

# 1 Ranking the effectiveness of worldwide COVID-19 govern- 2 ment interventions

3 Nils Haug<sup>1,2,\*</sup>, Lukas Geyrhofer<sup>2,\*</sup>, Alessandro Londei<sup>3,\*</sup>, Elma Dervic<sup>1,2</sup>, Amélie Desvars-Larrive<sup>2,4</sup>,  
4 Vittorio Loreto<sup>3,2,5</sup>, Beate Pinior<sup>2,4</sup>, Stefan Thurner<sup>1,2,6</sup>, Peter Klimek<sup>1,2,†</sup>

5 <sup>1</sup>*Medical University of Vienna, Section for Science of Complex Systems, CeMSIIS, Spitalgasse 23,*  
6 *1090 Vienna, Austria;*

7 <sup>2</sup>*Complexity Science Hub Vienna, Josefstädter Straße 39, 1080 Vienna, Austria;*

8 <sup>3</sup>*Sony Computer Science Laboratories, 75005 Paris, France;*

9 <sup>4</sup>*Unit of Veterinary Public Health and Epidemiology, Institute of Food Safety, Food Technology*  
10 *and Veterinary Public Health, University of Veterinary Medicine Vienna, Veterinärplatz 1, 1210*  
11 *Vienna, Austria;*

12 <sup>5</sup>*Physics Department, Sapienza University of Rome, 00185 Rome, Italy;*

13 <sup>6</sup>*Santa Fe Institute, Santa Fe, NM 87501, USA.*

14 \* Equal contributions.

15 † Correspondence: [peter.klimek@meduniwien.ac.at](mailto:peter.klimek@meduniwien.ac.at)

16

17 **Non-pharmaceutical interventions (NPIs) to mitigate the spread of SARS-CoV-2 were often**  
18 **implemented under considerable uncertainty and a lack of scientific evidence. Assessing**  
19 **the effectiveness of the individual interventions is critical to inform future preparedness re-**  
20 **sponse plans. Here we quantify the impact of 4,579 NPIs implemented in 76 territories on the**

21 **effective reproduction number,  $R_t$ , of COVID-19. We use a hierarchically coded data set of**  
22 **NPIs and propose a novel modelling approach that combines four computational techniques,**  
23 **which together allow for a worldwide consensus rank of the NPIs based on their effectiveness**  
24 **in mitigating the spread of COVID-19. We show how the effectiveness of individual NPIs**  
25 **strongly varies across countries and world regions, and in relation to human and economic**  
26 **development as well as different dimensions of governance. We quantify the effectiveness of**  
27 **each NPI with respect to the epidemic age of its adoption, i.e., how early into the epidemics.**  
28 **The emerging picture is one in which no one-fits-all solution exists, and no single NPI alone**  
29 **can decrease  $R_t$  below one and that a combination of NPIs is necessary to curb the spread of**  
30 **the virus. We show that there are NPIs considerably less intrusive and costly than lockdowns**  
31 **that are also highly effective, such as certain risk communication strategies and voluntary**  
32 **measures that strengthen the healthcare system. By allowing to simulate “what-if” scenarios**  
33 **at the country level, our approach opens the way for planning the most likely effectiveness of**  
34 **future NPIs.**

## 35 **1 Introduction**

36 In the absence of vaccines and antiviral medication, non-pharmaceutical interventions (NPIs)  
37 implemented in response to epidemic respiratory viruses are the only option to delay and moderate  
38 the spread of the virus in a population [\[1\]](#).

39 Confronted with the worldwide COVID-19 epidemic, most governments implemented bundles  
40 of highly restrictive, sometimes intrusive NPIs. Decisions had to be taken under rapidly changing  
41 epidemiological situations, despite a dramatic lack of scientific evidence on the individual and  
42 combined effectiveness of these measures [\[2\]\[3\]](#), degree of compliance of the population, and societal  
43 impact.

44 This strategy can be compared with hitting the infection curve with a blunt sledgehammer,  
45 hoping that some of the interventions might suppress transmission to a degree that drives the  
46 effective reproduction number,  $R_t$  (the average number of new infections caused by one infected  
47 individual), below one. The lack of data-driven evidence in support for the implemented NPIs may  
48 be best exemplified by the conflicting recommendations on the use of face masks. Both, the World  
49 Health Organization (WHO) and the German Robert Koch Institute (RKI) reversed their initial  
50 recommendations against masks after new evidence appeared, suggesting that optimum use of face  
51 masks is associated with a large reduction in risk of COVID-19 infection [\[4\]](#).

52 Government interventions may cause significant societal and economic damages and may  
53 affect individuals' behaviour, mental health and social security [\[5\]](#). Therefore, knowledge on the

54 most effective NPIs would allow stakeholders to judiciously and timely implement a package of  
55 key interventions to combat a potential resurgence of COVID-19 or any other future respiratory  
56 outbreak. As many countries rolled out several NPIs simultaneously, the challenge of disentangling  
57 the impact of each individual intervention arises.

58 To date, studies of the country-specific progression of the COVID-19 pandemic <sup>[6]</sup> have mostly  
59 explored the independent effects of a single category of interventions. These categories include  
60 travel restrictions <sup>[2,7]</sup>, social distancing <sup>[8-14]</sup>, or personal protective measures <sup>[11]</sup>. Some studies focused  
61 on a single country or even a town <sup>[14-18]</sup>. Some research has combined data from multiple countries  
62 but has pooled NPIs into rather broad categories <sup>[10,19]</sup>, which eventually limits the assessment of  
63 specific, potentially critical NPIs, that may be less costly and more effective than others.

64 Using a comprehensive, hierarchically coded, data set of 4,579 NPIs implemented in 76  
65 territories <sup>[20]</sup>, here we analyse the impact of government interventions on  $R_t$ , using harmonised  
66 results from a new multi-method approach consisting of (i) a case-control analysis (CC), (ii) a step  
67 function approach to LASSO time-series regression (LASSO), (iii) random forests (RF) and (iv)  
68 recurrent neural networks (RNN). We also investigate country-based control strategies as well as  
69 the impact of some selected country-specific metrics. We finally focus on the highly controversial  
70 nation-wide (or state-wide) lockdowns and investigate the relevance of this measure with respect to  
71 the timeliness of its implementation.

72 All approaches (i-iv) yield comparable rankings of the effectiveness of different categories of  
73 NPIs across their hierarchical levels. This remarkable agreement allows us to identify a consensus

74 set of specific NPIs that lead to a significant reduction of  $R_t$ . Further, we evaluate the heterogeneity  
75 of the effectiveness of individual NPIs in different territories. We find that time of implementation,  
76 already implemented measures, different dimensions of governance, as well as human and social  
77 development affect the effectiveness of NPIs to varying degrees.

## 78 **2 Results**

79 **Global approach.** Our results are based on a consolidated version of the CCCSL (CSH COVID19  
80 Control Strategies List)<sup>20</sup>. This data set provides a hierarchical taxonomy of NPIs on four levels:  
81 eight broad themes (level 1, L1) are divided into 53 categories of individual NPIs (level 2, L2) that  
82 include hundreds of subcategories (level 3, L3) and additional group codes (level 4, L4). In Figure 1  
83 we compare the results for the NPIs' effectiveness rankings for the four methods of our approach  
84 (i-iv) on L1 (themes); see Methods for how these ranks were obtained. A clear picture emerges:  
85 the themes of social distancing, healthcare & public health capacity, travel restrictions, and risk  
86 communication score as the top four intervention types in all methods. Social distancing is the  
87 highest ranked theme of NPIs in all but one method.

88 We next compare results obtained on L2 of the NPI data set, i.e., using the 53 individual NPI  
89 categories. A substantial number of interventions has a significant effect on  $R_t$  for the majority  
90 of the methods, see Figure 2 and Table 1. Seven NPI categories show significant impacts on  $R_t$   
91 in all four methods (given the number of significant results in each method, we would expect an  
92 overlap of 1.65 under completely unrelated results) while three out of four methods identify the

<b>L1 themes</b>	<b>CC</b>	<b>LASSO</b>	<b>RF</b>	<b>RNN</b>
Social distancing	1	1	1	3
Healthcare and public health capacity	2	2	4	1
Travel restriction	4	3	2	2
Risk communication	3	4	3	4
Resource allocation	6	5	5	5
Case identification, contact tracing and related measures	5	6	6	7
Returning to normal life	8	8	7	6
Environmental measures	7	7	8	8

Figure 1: Comparison of effectiveness rankings on the coarsest hierarchical level for the case-control analysis (CC), LASSO regression (LASSO), random forest regression (RF), and recurrent neural network analysis (RNN). All methods indicate that NPIs belonging to the themes of social distancing, healthcare & public health capacity, risk communication, and travel restrictions lead to the most significant reductions of  $R_t$  (these themes are ranked within the top four themes in all approaches).

93 same 14 NPI categories as significant. The pairwise Pearson correlations between the methods  
94 range between 0.3 and 0.7, with  $p$ -values between  $p = 0.029$  and  $p < 10^{-8}$ . Similar results hold for  
95 Spearman's correlation. In Table [S2](#) we list the subcategories (L3) belonging to the seven consensus  
96 categories.

97 A normalised score for each NPI category is obtained by rescaling the result of each method  
98 to range between zero (least effective) and one (most effective) and then averaging this score. The  
99 maximal (minimal) NPI score is therefore 100% (0%), meaning that the measure is the most (least)  
100 effective measure in each method. Amongst the seven consensual NPI categories, the largest impacts  
101 on  $R_t$  are displayed by the closure of educational institutions (with a score of 85% and estimates for  
102  $\Delta R_t$  ranging from  $-0.064$  to  $-0.34$ ), small gathering cancellations (72%,  $\Delta R_t$  between  $-0.082$   
103 and  $-0.23$ ) and border restrictions (57%,  $\Delta R_t$  between  $-0.024$  and  $-0.24$ ). We find seven other  
104 NPI categories consensually in three of our methods. These include risk communication activities  
105 to inform and educate the public (65%,  $\Delta R_t$  between  $-0.066$  and  $-0.31$ ) and mass gathering  
106 cancellations (51%,  $\Delta R_t$  between  $-0.032$  and  $-0.27$ ).

107 In Figure [3](#) we visualise the findings on the NPIs' effectiveness in a co-implementation  
108 network [20](#). Nodes correspond to categories (L2) with a size being proportional to their normalised  
109 score. Directed links from  $i$  to  $j$  indicate a tendency that countries implement NPI  $j$  after they  
110 implemented  $i$ . The network therefore illustrates the typical NPI implementation sequence of the 53  
111 countries and the steps within this sequence that contribute most to a reduction of  $R_t$ . For instance,  
112 there is a pattern where countries first cancel mass gatherings before moving on to cancellations

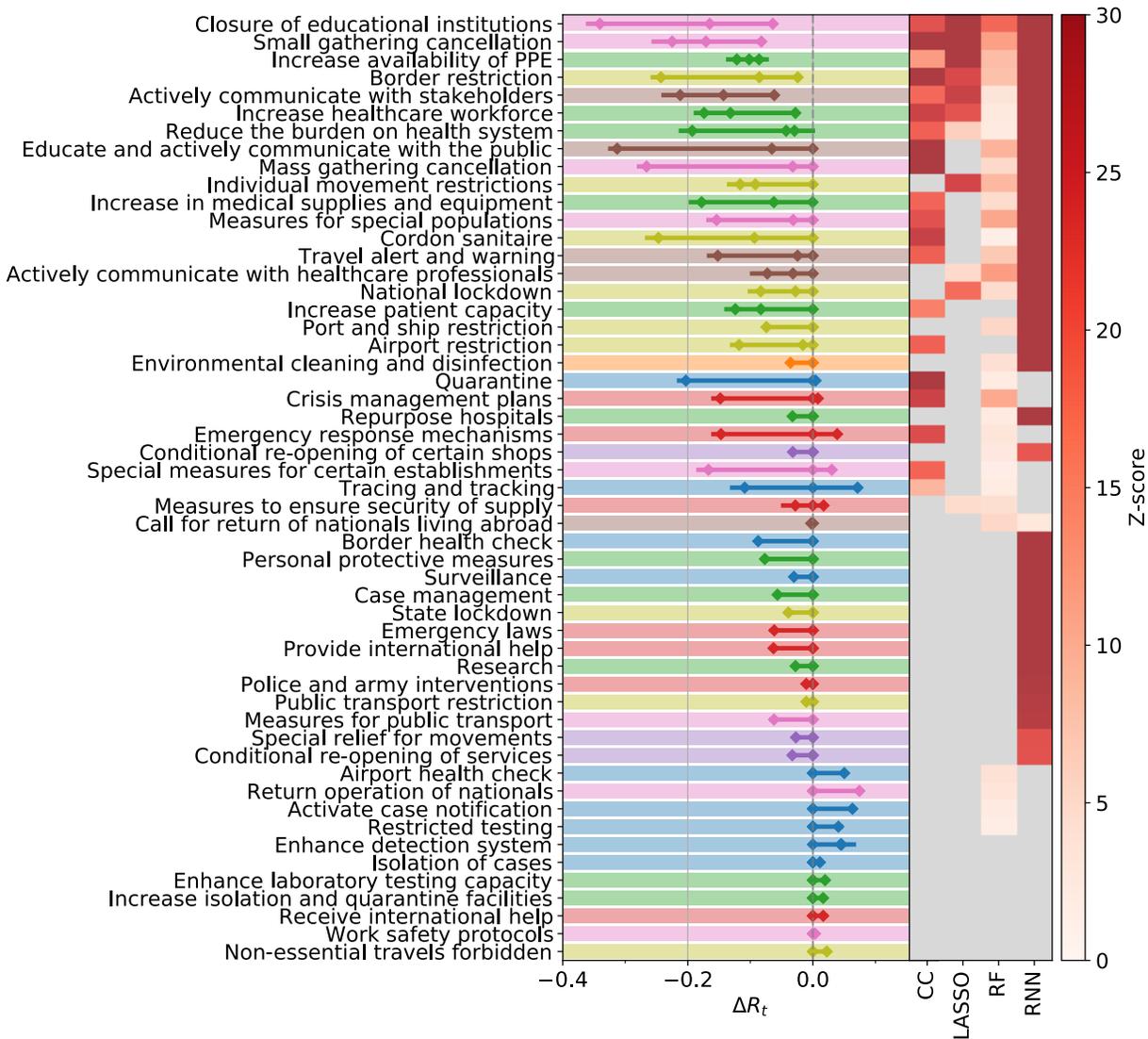


Figure 2: Decrease in the effective reproduction number,  $\Delta R_t$ , for the various NPIs at L2, as quantified by case-control analysis (CC), LASSO, and the recurrent neural network (RNN) regression. The left panel shows the combined 95% confidence interval of  $\Delta R_t$  for the most effective interventions across all included territories. The heatmap in the right panel shows the corresponding Z-scores of the measure effectiveness as determined by the four different methods. NPIs are ranked according to the number of methods agreeing on their impacts, from top (significant in all methods) to bottom (ineffective in all analyses). L1 themes are colour-coded as in Figure 1.

L2 category	Score	Consensus	$\Delta R_t^{CC}$	$\Delta R_t^{LASSO}$	Importance (RF)	$\Delta R_t^{RNN}$
Closure of educational institutions	85%	4	-0.165 (7)	-0.341 (4)	0.033 (3)	-0.064(1)
Small gathering cancellation	72%	4	-0.171 (5)	-0.225 (5)	0.018 (3)	-0.082 (1)
Educate and actively communicate with the public	65%	3	-0.313 (6)	0	0.012 (2)	-0.066 (1)
Border restriction	57%	4	-0.243 (6)	-0.086 (4)	0.010 (2)	-0.242 (7)
Increase availability of personal protective equipment (PPE)	56%	4	-0.086 (6)	-0.122 (3)	0.021 (4)	-0.102 (1)
Mass gathering cancellation	51%	3	-0.266 (6)	0	0.007 (2)	-0.0322 (8)
Measures for special populations	48%	3	-0.154 (7)	0	0.013 (2)	-0.0315 (9)
Increase healthcare workforce	43%	4	-0.174 (7)	-0.132 (6)	0.004 (3)	-0.0279 (9)
Actively communicate with stakeholders	43%	4	-0.143 (7)	-0.212 (8)	0.003 (2)	-0.062 (1)
Individual movement restrictions	43%	3	0	-0.116 (5)	0.032 (6)	-0.092 (1)
Increase in medical supplies and equipment	40%	3	-0.178 (9)	0.000 (2)	0.009 (3)	-0.062 (1)
Reduce the burden on health system	35%	4	-0.193 (9)	-0.043 (6)	0.003 (2)	-0.029 (1)
Cordon sanitaire	35%	3	-0.247 (9)	0	0.002 (1)	-0.093 (2)
National lockdown	30%	3	0	-0.083 (4)	0.010 (3)	-0.028 (1)

Table 1: Comparison of effectiveness rankings on L2. Out of the 53 different NPI categories, all four methods show significant results for seven NPIs (consensus 4); three methods agree on seven further NPIs (consensus 3). We report the average normalized score, the observed reduction in  $R_t$  for the various methods and the NPI importance for the random forest. The numbers in brackets give the amount by which the last digit of the corresponding number outside the brackets fluctuates within the 95% confidence interval.

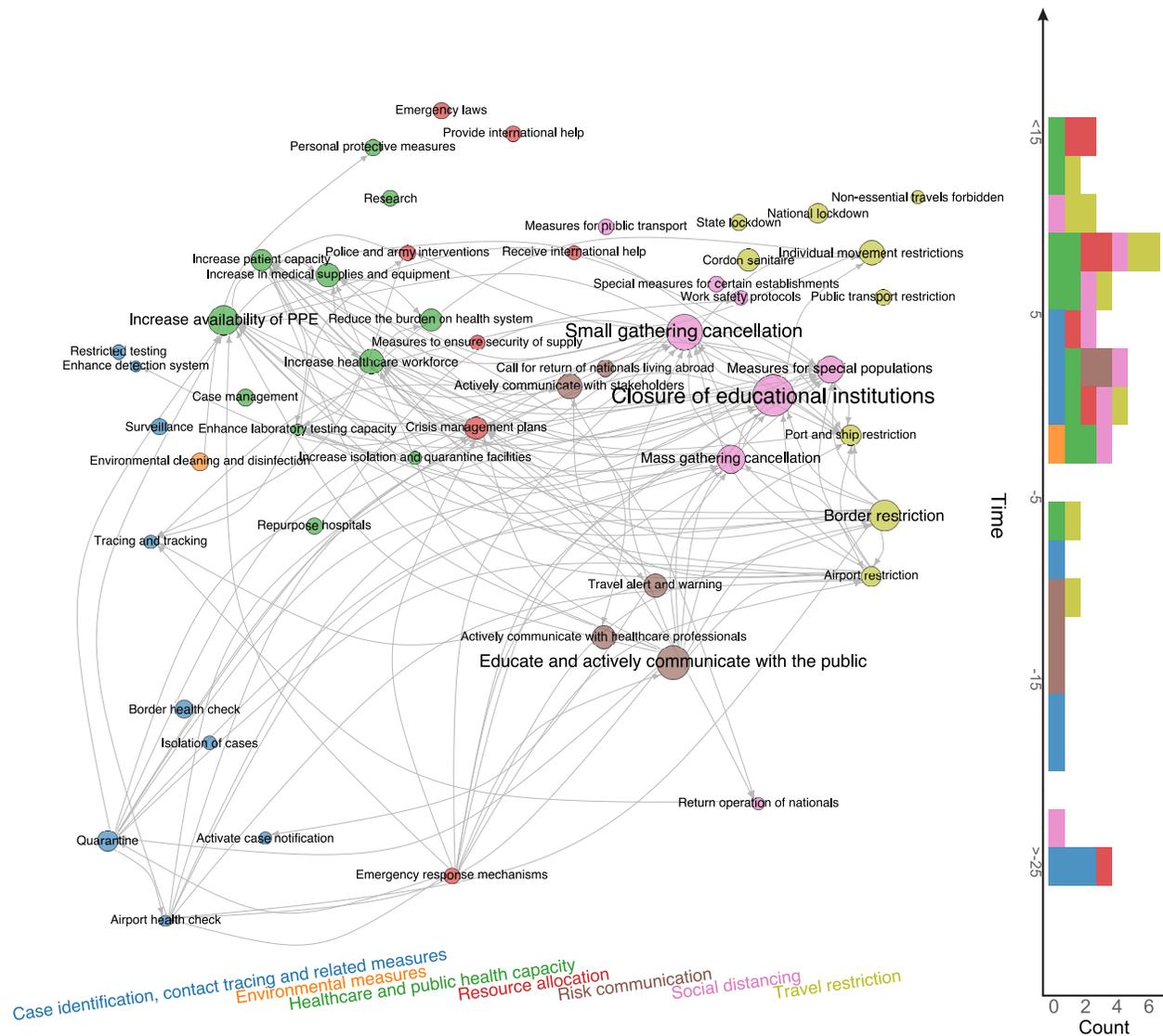


Figure 3: Time-ordered NPI co-implementation network across countries. Nodes are categories (L2) with colour indicating the theme (L1) and size being proportional to the average effectiveness of the intervention. Arrows from nodes  $i$  to  $j$  represent that countries which have already implemented intervention  $i$  tend to implement intervention  $j$  later in time. Nodes are positioned vertically according to their average time of implementation (measured relative to the day where the country reached 30 confirmed cases) and horizontally according to their L1 theme.

113 of specific types of small gatherings, where the latter associates on average with more substantial  
114 reductions in  $R_t$ . Education and active communication is a powerfully effective early measure  
115 (implemented around 15 days before 30 cases were reported). Most social distancing and travel  
116 restriction measures (i.e., closure of educational institutions, work safety protocols, cordon sanitaire,  
117 individual movement restrictions, complete lockdown) are typically implemented within the first  
118 two weeks after reaching 30 cases with varying impacts; see also Figure 2.

119 By analysing the third and fourth levels of the coding hierarchy, the CC approach makes  
120 it possible to assess the effectiveness of NPIs related to the use of face masks. Increasing the  
121 availability of face masks for healthcare professionals shows a significant impact on the  $R_t$  ( $\Delta R_t =$   
122  $-0.07(1)$ ). There is also a considerable effect of communicating with the public and promoting  
123 the use of face masks ( $\Delta R_t = -0.19(2)$ ). However, their mandatory use in public settings shows a  
124 weaker effect ( $\Delta R_t = -0.13(3)$ ). These three mask-related NPIs have typically been implemented  
125 at different times of the epidemic. Promoting the use of face masks started on average three days  
126 after 30 cases were counted in a country, measures to increase their availability in healthcare settings  
127 were implemented 11 days afterwards, and the mandatory use of face masks in public environments  
128 came 22 days later.

129 **Country-level approach.** A sensitivity check of our results with respect to the removal of individ-  
130 ual continents from the analysis also indicates substantial variations between different regions of  
131 the world in terms of NPI effectiveness, see SI. To further quantify how much the effectiveness of  
132 an NPI depends on the particular territory (country or US state) where it has been introduced, we

133 measure the heterogeneity of the NPI rankings in different territories through an entropic approach  
134 in the RNN method; see Methods. Fig. 4 shows the normalised entropy of each NPI category versus  
135 its rank. A value of entropy close to zero implies that the corresponding NPI has a similar rank  
136 relative to all other NPIs in all territories. In other words, the effectiveness of the NPI does not  
137 depend on the specific country or state. On the contrary, a high value of the normalised entropy  
138 signals that the same NPI performs very differently in different regions.

139 The values of the normalised entropies for many NPIs are far from being one and below  
140 the corresponding values obtained through a temporal reshuffling of the NPIs in each country.  
141 The effectiveness of many NPIs therefore is, first, significant and, second, heavily dependent on  
142 the local context, which is a combination of socio-economic features and NPIs already adopted.  
143 We further explore this interplay of NPIs with socio-economic factors by analysing the effects  
144 of demographic and socio-economic covariates, as well as indicators for governance, human and  
145 economic development in the CC method, see Supplementary Information (SI). We find a robust  
146 tendency that NPIs are less effective in countries with high levels of human development (as  
147 quantified by the Human Development Index), governance-related accountability and political  
148 stability (as quantified by World Governance Indicators provided by the World Bank).

149 We then focus on the whether the efficacy of specific NPI depends on its epidemic age, i.e.,  
150 how early it was adopted in each specific country. Without loss of generality, here we focus on the  
151 NPI of a *National lockdown* (or *State lockdown* in the US) and we refer to the SI for the complete  
152 treatment of all the other NPIs. As it is known, the complete lockdown triggered a lot of debates

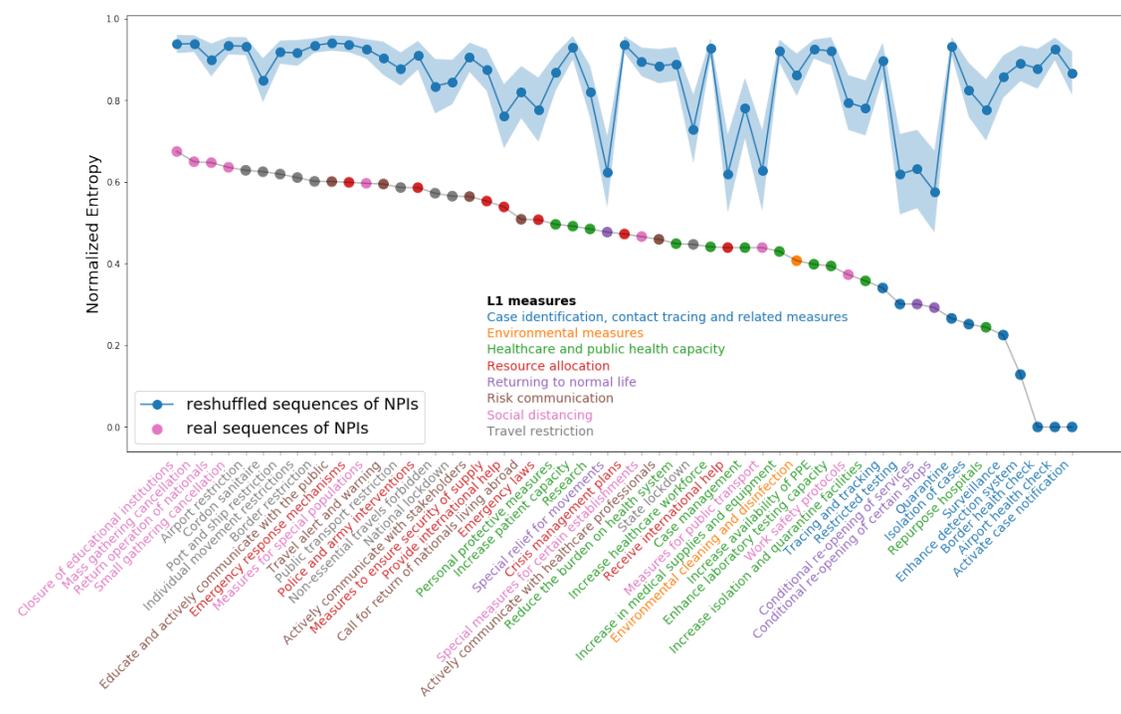


Figure 4: Normalised entropy vs rank for all the NPIs at L2. Each NPI name is coloured according to its theme of belonging (L1) as indicated in the legend. The blue curve represents the same information obtained out of a reshuffled data set of NPIs.

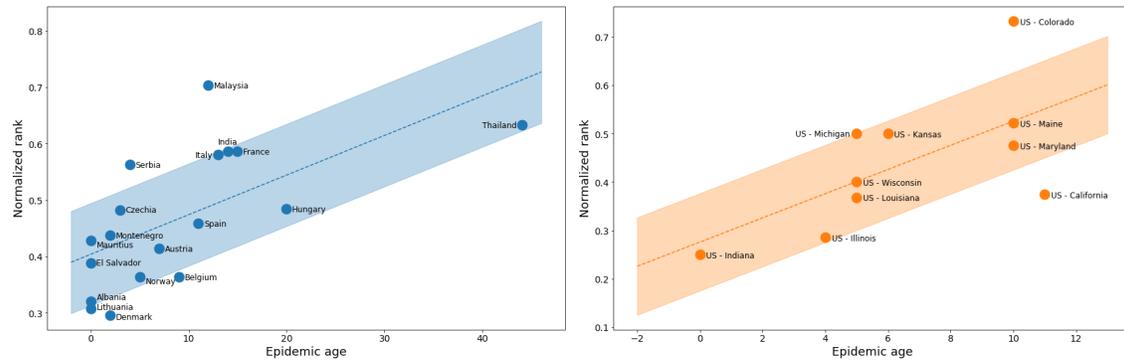


Figure 5: Normalised ranking position of *National lockdown* (left) or *State lockdown* (right) versus the epidemic age (number of days since 30 confirmed cases) of the NPI in each country/state in which it has been adopted. To avoid biases connected to the epidemic age, we compute the ranking from the relative variation of  $R_t$ , i.e.,  $\Delta R_t / R_t$ .

153 about its efficacy and its real necessity, also considering its level of invasiveness. In the RNN  
 154 approach, we assess the relative effectiveness of the lockdown measure in different countries. Here  
 155 relative effectiveness means the relative position in the ranking of a specific country, i.e., the ranking  
 156 position normalised with the number of NPIs adopted in that country. To this end we repeat the  
 157 same knockout procedure adopted above to extract the NPI rankings in the framework of the RNN  
 158 method (see SI) for each country separately and compute, for each country, the normalised ranking  
 159 position of *National lockdown* (or *State lockdown* in the US). Fig. 5 reports the normalised rank in  
 160 each country/state as a function of the epidemic age of *National lockdown* (left) or *State lockdown*  
 161 (right) in that country/state.

162 There is a strong correlation between the effectiveness of the national/state lockdown and the  
 163 epidemic age of its implementation. We can conclude that the epidemic age is highly relevant when

164 national or state lockdown are concerned. A short synthesis for that is "the earlier, the better". We  
165 refer to the Supplementary Information for the report of this analysis applied to all NPIs.

### 166 **3 Discussion**

167 To give a better synthesis, in the following, we discuss our main findings organised according to the  
168 themes at the top level of NPIs hierarchy.

169 **Social distancing.** Bans of small gatherings (gatherings of 50 persons or less) and the closure of  
170 educational institutions have a more substantial effect on  $R_t$  (but are also more intrusive to our daily  
171 lives) than the prohibition of mass gatherings, measures targeting special populations (e.g., elderly,  
172 vulnerable populations, hospitalized patients, prisoners or more exposed non-healthcare profes-  
173 sionals) or adaptive measures for certain establishments (e.g., places of worship, administrative  
174 institutions, entertainment venues, nursing homes). In two recent studies based on smaller numbers  
175 of countries, school closures had been attributed only a little effect on the spread of COVID-19 [9,10](#).  
176 Social distancing measures are less effective in countries with a high population density and a  
177 high degree of citizen participation in the government, as well as freedom of expression or free  
178 media (WGI Voice & Accountability). The country-level analysis confirms that these NPIs have  
179 a particularly high entropy, meaning that their effectiveness varies indeed substantially across  
180 countries. An exception to that are the measures for public transport and work safety protocols,  
181 where the latter mostly refers to mandatory guidelines for, e.g., physical barriers or fever checks at  
182 workplaces. These two social distancing measures have a low effectiveness rank (little significance

183 across the methods) and low entropy, meaning that they had no impact on  $R_t$  consistently across  
184 most countries.

185 **Healthcare and public health capacity.** An increase in the availability of personal protective  
186 equipment (PPE) to the healthcare workforce, together with measures aiming to reduce the num-  
187 ber of non-COVID-19 or non-critical COVID-19 patients in medical centres and hospitals (by  
188 promoting self-isolation of mildly symptomatic patients, setting up health hotlines, etc.) are also  
189 essential building blocks of successful containment strategies. All of these measures combine high  
190 effectiveness of early implementation and low entropy, meaning that they are similarly effective in  
191 most countries. Consequently, they also show less or no consistent correlations with most of the  
192 country-level development or governance indicators. There is one notable exception to this general  
193 trend, namely the increased availability of PPE, which positively and strongly correlates with the  
194 control of corruption. Indeed, there are increased news reports currently on scandals related to  
195 government procurement of PPE [21-23](#).

196 **Travel restrictions.** Different types of travel restrictions also show significant effects, in particular  
197 border restrictions (e.g., border closure, border controls), individual movement restrictions (e.g.,  
198 curfews, the prohibition of non-essential activities) and cordons sanitaires (containment zones).  
199 The high effectiveness of border restrictions is driven by European countries (its impact on  $R_t$   
200 turns insignificant in two of our methods after removing all European countries); most likely for  
201 geographic reasons. This finding is in line with a high entropy score of border, airport, port and ship  
202 as well as individual movement restrictions.

203 Effectiveness of ultimate measures such as stay-at-home orders or lockdowns is still con-  
204 troversial. Recent studies suggest that a national lockdown reduces  $R_t$  by an average of 5%<sup>[9]</sup> to  
205 80%<sup>[10]</sup>, whereas other interventions seem to reduce the virus spread by 5%<sup>[10]</sup> to 30%<sup>[9]</sup>. In some  
206 countries or territories, the effect of a lockdown decided in the late stage of the epidemic may not be  
207 more effective than previously implemented bans on gatherings<sup>[9,10,24]</sup>. Our analysis highlights the  
208 importance of early national lockdowns by showing how the relative effectiveness of that measure  
209 correlates with the epidemic age of its adoption. However, the reduced effectiveness of lockdowns  
210 at higher epidemic age, as observed in Fig. 5, does not necessarily imply that taking this NPI late is  
211 useless.

212 **Risk communication.** In terms of risk communication, we find that pro-active communication with  
213 stakeholders from the private sector (e.g., business owners or chief executive officers) to promote  
214 voluntary safety protocols in enterprises, businesses, event organization, government administrations,  
215 etc., shows a significant effect in each of the four analyses, mainly when implemented early. Three  
216 out of four approaches also indicate a substantial impact of public health communication strategies  
217 (i.e., non-binding NPIs) encouraging citizen engagement and empowering them with information.  
218 The voluntary promotion of the use of face masks is also an efficient risk communication strategy.  
219 To better understand this, we performed an in-depth assessment of the impact of NPIs related  
220 to face masks. Masks are captured in total in three different interventions: the first one aims  
221 to increase their availability to health workers (increase the availability of PPE), which not only  
222 shows a protective effect for the personnel but also contributes to curbing the spread of the disease  
223 by diminishing the role of hospitals as sources of infection,  $\Delta R_t = -0.07(1)$ . The two other

224 mask-related NPIs concern the use of face masks in the general population. Some governments have  
225 actively promoted the wearing of masks through information campaigns on respiratory etiquette.  
226 This measure was typically implemented early in the epidemic (on average, three days after reaching  
227 30 cases). Other countries have enacted the use of masks as mandatory, generally at later stages  
228 of the epidemic (22 days after 30 cases), often accompanied by information campaigns (and  
229 sometimes enforced by police sanctions). The early promotion of the use of face masks shows a  
230 significant effect ( $\Delta R_t = -0.19(2)$ ), whereas late mandatory use appears to have an attenuated  
231 impact ( $\Delta R_t = -0.13(3)$ ). This picture suggests that face masks are useful [4](#)[25](#) in particular when  
232 they are immediately available (as stocks for example) for health workers and the public, making it  
233 possible to promote their general use. Many European countries first had to procure masks before  
234 establishing their usage mandatory to enforce compliance [26](#). This circumstance might have reduced  
235 their effectiveness.

236 **Resource allocation.** Measures for resource allocation show limited impacts on  $R_t$  in our analysis  
237 (e.g., police and army interventions being insignificant in all studies) with relatively high entropy,  
238 meaning that country-level effects are important. Surprisingly, the implementation of crisis manage-  
239 ment plans turns out to be highly effective, except for the Americas. After removing countries from  
240 North and South America from the analyses, all four of our methods agree on significant effects of  
241 crisis management plans with an  $\Delta R_t$  of down to  $-0.3$ , suggesting a lack of *effective* crisis plans  
242 in American territories. For instance, US states had to focus on providing health insurance and  
243 economic stimulus as well as facilitating administrative procedures, while European countries could  
244 develop their plans on top of a stronger socio-economic basis [27](#)[28](#). Crisis management plans are

245 also more effective in countries with a non-participatory government, meaning that countries with  
246 increasingly authoritarian practices might be at an advantage at implementing such policies, as can  
247 be seen in the swift response of Singapore <sup>29</sup>.

248 **Case identification, contact tracing and related measures.** NPIs related to case identification  
249 and contact tracing show some of the lowest effectiveness ranks and in some cases even increase  $R_t$ ,  
250 consistently across most countries (NPIs with the five lowest entropy scores all belong to this theme).  
251 This result is to be expected, as, e.g., increased testing and faster contact tracing will on the short-run  
252 increase the numbers of found cases in return for reduced numbers in the long run. We do not assess  
253 such long-term effects (over timespans of more than a month) in the current work. Furthermore,  
254 note that our analysis considers mostly data from March and April 2020 where many countries  
255 experienced surges of case numbers that most likely hindered effective contact tracing and other  
256 case identification measures. This also applies to the relative ineffectiveness of quarantining people  
257 who either are infected or were exposed to infected persons, while the promotion of self-isolation of  
258 people with symptoms belonged to one of the most effective NPI. This result confirms a tendency  
259 in our results where voluntary measures are more effective than similar mandatory ones.

260 **Strengths & Limitations.** The assessment of the effectiveness of NPIs is statistically challenging,  
261 as measures were typically implemented simultaneously and because their impact might well depend  
262 on the particular implementation sequence. Our four methodological approaches are susceptible to  
263 resulting biases in different ways. While some methods might be prone to overestimating effects  
264 from an NPI due to insufficient adjustments for confounding effects from other measures, other

265 methods might underestimate the contribution of an NPI by assigning its impact to a highly correlated  
266 NPI. As a consequence, estimates of  $\Delta R_t$  might vary substantially across different methods, whereas  
267 the agreement on the significance of individual NPIs is much more pronounced. The strength of our  
268 study, therefore, lies in the harmonization of these four independent methodological approaches,  
269 combined with the usage of an extensive data set on NPIs. This study design substantially reduces  
270 the risk of idiosyncratic results due to model biases. Moreover, whereas previous studies often  
271 subsumed a wide range of social distancing and travel restriction measures under a single entity, our  
272 analysis contributes to a more fine-grained understanding of each NPI.

273 It is also essential to highlight the limitations of our approach. The CCCSL data set features  
274 non-homogeneous data quality and completeness across the different territories and data collection  
275 could be biased by the data collector (native versus non-native) as well as the information communi-  
276 cated by governments. Moreover, the coding system presents some drawbacks, notably because  
277 some interventions could belong to more than one category but are only recorded once. Compliance  
278 with NPIs is crucial for their effectiveness, yet we assumed a comparable degree of compliance by  
279 each population. Additionally, we neither took into account the stringency of NPI implementation  
280 nor potential variations of NPI effectiveness over time, except for the dependency on the epidemic  
281 age of its adoption.

282 To compute  $R_t$ , we used time-series of the number of confirmed COVID-19 cases<sup>30</sup>. This  
283 approach may over-represent patients with severe symptoms and may be biased by variations in  
284 testing and reporting policies among countries. We assumed a constant serial interval (average

285 time-span between primary and secondary infection). However, this number shows considerable  
286 variations in the literature<sup>[31]</sup> and depends on measures such as social distancing and self-isolation.

287 Our analysis provides a large-scale quantification of the effectiveness of different NPIs in  
288 mitigating the transmission of COVID-19. The consensus displayed by the four different methods  
289 guarantees the robustness of the results.

## 290 4 Conclusions

291 Here we presented the outcome of an extensive analysis on the impact of 4,579 individual NPIs  
292 ( $N = 53$ ) on the effective reproduction number  $R_t$  of COVID-19 in 76 territories worldwide. The  
293 adoption of the CCCSL data set<sup>[20]</sup> of NPIs, makes of our study the largest on NPI effectiveness to  
294 date<sup>[10,19,32,33]</sup>.

295 Our study dissects the entangled packages of NPIs<sup>[20]</sup> and gives new insights into their effectiveness.  
296 Our findings suggest that there is no silver bullet to efficiently reduce the burden of a potential  
297 second COVID-19 wave or any similar future viral respiratory epidemics through NPIs. Instead, we  
298 identify several decisive interventions that significantly contribute to reducing  $R_t$  below one, though  
299 none of these interventions alone would be enough to stop the epidemic. At the global level, our  
300 results reveal a worldwide consensus rank of the NPIs based on their effectiveness in mitigating the  
301 spread of COVID-19. For instance, i) social distancing, ii) travel restrictions, and iii) healthcare and  
302 public health capacity (e.g., reduce the burden on the healthcare system by encouraging self-initiated  
303 isolation of people symptoms) have a particularly strong effect on the reduction of  $R_t$ . This finding

304 is in line with results from other studies which used different statistical approaches, epidemiological  
305 metrics, geographic coverage, and classifications of NPIs [\[27-13\]](#). More importantly, the possibility  
306 to focus the analysis on individual countries, reveals that the efficacy of individual NPIs strongly  
307 varies across countries and world regions, and in relation to human and economic development as  
308 well as different dimensions of governance. This finding is significant because it implies that the  
309 impact of specific NPIs in a particular area is the complex outcome of the non-trivial combination  
310 of socio-economic features and the NPIs already adopted. The dependency on the local context  
311 is not surprising, and it points to a high heterogeneity across countries and a non-independence  
312 among the different NPIs. Another source of complexity is represented by the dependency on the  
313 effectiveness of each NPIs on the epidemic age of its adoption. In simple words, the same NPI in  
314 the same country can have a drastically different impact if taken early or later on.

315         The emerging picture is one in which no one-fits-all solution exists, and no single NPI alone  
316 can decrease  $R_t$  below one. Instead, in the absence of a vaccine, a resurgence of COVID-19 cases  
317 can only be stopped by a suitable combination of NPIs, each tailored to the specific country and  
318 its epidemic age. These measures must be enacted together — and with the best timing — to be  
319 maximally effective on the spread of COVID-19 and enable a faster re-opening.

320         One of the consequences of this picture is that NPIs considerably less intrusive and less  
321 costly than lockdowns could be highly effective, such as specific risk communication strategies  
322 and voluntary measures that strengthen the healthcare system. In fact, we found several instances  
323 where non-binding governmental guidelines strongly outperformed their mandatory counterparts

324 (e.g., masks, self-isolation, work safety). These results call for a strong effort to simulate "what-if"  
325 scenarios at the country level for planning the most likely effectiveness of future NPIs, and, thanks  
326 to the possibility to go down to the level of single countries and specific local contexts, our approach  
327 is the first contribution to this end.

## 328 **References**

- 329 1. N. L. Qualls *et al.*, "Community mitigation guidelines to prevent pandemic influenza — United  
330 States, 2017," *MMWR Recommendations and reports*, vol. 66, no. 1, 2017.
- 331  
332 2. H. Tian, Y. Liu, Y. Li, C. H. Wu, B. Chen, *et al.*, "An investigation of transmission control  
333 measures during the first 50 days of the COVID-19 epidemic in China," *Science*, vol. 368,  
334 no. 6491, pp. 638–642, 2020.
- 335 3. S. Chen, J. Yang, W. Yang, , C. Wang, and T. Bärnighausen, "COVID-19 control in China  
336 during mass population movements at New Year," *The Lancet*, vol. 395, no. 10226, pp. 764–766,  
337 2020.
- 338 4. D. K. Chu *et al.*, "Physical distancing, face masks, and eye protection to prevent person-to-  
339 person transmission of SARS-CoV-2 and COVID-19: a systematic review and meta-analysis,"  
340 *Lancet*, vol. 6736, no. 20, pp. 1–15, 2020.
- 341 5. B. Pfefferbaum and C. S. North, "Mental health and the covid-19 pandemic," *New England*  
342 *Journal of Medicine*, vol. 0, no. 0, p. null, 0.

- 343 6. Johns Hopkins University of Medicine, “COVID-19 dashboard by the Center for Systems  
344 Science and Engineering (CSSE) at Johns Hopkins University of Medicine.” (Accessed:  
345 2020-06-04).
- 346 7. M. Chinazzi *et al.*, “The effect of travel restrictions on the spread of the 2019 novel coronavirus  
347 (COVID-19) outbreak.,” *Science*, vol. 368, no. 6489, pp. 395–400, 2020.
- 348 8. A. Arenas, W. Cota, C. Granell, and B. Steinegger, “Derivation of the effective reproduction  
349 number R for COVID-19 in relation to mobility restrictions and confinement,” *medRxiv*, 2020.
- 350 9. N. Banholzer *et al.*, “Estimating the impact of non-pharmaceutical interventions on documented  
351 infections with COVID-19: A cross-country analysis,” *medRxiv*, 2020.
- 352 10. S. Flaxman *et al.*, “Estimating the effects of non-pharmaceutical interventions on COVID-19 in  
353 Europe,” *Nature*, Jun 2020.
- 354 11. A. Teslya *et al.*, “Impact of self-imposed prevention measures and short-term government  
355 intervention on mitigating and delaying a COVID-19 epidemic,” *medRxiv*, 2020.
- 356 12. J. Wang, K. Tang, K. Feng, and W. Lv, “When is the COVID-19 Pandemic Over? Evidence  
357 from the Stay-at-Home Policy Execution in 106 Chinese Cities,” *SSRN*, 2020.
- 358 13. J.-P. R. Soucy *et al.*, “Estimating effects of physical distancing on the covid-19 pandemic using  
359 an urban mobility index,” *medRxiv*, 2020.

- 360 14. K. Prem, Y. Liu, *et al.*, “The effect of control strategies to reduce social mixing on outcomes  
361 of the COVID-19 epidemic in Wuhan, China: a modelling study,” *The Lancet Public Health*,  
362 vol. 5, no. 5, pp. e261–e270, 2020.
- 363 15. M. U. Kraemer *et al.*, “The effect of human mobility and control measures on the COVID-19  
364 epidemic in China,” *Science*, vol. 497, no. 6490, pp. 493–497, 2020.
- 365 16. M. Gatto *et al.*, “Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency  
366 containment measures,” *PNAS*, vol. 117, no. 19, pp. 10484–10491, 2020.
- 367 17. L. Lorch, W. Trouleau, S. Tsirtsis, A. Szanto, B. Schölkopf, and M. Gomez-Rodriguez, “A spa-  
368 tiotemporal epidemic model to quantify the effects of contact tracing, testing, and containment,”  
369 *arXiv*, 2020.
- 370 18. J. Dehning, J. Zierenberg, *et al.*, “Inferring change points in the spread of COVID-19 reveals  
371 the effectiveness of interventions,” *Science*, vol. 9789, no. May, pp. 1–15, 2020.
- 372 19. S. Hsiang *et al.*, “The effect of large-scale anti-contagion policies on the COVID-19 pandemic,”  
373 *Nature*, 2020.
- 374 20. A. Desvars-Larrive *et al.*, “CCCSL: Complexity Science Hub Covid-19 Control Strategies List.”  
375 <https://github.com/amel-github/covid19-interventionmeasures>,  
376 2020. Accessed: 2020-05-15.
- 377 21. Bloomberg, “Slovene economy minister survives ouster bid over ppe scandal,” 2020. Accessed:  
378 2020-07-02.

- 379 22. India Today, “Whiff of a scandal,” 2020. Accessed: 2020-07-02.
- 380 23. The Telegraph, “Ppe: Government counted each glove as single item to reach one billion total,  
381 investigation shows,” 2020. Accessed: 2020-07-02.
- 382 24. R. A. J. Post, M. Regis, Z. Zhan, and E. R. van den Heuvel, “How did governmental interven-  
383 tions affect the spread of covid-19 in european countries?,” *medRxiv*, 2020.
- 384 25. C. J. Worby and H.-H. Chang, “Face mask use in the general population and optimal resource  
385 allocation during the COVID-19 pandemic,” *medRxiv*, 2020.
- 386 26. Al Jazeera News, “Which countries have made wearing face masks compulsory?,” 2020.  
387 Accessed: 2020-06-10.
- 388 27. Business Insider, “14 countries that are paying their workers during quarantine,” 2020. Ac-  
389 cessed: 2020-07-03.
- 390 28. BBC News, “Coronavirus bailouts: Which country has the most generous deal?,” 2020. Ac-  
391 cessed: 2020-07-03.
- 392 29. J. E. Wong, Y. S. Leo, and C. C. Tan, “Covid-19 in singapore—current experience: critical  
393 global issues that require attention and action,” *Jama*, vol. 323, no. 13, pp. 1243–1244, 2020.
- 394 30. Johns Hopkins University of Medicine, “2019 novel coronavirus COVID-19 (2019-nCoV)  
395 data repository by Johns Hopkins CSSE.” [https://github.com/CSSEGISandData/  
396 COVID-19](https://github.com/CSSEGISandData/COVID-19). Accessed: 2020-05-20.

- 397 31. J. Griffin *et al.*, “A rapid review of available evidence on the serial interval and generation time  
398 of COVID-19,” *medRxiv*, 2020.
- 399 32. J. M. Brauner, S. Mindermann, M. Sharma, *et al.*, “The effectiveness and perceived burden of  
400 nonpharmaceutical interventions against COVID-19 transmission : a modelling study with 41  
401 countries,” *medRxiv*, 2020.
- 402 33. C. Cheng, J. Barceló, A. S. Hartnett, R. Kubinec, and L. Messerschmidt, “Covid-19 government  
403 response event dataset (corononet v. 1.0),” *Nature Human Behaviour*, pp. 1–13, 2020.
- 404 34. S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural computation*, vol. 9,  
405 no. 8, pp. 1735–1780, 1997.
- 406 35. A. Cori, N. M. Ferguson, C. Fraser, and S. Cauchemez, “A new framework and software to  
407 estimate time-varying reproduction numbers during epidemics,” *American Journal of Epidemi-*  
408 *ology*, vol. 178, no. 9, pp. 1505–1512, 2013.
- 409 36. F. Valka and C. Schuler, “Estimation and interactive visualization of the time-varying reproduc-  
410 tion number  $R_t$  and the time-delay from infection to estimation,” *preprint*, 2020.
- 411 37. “World population prospects 2019.” <https://population.un.org/wpp/>  
412 [DataQuery/](#). Accessed: 2020-04-24.
- 413 38. “World bank international comparison program database.” [https://data.worldbank](https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD)  
414 [org/indicator/NY.GDP.PCAP.PP.CD](#). Accessed: 2020-04-24.

- 415 39. P. Conceicao *et al.*, *Human Development Report 2019*. United Nations Development Pro-  
416 gramme, 2019.
- 417 40. “World governance indicators.” [https://info.worldbank.org/governance/  
418 wgi/](https://info.worldbank.org/governance/wgi/). Accessed: 2020-04-24.
- 419 41. R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical  
420 Society: Series B (Methodological)*, vol. 58, no. 1, pp. 267–288, 1996.
- 421 42. J. Taylor and R. J. Tibshirani, “Statistical learning and selective inference,” *Proceedings of the  
422 National Academy of Sciences*, vol. 112, no. 25, pp. 7629–7634, 2015.
- 423 43. T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining,  
424 Inference, and Prediction*. Springer, 2009.
- 425 44. F. Pedregosa *et al.*, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning  
426 Research*, vol. 12, pp. 2825–2830, 2011.

## 427 **5 Methods**

### 428 **Data**

429 **NPI data.** We use the publicly available Complexity Science Hub Vienna COVID-19  
430 Control Strategies list (CCCSL) data set on NPIs<sup>[20]</sup>. Therein, NPIs are categorised using a four-  
431 level hierarchical coding scheme: L1 defines the theme of the NPI: “Case identification, contact

432 tracing and related measures”, “Environmental measures”, “Healthcare and public health capacity”,  
433 “Resource allocation”, “Returning to normal life”, “Risk communication”, “Social distancing” and  
434 “Travel restriction”. Each L1 (theme) is composed of several categories (L2 of the coding scheme),  
435 that contain subcategories (L3) which are further subdivided to group codes (L4). The data set  
436 covers 52 countries; data for the USA is available at the state level (24 states). This makes a total of  
437 76 territories. The time window for which data is available varies by territory. In this analysis, we  
438 use a static version of the CCCSL, retrieved on 15 May 2020, presenting 4,579 NPIs. A glossary  
439 of the codes is provided on [github](#). For each country, we use the data until the day to which the  
440 measures have been reliably updated. NPIs that have been implemented in less than five territories  
441 are not considered, leading to a final number of 3,850 NPIs of 53 different L2 categories to be used  
442 in the analyses.

443 **COVID-19 case data.** To estimate the effective reproduction number  $R_t$ , and growth rates  
444 of the number of COVID-19 cases, we use time series of the number of confirmed COVID-19 cases  
445 in the 76 considered territories <sup>30</sup>. To control for weekly fluctuations, we smooth the time series  
446 by computing the rolling average using a Gaussian window with a standard deviation of two days,  
447 truncated at a maximum window size of 15 days.

448 **Regression techniques.** We apply four different statistical approaches to quantify the impact of a  
449 NPI  $M$  on the reduction of  $R_t$  (see details in the Supplementary Information).

450       **Case-control analysis.** The case-control analysis (CC) considers each single category (L2)  
451 or subcategory (L3)  $M$  separately and compares the difference  $\Delta R_t$  in the  $R_t$  between all countries  
452 that implemented  $M$  (cases) with those that did not implement it (controls) during the observation  
453 window. The comparison is made via a linear regression model adjusting for (i) epidemic age (days  
454 after the country has reached 30 confirmed cases), (ii) the value of  $R_t$  before  $M$  takes effect, (iii)  
455 total population, (iv) population density, (v) the total number of NPIs implemented and (vi) number  
456 of NPIs implemented in the same category as  $M$ . With this design, we investigate the time delay of  
457  $\tau$  days between implementing  $M$  and observing  $\Delta R_t$ , as well as additional country-based covariates  
458 that quantify other dimensions of governance and human and economic development. Estimates for  
459  $R_t$  are averaged over delays between 1 and 28 days.

460       **Step function Lasso regression.** In this approach, we assume that without any intervention,  
461 the reproduction factor is constant and deviations from this constant are caused by a time-delayed  
462 onset of each NPI on L2 (categories) of the hierarchical data set. We use a Lasso regularization  
463 approach combined with a meta parameter search to select a reduced set of NPIs that best describe  
464 the observed  $\Delta R_t$ . Estimates for the changes of  $\Delta R_t$  attributable to NPI  $M$  are obtained from  
465 country-wise cross-validation.

466       **Random forest regression.** We perform a random forest (RF) regression, where the NPIs  
467 implemented in a country are used as predictors for  $R_t$ , time-shifted  $\tau$  days into the future. Here,  $\tau$   
468 accounts for the time delay between implementation and onset of the effect of a given NPI. Similar

469 to the Lasso regression, the assumption underlying the random forest approach is that without  
470 changes in interventions, the effective reproduction number in a country remains constant. But  
471 contrary to the two methods described above, the random forest represents a nonlinear model,  
472 meaning that the effects of individual NPIs on  $R_t$  do not need to add up linearly. The importance of  
473 a NPI is defined as the decline in the predictive performance of the random forest on unseen data if  
474 the data concerning that NPI is replaced by noise, also called permutation importance.

475 **Recurrent Neural Networks Modeling.** Recurrent neural networks (RNN) built using  
476 Long Short-Term Memory (LSTM) cells<sup>34</sup> have proven to be suitable models for dynamic processes  
477 such as epidemic propagation due to their ability to recall past events. Here we use a RNN with  
478 identical input as described for the random forest regression, along with the values of  $R_t$ . The  
479 best performing network (least mean square error in country-wise cross-validation) is identified  
480 as having two hidden layers of 100 neurons, one output described by a linear output layer, and 54  
481 inputs (corresponding to each category and  $R_t$ ). To quantify the impact of a measure  $M$  on  $R_t$ , we  
482 use the trained RNN as a predictive model and compare simulations without any measure (reference)  
483 to simulations where one measure is presented at a time to assess  $\Delta R_t$ . To reduce overfitting effects,  
484 we report results from an ensemble of RNNs trained to similar precision levels.

485 **Estimation of the effective reproduction number.** We use the R package EpiEstim<sup>35</sup> with a  
486 sliding time window of 7 days to estimate the time series of the effective reproduction number  $R_t$   
487 for every country. We choose an uncertain serial interval following a probability distribution with a  
488 mean of 4.46 days and a standard deviation of 2.63 days<sup>36</sup>.

489 **Ranking of NPIs.** For each of the four methods (CC, Lasso regression and RNN), we rank the NPI  
490 categories in descending order according to their impact, i.e., the estimated degree to which they  
491 lower  $R_t$  or their feature importance (RF). To obtain a ranking of the eight different themes (L1)  
492 of NPIs, we sum the impacts of the 5 highest ranked categories of each theme and then rank the  
493 themes according to this cumulative impact.

494 **Co-implementation network.** We construct the NPI co-implementation network as previously  
495 described [20](#). If there is a statistical tendency that a country implementing NPI  $i$  also implements  
496 NPI  $j$  later in time, we draw a directed link from  $i$  to  $j$ . Nodes are placed on the  $y$ -axis according to  
497 the average epidemic age at which the corresponding NPI is implemented; they are grouped on the  
498  $x$  axis by their L1 theme. Node colours correspond to themes. The effectiveness scores for all NPIs  
499 are rescaled between zero and one for each method; node size is proportional to the rescaled scores,  
500 averaged over all methods.

501 **Entropic country-level approach.** Each territory can be characterised by its socio-economic  
502 conditions and the unique temporal sequence of NPIs adopted. To quantify the NPI effect, we  
503 measure the heterogeneity of the overall rank of a NPI amongst the countries that have taken that  
504 NPI. To compare countries which have implemented different numbers of NPIs, we consider the  
505 normalised rankings, where the ranking position is divided by the number of elements in the ranking  
506 list (i.e., the number of NPIs taken in a specific country). We then bin the interval  $[0, 1]$  of the  
507 normalised rankings into 10 subintervals and compute for each NPI the entropy of the distribution

508 of occurrences of the NPI in the different normalised rankings per country:

$$S(NPI) = -\frac{1}{\log(10)} \sum_i p_i \log(p_i), \quad (1)$$

509 where  $p_i$  is the probability that the considered NPI appeared in the  $i$ -th bin in the normalised  
510 rankings of all countries. To assess the confidence of these entropic values, results are compared  
511 with expectations from a temporal reshuffling of the data. For each country, we keep the same NPIs  
512 adopted but reshuffle the timestamps of their adoption.

### 513 **Acknowledgements**

514 We thank Alexandra Roux for her contribution to the coding of the interventions recorded in  
515 the data set used in this study. We thank David Garcia, Vito D.P. Servedio, David Hofmann for  
516 their contribution in the early stage of this work. NH would like to thank Luis Haug for helpful  
517 discussions. This work was funded by the Austrian Science Promotion Agency, FFG project  
518 under 857136, the WWTF under COV 20-001, COV 20-017 and MA16-045, the Medizinisch-  
519 Wissenschaftlichen Fonds des Bürgermeisters der Bundeshauptstadt Wien under CoVid004, and the  
520 project VET-Austria, a cooperation between the Austrian Federal Ministry of Social Affairs, Health,  
521 Care and Consumer Protection, the Austrian Agency for Health and Food Safety and the University  
522 of Veterinary Medicine Vienna.

523 **Author contributions**

524 NH, LG, AL, VL, PK conceived and performed the analyses. VL, ST, PK supervised the study.

525 ED contributed additional tools. NH, LG, AL, ADL, BP and PK wrote the first draft of the paper.

526 ADL supervised the data collection on NPIs. All authors (NH, LG, AL, ED, ADL, VL, BP, ST, PK)

527 discussed the results and contributed to the revision of the final manuscript.

528 **Competing interests**

529 The authors declare no competing interests.