values: (i) accurate reporting (the baseline, no statistically significant difference between reported and estimated Covid deaths); (ii) overreporting ($MRR < 0$); and (iii) underreporting ($MRR > 0$).

The explanatory variables aim to capture some of the motivations for information manipulation discussed in the introduction, focusing primarily on possible motivations for overreporting (we examine underreporting in detail in the main text).

- *Demonstrations* from Varieties of Democracy Institute (2021) measures the scope and frequency of mass political events such as demonstrations, protests and riots in 2019. If governments seek to exaggerate the severity of the situation in order to quell dissent, we would expect the likelihood of choosing to overreport to increase with this measure.

- *State Legitimacy* from Fund for Peace (2017) scores the "representativeness and openness of government and its relationship with its citizenry" in 2019. It serves a similar purpose to *Demonstrations* since it scores countries on whether and to what extent "demonstrations or riots or uprisings occurred". However, it also scores other properties such as "openness and fairness of the political process (are elections perceived to be free and fair?)" and thus might capture more than just demonstrations and riots.

- *Foreign Aid* from World Bank (2021b) measures the share of Official Development Assistance (ODA) to gross national income in 2015-2019 (results are unchanged using aid in 2019 only). While not directly indicative of an incentive to signal helplessness during the pandemic, it is indicative of the inclination to request and receive such aid.

- *Covid Stringency* from (Hale et al., 2021) measures the government's early non-pharmaceutical policies to combat Covid, such as travel restrictions, school-closures, social-distancing and masking requirements. We use the value of the index from the beginning of the pandemic (last day of March 2020), to test the hypothesis that governments might overreport Covid deaths to ex-post justify the strong restrictions they imposed. A similar mechanism would suggest underreporting if the initial response was too lax.

Table C1 reports the results. After adjusting for capacity, we find no evidence that domestic unrest is significantly associated with the likelihood of overreporting. In fact, it is positively associated with the likelihood of underreporting Covid deaths. Contrary to expectations, pre-pandemic dependency on foreign aid is *negatively* related to overreporting, as only 4 of the countries that received foreign aid have overreported. The stringency of the initial response to Covid is positively associated with the likelihood of both over and underreporting, but neither association is statistically different from zero.

Table C1: Over or Under Reporting of Covid

Dependent Variable: Reporting Accuracy (Accurate-Reporting (baseline), Over-Reporting and Under-Reporting)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Over-Reporting								
Demonstrations	-0.202	-0.145						
	(0.456)	(0.370)						
State Legitimacy			-0.0487	-0.00263				
			(0.208)	(0.232)				
Foreign Aid (% GNI)					-0.709	-1.851**		
					(0.451)	(0.855)		
Covid Stringency							0.0263	0.0349
							(0.0429)	(0.0374)
Under-Reporting								
Demonstrations	0.229	0.197						
	(0.149)	(0.150)						
State Legitimacy			0.480***	0.611***				
			(0.105)	(0.152)				
Foreign Aid (% GNI)					0.0272	-0.0403		
					(0.145)	(0.139)		
Covid Stringency							0.0217	0.0163
							(0.0150)	(0.0178)
Controls:								
DRC	✓	✓	✓	✓	✓	✓	✓	✓
PWC	✓	✓	✓	✓	✓	✓	✓	✓
SPI		✓		✓		✓		✓
Tax Revenue (% GDP)		✓		✓		✓		✓

Notes: N = 131. Multinomial Logistic regression coefficents. Demonstrations is the scope and frequency of mass political events such as demonstrations, protests and riots in 2019. State Legitimacy scores the "representativeness and openness of government and its relationship with its citizenry", including whether and to what extent "demonstrations or riots or uprisings occurred" in 2019. Foreign Aid measures the share of Official Development Assistance (ODA) to gross national income in 2015-2019. Covid Stringency measurs the non-pharmaceutical policies to combat Covid, such as travel restrictions, school-closures, social-distancing and masking requirements. All regressions include dummies for missing data. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

In summary, while theoretically the direction of information manipulation when facing a crisis can go both ways, in the case of Covid the empirical finding is very sharp: regimes were much more likely to conceal Covid, and there is little evidence for systematic overreporting in countries facing domestic unrest or dependent on foreign aid. This helps shed light on the mechanisms that might lead governments to exaggerate threats. On the one hand, our result suggests a limited tendency to purposefully exaggerate the gravity of a crisis in order to induce a "flight to safety" (Quattrone and Tversky, 1988; Bisbee and Honig, 2022), or to create an atmosphere of emergency that can legitimize repressive measures (Sturm, 1949; Davis and Silver, 2004), consistent with Lowande and Rogowski (2021). On the other hand, this result is not particularly informative regarding the "rally

around the flag" mechanism, as a pandemic is not an outgroup-related threat that can effectively be harnessed to rally the population (although some have tried). Similarly, our result does not rule out attempts to instill fear in order to suppress dissent (Young, 2019), as Covid deaths do not signal the government's repressive capacity.

D Appendix Tables & Figures

Table D1: Correlates of Misinformation: Descriptive Statistics

Variable	N	Min	Mean	SD	Max	Lowest	Highest	Sources
Macroeconomic Incentives								
Tourism	114	0.91	5.65	5.95	38.92	Uzbekistan	Maldives	World Travel & Tourism Council (2021)
FDI	117	-7.16	2.83	7.93	79.52	Cyprus	Malta	World Bank (2021b)
Unemployment	113	0.10	7.08	4.93	25.54	Qatar	South Africa	World Bank (2021b)
Culture								
Interpersonal Trust	98	2.75	24.97	16.10	73.88	Albania	Denmark	Haerpfer et al. (2020); Latinobarometro (2018)
Individualism	93	6.00	41.26	22.11	91.00	Guatemala	United States	Hofstede et al. (2010)
Religiosity	99	14.49	65.08	25.07	100.00	Japan	Egypt	Haerpfer et al. (2020); LAPOP Lab (2019)
Audience Sophistication								
Tertiary Education	97	0.46	17.07	9.64	46.89	Maldives	United Arab Emirates	Barro and Lee (2013)
HDI Education Index	114	0.50	0.76	0.11	0.94	Bhutan	Germany	UNDP (2020)
Internet Access	127	21.96	75.00	17.95	99.70	Tajikistan	Bahrain	World Bank (2021b)
Institutions								
Clean Elections	108	0.00	69.13	27.77	97.50	Palestine	Germany	Varieties of Democracy Institute (2021)
Executive Constraints	108	3.15	68.18	27.20	98.40	Nicaragua	Denmark	Varieties of Democracy Institute (2021)
Freedom of the Press	112	-74.93	-31.07	14.34	-7.82	Vietnam	Norway	Reporters Without Borders (2019)
Political Cycle & Legacies								
Elections during 2020-21	134	0.00	0.57	0.50	1.00			Varieties of Democracy Institute (2021); IFES (2023)
Communist Legacy	134	0.00	0.25	0.44	1.00			Bjørnskov and Rode (2020)
Colonized Legacy	134	0.00	0.54	0.50	1.00			Becker (2019)
Additional Factors								
GDP per Capita (log)	123	8.20	10.18	0.78	11.84	Tajikistan	Liechtenstein	World Bank (2021b)
Population≥65	127	1.17	12.40	6.60	35.62	Qatar	Monaco	World Bank (2021b)

Table D2: Colonial Legacy
Dependent Variable: Misreporting Rate

	(1)	(2)	(3)	(4)
Colonized Legacy	-0.135	-0.0658	-0.318*	-0.0487
	(0.180)	(0.152)	(0.177)	(0.161)
Macroeconomic Incentives (1st PC)		-0.0155		-0.0144
		(0.0542)		(0.0600)
Culture (1st PC)		-0.0328		-0.0262
		(0.108)		(0.124)
Audience Sophistication (1st PC)		-0.284***		-0.322***
		(0.106)		(0.119)
Institutions (1st PC)		-0.354***		-0.359***
		(0.0827)		(0.0893)
Elections during 2020-21		0.417***		0.420***
		(0.137)		(0.140)
Communist Legacy		0.442**		0.403*
		(0.199)		(0.209)
Controls:				
DRC	✓	✓	✓	✓
PWC	✓	✓	✓	✓
SPI			✓	✓
Tax Revenue (% GDP)			✓	✓
R^2	0.108	0.507	0.170	0.513
R^2 Partial	0.00230	0.000492	0.00933	0.000349

Notes: OLS. N=134. All regressions include dummies for missing data. The dependent variable is Misreporting Rate. Colonized Legacy equals 1 for countries that had been colonized (Becker, 2019). R^2 Partial is the R^2 attributable to Colonized Legacy (Huettner and Sunder, 2012). Continuous coefficients are standardized and discrete coefficients are y-standardized. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D3: Correlates of Misinformation with Extended Capacity Controls

Dependent Variable: Misreporting Rate

	Macroeconomic Incentives				Culture			
	International Tourism	Unemployment	Net FDI	1st PC	Interpersonal Trust	Individualism	Religiosity	1st PC
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Std. β	-0.140	-0.0238	-0.00678	-0.0528	-0.260***	-0.281***	0.352***	-0.323***
	(0.113)	(0.0882)	(0.0346)	(0.0594)	(0.0787)	(0.0874)	(0.0779)	(0.0828)
Controls:								
DRC	✓	✓	✓	✓	✓	✓	✓	✓
PWC	✓	✓	✓	✓	✓	✓	✓	✓
SPI	✓	✓	✓	✓	✓	✓	✓	✓
Tax Revenue (% GDP)	✓	✓	✓	✓	✓	✓	✓	✓
R^2	0.177	0.154	0.152	0.168	0.256	0.208	0.283	0.234
R^2 Partial	0.0133	0.000643	0.000464	0.00211	0.0837	0.0743	0.116	0.0974
Adj. p_{val}	1	1	1	1	.0204	.0266	< 0.001	.0024

	Audience Sophistication				Institutions			
	Tertiary Education	HDI Education Index	Internet Access	1st PC	Clean Elections	Executive Constraints	Press Freedom	1st PC
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Std. β	-0.136	-0.328***	-0.340***	-0.417***	-0.470***	-0.576***	-0.437***	-0.551***
	(0.0837)	(0.111)	(0.120)	(0.0930)	(0.117)	(0.0918)	(0.0997)	(0.0898)
Controls:								
DRC	✓	✓	✓	✓	✓	✓	✓	✓
PWC	✓	✓	✓	✓	✓	✓	✓	✓
SPI	✓	✓	✓	✓	✓	✓	✓	✓
Tax Revenue (% GDP)	✓	✓	✓	✓	✓	✓	✓	✓
R^2	0.167	0.206	0.233	0.261	0.299	0.364	0.262	0.342
R^2 Partial	0.0171	0.0675	0.0908	0.117	0.167	0.224	0.143	0.203
Adj. p_{val}	1	.0617	.0867	< 0.001	.0015	< 0.001	< 0.001	< 0.001

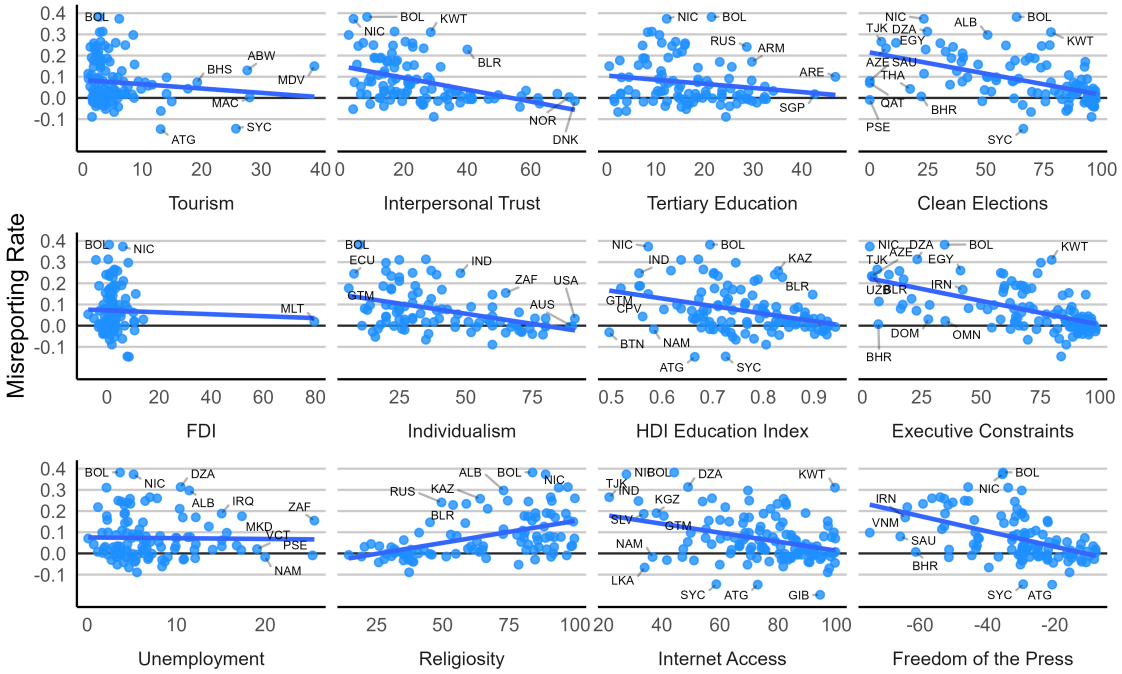
Notes: OLS. N=134. All regressions include dummies for missing data. Coefficients are standardized. R^2 Partial is the R^2 attributable to the shown explanatory variables (Huettner and Sunder, 2012). International Tourism is share of GDP in 2018 (latest avilable year). FDI Net is difference between inbound to outbound foreign direct investment, as share of GDP in 2019. Unemployment is rate in 2019. Interpersonal Trust is share of population who agree that "Most people can be trusted". Individualism is "The degree of interdependence a society maintains among its members". Religiosity is share of people who answer that religion in their life is very or rather important. Tertiary Education is share of population aged 25-64 with completed tertiary education. HDI Education Index is average of mean years of schooling and expected years of schooling in the population. Internet Access is share of population with internet access. Clean Elections measures "To what extent are elections free and fair". Exectuive Constraints is average of "Judicial Constraints" and "Legislative Constraints" on the executive indices. Freedom of the Press measures the degree to which journalists and reporters are able to operate and report freely. 1st PC accounts for 43%, 76%, 74% and 88% of the common variation in domains Macroeconomic Incentives, Culture, Audience Sophistication and Institutions, respectively. Variables sources and descriptive statistics are in Appendix Table D1. Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table D4: Correlates of Misinformation with Extended Capacity Controls (Continued)

	$\frac{\text{Excess - Reported}}{\text{Expected}}$				$\frac{\text{Excess - Reported}}{\text{Population}}$	$\log(\frac{\text{Excess}}{\text{Reported}})$
	(1)	(2)	(3)	(4)	(5)	(6)
Macroeconomic Incentives (1st PC)	-0.0136	-0.0199	-0.00610	-0.0153	-0.0253	0.00275
	(0.0600)	(0.0601)	(0.0632)	(0.0612)	(0.0461)	(0.0463)
Culture (1st PC)	-0.0228	-0.0273	-0.0170	-0.145	-0.137	0.142
	(0.122)	(0.120)	(0.124)	(0.131)	(0.114)	(0.121)
Audience Sophistication (1st PC)	-0.319***	-0.353***	-0.277*	-0.201	-0.149	-0.352***
	(0.119)	(0.120)	(0.146)	(0.127)	(0.119)	(0.126)
Institutions (1st PC)	-0.359***	-0.386***	-0.355***	-0.331***	-0.352***	-0.472***
	(0.0881)	(0.0953)	(0.0891)	(0.0997)	(0.111)	(0.0845)
Elections	0.416***	0.429***	0.419***	0.482***	0.244*	0.279*
	(0.141)	(0.141)	(0.144)	(0.151)	(0.138)	(0.148)
Communist Legacy	0.428**	0.357*	0.388*	0.399**	0.865***	0.310
	(0.192)	(0.213)	(0.201)	(0.180)	(0.213)	(0.193)
Population¿65 (%)		0.135				
		(0.144)				
GDP Per Capita (log)			-0.0726			
			(0.142)			
Controls:						
DRC	✓	✓	✓	✓	✓	✓
PWC	✓	✓	✓	✓	✓	✓
SPI	✓	✓	✓	✓	✓	✓
Tax Revenue (% GDP)	✓	✓	✓	✓	✓	✓
Sample	Main	Main	Main	Extended	Main	Main
Observations	134	134	134	137	134	134
R^2	0.513	0.520	0.515	0.469	0.485	0.464

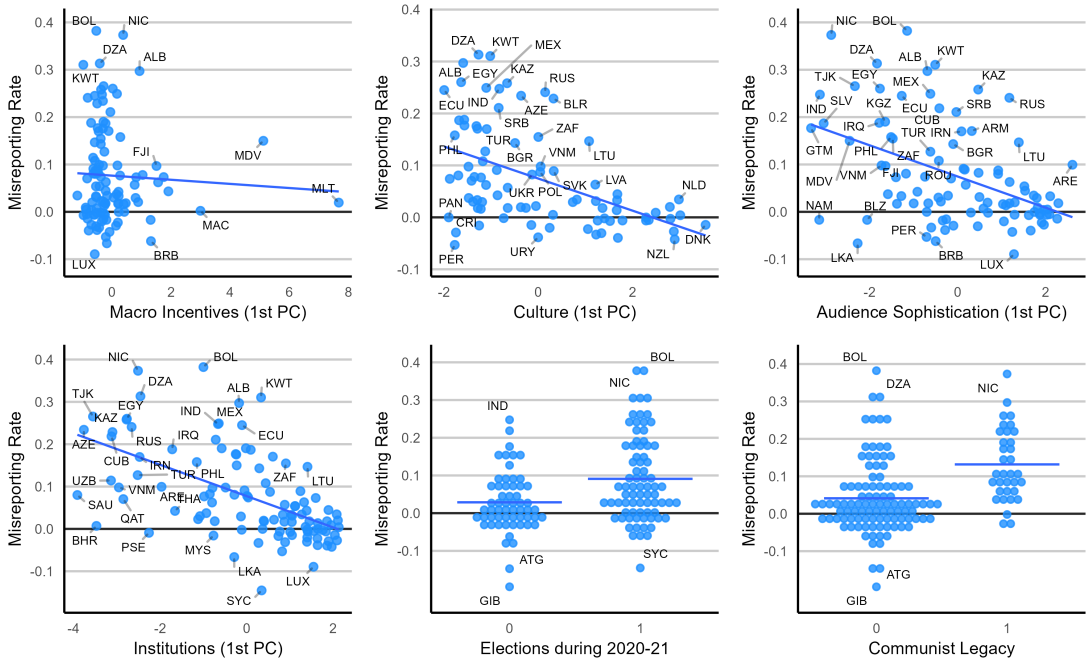
Notes: OLS. All regressions include dummies for missing data. Continuous coefficients are standardized and discrete coefficients are y-standardized. 1st PC accounts for 43%, 76%, 74% and 88% of the common variation in domains Macroeconomic Incentives, Culture, Audience Sophistication and Institutions, respectively. Elections during 2020-21 equals 1 for countries that had national parliamentary or presidential elections in 2020-21. Communist Legacy equals 1 for countries that had a communist or socialist regime (lowest duration is Nicaragua 1979-1989 (Bjørnskov and Rode, 2020)). Square brackets in column 1 show the R^2 Partial attributable to each variable (Huettner and Sunder, 2012). Robust standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Figure D1: Correlates of Misinformation: Factors



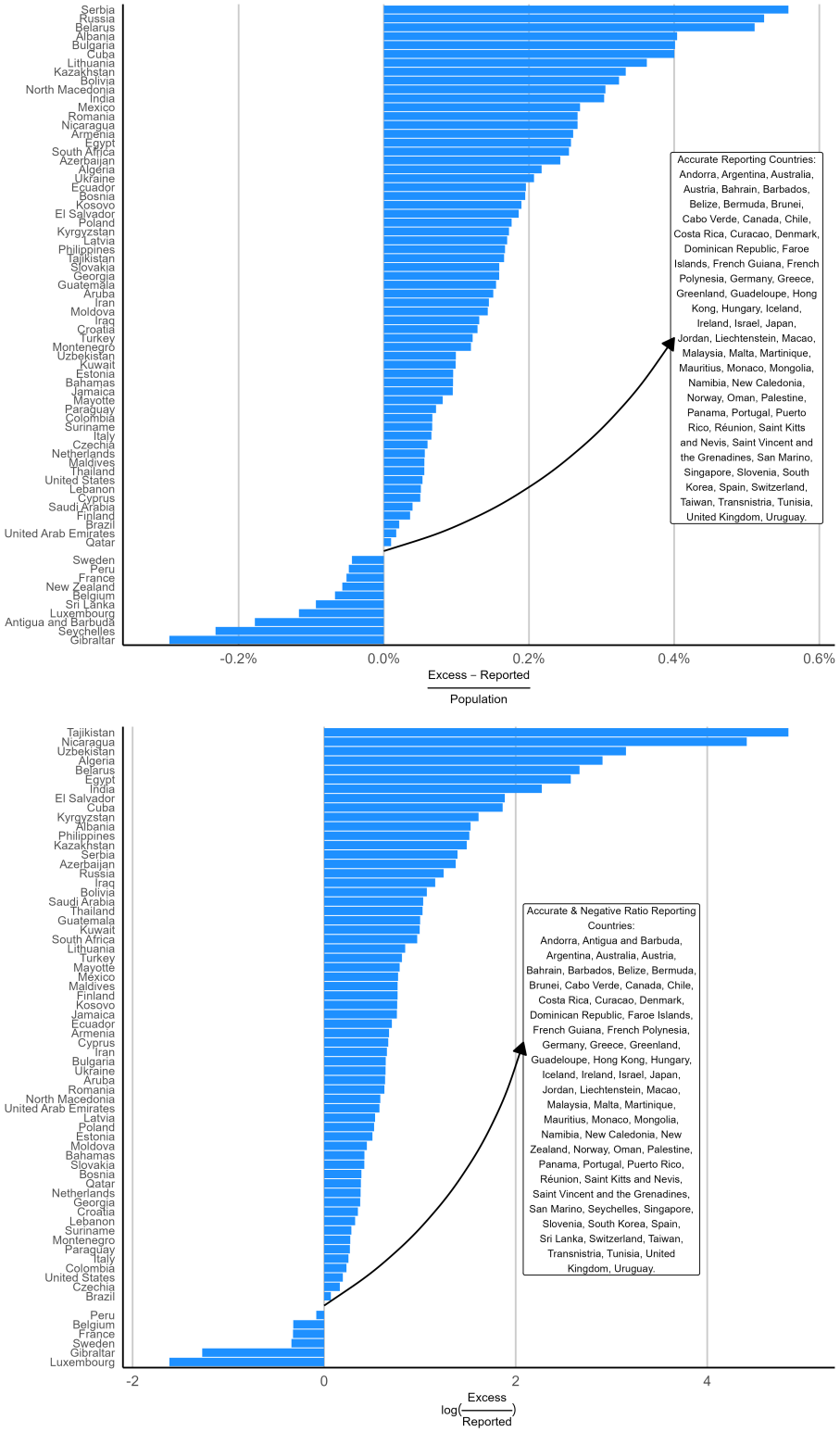
Notes: Raw data. The figure shows the Misreporting Rate (MRR) on the y-axis against the factors from Table 2 on the x-axis. Blue line is linear fit.

Figure D2: Correlates of Misinformation (Continued): Factors



Notes: The figure shows the Misreporting Rate (MRR) on the y-axis against the principal components and variables from Table 3 on the x-axis. In the scatterplots, blue line is linear fit. In the dotplots, blue line is mean MRR by group.

Figure D3: Alternative Covid Misreporting measures by country



Notes: The figure shows alternative Covid Misreporting measures. Top panel normalizes the difference between excess and reported Covid deaths by the exposed population size for the relevant time period. Accurate reporting countries presented in text inset. Bottom panel is the logged ratio between excess and reported Covid deaths. Accurate reporting countries and countries where ratio is negative or zero are shown in text inset.