Check for updates

Ranking the effectiveness of worldwide COVID-19 government interventions

Nils Haug [1,2,7], Lukas Geyrhofer [2,7], Alessandro Londei [3], Elma Dervic [1,2], Amélie Desvars-Larrive [2,4], Vittorio Loreto [2,3,5], Beate Pinior [2,4], Stefan Thurner [1,2,6] and Peter Klimek [1,2] ✉

Assessing the effectiveness of non-pharmaceutical interventions (NPIs) to mitigate the spread of SARS-CoV-2 is critical to inform future preparedness response plans. Here we quantify the impact of 6,068 hierarchically coded NPIs implemented in 79 territories on the effective reproduction number, R_t, of COVID-19. We propose a modelling approach that combines four computational techniques merging statistical, inference and artificial intelligence tools. We validate our findings with two external datasets recording 42,151 additional NPIs from 226 countries. Our results indicate that a suitable combination of NPIs is necessary to curb the spread of the virus. Less disruptive and costly NPIs can be as effective as more intrusive, drastic, ones (for example, a national lockdown). Using country-specific 'what-if' scenarios, we assess how the effectiveness of NPIs depends on the local context such as timing of their adoption, opening the way for forecasting the effectiveness of future interventions.

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

Confronted with the worldwide COVID-19 epidemic, most governments have implemented bundles of highly restrictive, sometimes intrusive, NPIs. Decisions had to be taken under rapidly changing epidemiological situations, despite (at least at the very beginning of the epidemic) a lack of scientific evidence on the individual and combined effectiveness of these measures[2–4], degree of compliance of the population and societal impact.

Government interventions may cause substantial economic and social costs[5] while affecting individuals' behaviour, mental health and social security[6]. Therefore, knowledge of the most effective NPIs would allow stakeholders to judiciously and timely implement a specific sequence of key interventions to combat a resurgence of COVID-19 or any other future respiratory outbreak. Because many countries rolled out several NPIs simultaneously, the challenge arises of disentangling the impact of each individual intervention.

To date, studies of the country-specific progression of the COVID-19 pandemic[7] have mostly explored the independent effects of a single category of interventions. These categories include travel restrictions[2,8], social distancing[9–12] and personal protective measures[13]. Additionally, modelling studies typically focus on NPIs that directly influence contact probabilities (for example, social distancing measures[18], social distancing behaviours[12], self-isolation, school closures, bans on public events[20] and so on). Some studies focused on a single country or even a town[14–18] while other research combined data from multiple countries but pooled NPIs into rather broad categories[15,19–21], which eventually limits the assessment of specific, potentially critical, NPIs that may be less costly and more effective than others. Despite their widespread use, relative ease of implementation, broad choice of available tools and their importance in developing countries where other measures (for example, increases in healthcare capacity, social distancing or enhanced testing) are difficult to implement[22], little is currently known about the effectiveness of different risk-communication strategies. An accurate assessment of communication activities requires information on the targeted public, means of communication and content of the message.
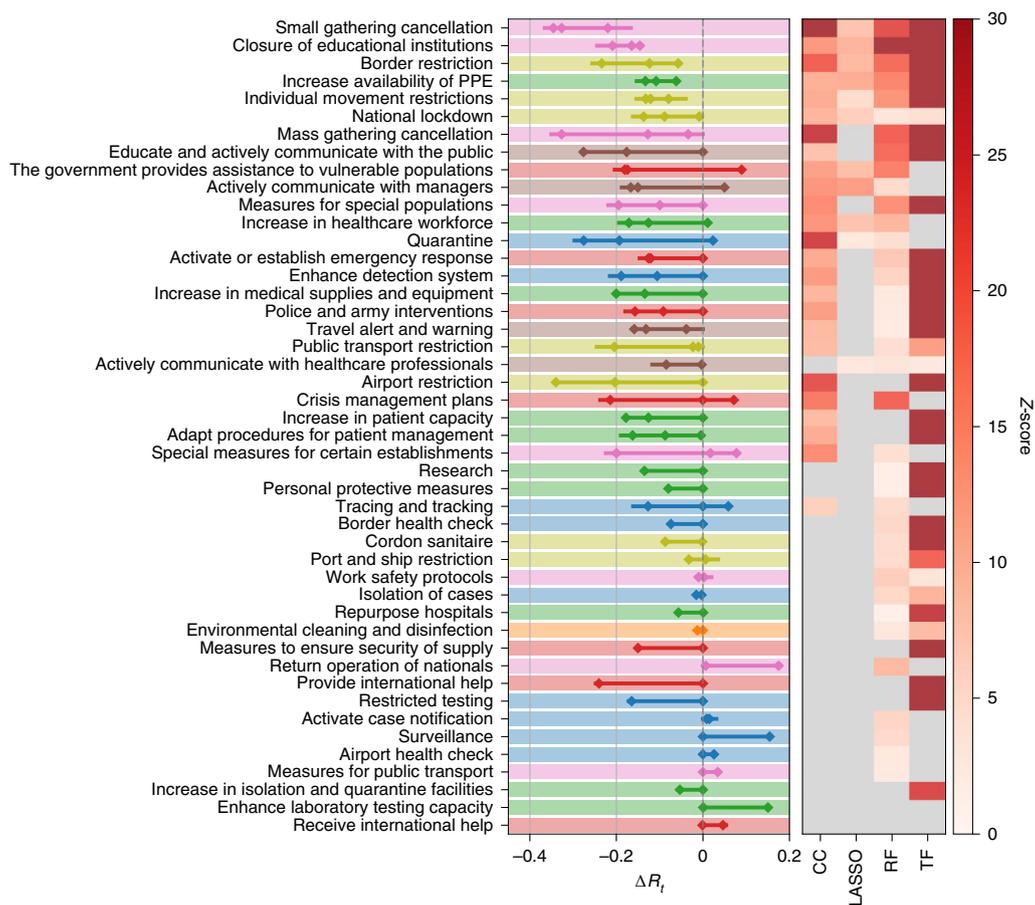
Using a comprehensive, hierarchically coded dataset of 6,068 NPIs implemented in March–April 2020 (when most European countries and US states experienced their first infection waves) in 79 territories[23], here we analyse the impact of government interventions on R_t using harmonized results from a multi-method approach consisting of (1) a case-control analysis (CC), (2) a step function approach to LASSO time-series regression (LASSO), (3) random forests (RF) and (4) transformers (TF). We contend that the combination of four different methods, combining statistical, inference and artificial intelligence classes of tools, also allows assessment of the structural uncertainty of individual methods[24]. We also investigate country-specific control strategies as well as the impact of selected country-specific metrics.

All the above approaches (1–4) yield comparable rankings of the effectiveness of different categories of NPIs across their hierarchical levels. This remarkable agreement allows us to identify a consensus set of NPIs that lead to a significant reduction in R_t. We validate this consensus set using two external datasets covering 42,151 measures in 226 countries. Furthermore, we evaluate the heterogeneity of the effectiveness of individual NPIs in different territories. We find that the time of implementation, previously implemented measures, different governance indicators[25], as well as human and social development affect the effectiveness of NPIs in countries to varying degrees.

Results

Global approach. Our main results are based on the Complexity Science Hub COVID-19 Control Strategies List (CCCSL)[23]. This dataset provides a hierarchical taxonomy of 6,068 NPIs, coded on four levels, including eight broad themes (level 1, L1) divided into

Fig. 1 | Change in R_t (ΔR_t) for 46 NPIs at L2, as quantified by CC analysis, LASSO and TF regression. The left-hand panel shows the combined 95% confidence intervals of ΔR_t for the most effective interventions across all included territories. The heatmap in the right-hand panel shows the corresponding Z-scores of measure effectiveness as determined by the four different methods. Grey indicates no significantly positive effect. 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 Supplementary Fig. 1.

63 categories of individual NPIs (level 2, L2) that include >500 sub-categories (level 3, L3) and >2,000 codes (level 4, L4). We first compare the results for NPI effectiveness rankings for the four methods of our approach (1–4) on L1 (themes) (Supplementary Fig. 1). A clear picture emerges where the themes of social distancing and travel restrictions are top ranked in all methods, whereas environmental measures (for example, cleaning and disinfection of shared surfaces) are ranked least effective.

We next compare results obtained on L2 of the NPI dataset—that is, using the 46 NPI categories implemented more than five times. The methods largely agree on the list of interventions that have a significant effect on R_t (Fig. 1 and Table 1). The individual rankings are highly correlated with each other ($P = 0.0008$; Methods). Six NPI categories show significant impacts on R_t in all four methods. In Supplementary Table 1 we list the subcategories (L3) belonging to these consensus categories.

A normalized score for each NPI category is obtained by rescaling the result within each method to range between zero (least effective) and one (most effective) and then averaging this score. The maximal (minimal) NPI score is therefore 100% (0%), meaning that the measure is the most (least) effective measure in each method. We show the normalized scores for all measures in the CCSL dataset in Extended Data Fig. 1, for the CoronaNet dataset in Extended Data Fig. 2 and for the WHO Global Dataset of Public Health and Social Measures (WHO-PHSM) in Extended

Data Fig. 3. Among the six full-consensus NPI categories in the CCSL, the largest impacts on R_t are shown by small gathering cancellations (83%, ΔR_t between −0.22 and −0.35), the closure of educational institutions (73%, and estimates for ΔR_t ranging from −0.15 to −0.21) and border restrictions (56%, ΔR_t between −0.057 and −0.23). The consensus measures also include NPIs aiming to increase healthcare and public health capacities (increased availability of personal protective equipment (PPE): 51%, ΔR_t −0.062 to −0.13), individual movement restrictions (42%, ΔR_t −0.08 to −0.13) and national lockdown (including stay-at-home order in US states) (25%, ΔR_t −0.008 to −0.14).

We find 14 additional NPI categories consensually in three of our methods. These include mass gathering cancellations (53%, ΔR_t between −0.13 and −0.33), risk-communication activities to inform and educate the public (48%, ΔR_t between −0.18 and −0.28) and government assistance to vulnerable populations (41%, ΔR_t between −0.17 and −0.18).

Among the least effective interventions we find: government actions to provide or receive international help, measures to enhance testing capacity or improve case detection strategy (which can be expected to lead to a short-term rise in cases), tracing and tracking measures as well as land border and airport health checks and environmental cleaning.

In Fig. 2 we show the findings on NPI effectiveness in a co-implementation network. Nodes correspond to categories (L2)

Table 1 | Comparison of effectiveness rankings on L2

| L2 category	Score (%)	Consensus	ΔR_t^{CC}	ΔR_t^{LASSO}	Importance (RF)	ΔR_t^{TF}
Small gathering cancellation	83	4	−0.35 (2)	−0.22 (5)	0.020 (2)	−0.327 (3)
Closure of educational institutions	73	4	−0.16 (2)	−0.21 (4)	0.028 (2)	−0.146 (2)
Border restriction	56	4	−0.23 (2)	−0.12 (2)	0.017 (2)	−0.057 (2)
Increased availability of PPE	51	4	−0.11 (2)	−0.13 (2)	0.012 (1)	−0.062 (2)
Individual movement restrictions	42	4	−0.13 (2)	−0.08 (3)	0.017 (2)	−0.121 (2)
National lockdown	25	4	−0.14 (3)	−0.09 (2)	0.0020 (9)	−0.008 (3)
Mass gathering cancellation	53	3	−0.33 (2)	0	0.012 (1)	−0.127 (2)
Educate and actively communicate with the public	48	3	−0.18 (4)	0	0.018 (2)	−0.276 (2)
The government provides assistance to vulnerable populations	41	3	−0.17 (3)	−0.18 (4)	0.009 (1)	0.090 (3)
Actively communicate with managers	40	3	−0.15 (2)	−0.20 (4)	0.004 (2)	−0.050 (2)
Measures for special populations	37	3	−0.19 (2)	0	0.008 (1)	−0.100 (2)
Increase healthcare workforce	35	3	−0.17 (20)	−0.13 (3)	0.030 (8)	0.011 (2)
Quarantine	30	3	−0.28 (2)	−0.2 (1)	0.0023 (9)	0.023 (2)
Activate or establish emergency response	29	3	−0.13 (2)	0	0.0037 (9)	−0.121 (2)
Enhance detection system	25	3	−0.19 (3)	0	0.0032 (9)	−0.106 (2)
Increase in medical supplies and equipment	25	3	−0.13 (3)	−0.004 (3)	0.003 (2)	−0.200 (3)
Police and army interventions	23	3	−0.16 (2)	0	0.003 (2)	−0.091 (2)
Travel alert and warning	20	3	−0.13 (3)	0.0 (1)	0.002 (1)	−0.159 (3)
Public transport restriction	13	3	00.20 (4)	−0.01 (7)	0.004 (1)	−0.023 (3)
Actively communicate with healthcare professionals	11	3	0	−0.08 (4)	0.003 (1)	−0.003 (2)

Out of the 46 NPI categories, all four methods show significant results for six NPIs (consensus 4) while three methods agree on 14 further NPIs (consensus 3). We report the average normalized score, the observed reduction in R_t for the various methods and NPI importance for RF. Numbers in parentheses denote half of the amount by which the last digit of the corresponding number outside the parentheses fluctuates within the 95% confidence interval.

with size being proportional to their normalized score. Directed links from i to j indicate a tendency that countries implement NPI j after they have implemented i. The network therefore illustrates the typical NPI implementation sequence in the 56 countries and the steps within this sequence that contribute most to a reduction in R_t. For instance, there is a pattern where countries first cancel mass gatherings before moving on to cancellations of specific types of small gatherings, where the latter associates on average with more substantial reductions in R_t. Education and active communication with the public is one of the most effective 'early measures' (implemented around 15 days before 30 cases were reported and well before the majority of other measures comes). Most social distancing (that is, closure of educational institutions), travel restriction measures (that is, individual movement restrictions like curfew and national lockdown) and measures to increase the availability of PPE are typically implemented within the first 2 weeks after reaching 30 cases, with varying impacts on R_t; see also Fig. 1.

Within the CC approach, we can further explore these results on a finer hierarchical level. We show results for 18 NPIs (L3) of the risk-communication theme in Supplementary Information and Supplementary Table 2. The most effective communication strategies include warnings against travel to, and return from, high-risk areas ($\Delta R_t^{CC} = -0.14$ (1); the number in parenthesis denotes the standard error) and several measures to actively communicate with the public. These include to encourage, for example, staying at home ($\Delta R_t^{CC} = -0.14$ (1)), social distancing ($\Delta R_t^{CC} = -0.20$ (1)), workplace safety measures ($\Delta R_t^{CC} = -0.18$ (2)), self-initiated isolation of people with mild respiratory symptoms ($\Delta R_t^{CC} = -0.19$ (2)) and information campaigns ($\Delta R_t^{CC} = -0.13$ (1)) (through various channels including the press, flyers, social media or phone messages).

Validation with external datasets. We validate our findings with results from two external datasets (Methods). In the WHO-PHSM dataset[26] we find seven full-consensus measures (agreement on significance by all methods) and 17 further measures with three agreements (Extended Data Fig. 4). These consensus measures show a large overlap with those (three or four matches in our methods) identified using the CCCSL, and include top-ranked NPI measures aiming at strengthening the healthcare system and testing capacity (labelled as 'scaling up')—for example, increasing the healthcare workforce, purchase of medical equipment, testing, masks, financial support to hospitals, increasing patient capacity, increasing domestic production of PPE. Other consensus measures consist of social distancing measures ('cancelling, restricting or adapting private gatherings outside the home', adapting or closing 'offices, businesses, institutions and operations', 'cancelling, restricting or adapting mass gatherings'), measures for special populations ('protecting population in closed settings', encompassing long-term care facilities and prisons), school closures, travel restrictions (restricting entry and exit, travel advice and warning, 'closing international land borders', 'entry screening and isolation or quarantine') and individual movement restriction ('stay-at-home order', which is equivalent to confinement in the WHO-PHSM coding). 'Wearing a mask' exhibits a significant impact on R_t in three methods (ΔR_t between −0.018 and −0.12). The consensus measures also include financial packages and general public awareness campaigns (as part of 'communications and engagement' actions). The least effective measures include active case detection, contact tracing and environmental cleaning and disinfection.

The CCCSL results are also compatible with findings from the CoronaNet dataset[27] (Extended Data Figs. 5 and 6). Analyses show four full-consensus measures and 13 further NPIs with an
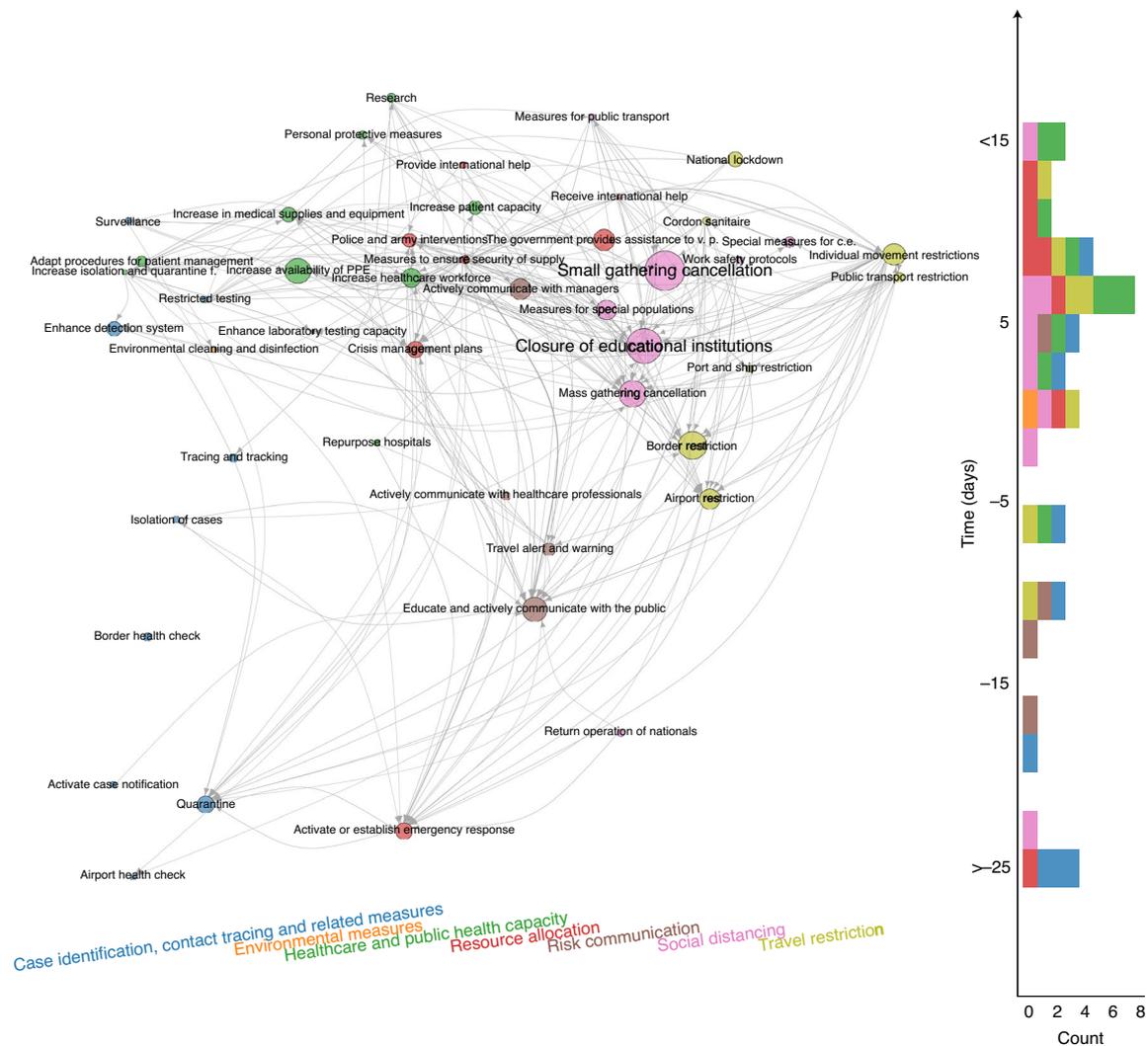
Fig. 2 | Time-ordered NPI co-implementation network across countries. Nodes are categories (L2), with colours indicating the theme (L1) and size being proportional to the average effectiveness of the intervention. Arrows from nodes *i* to *j* denote that those countries which have already implemented 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 that country reached 30 confirmed cases), and horizontally according to their L1 theme. The stacked histogram on the right shows the number of implemented NPIs per time period (epidemic age) and theme (colour). v.p., vulnerable populations; c.e., certain establishments; quarantine f., quarantine facilities.

agreement of three methods. These consensus measures include heterogeneous social distancing measures (for example, restriction and regulation of non-essential businesses, restrictions of mass gatherings), closure and regulation of schools, travel restrictions (for example, internal and external border restrictions), individual movement restriction (curfew), measures aiming to increase the healthcare workforce (for example, 'nurses', 'unspecified health staff') and medical equipment (for example, PPE, 'ventilators', 'unspecified health materials'), quarantine (that is, voluntary or mandatory self-quarantine and quarantine at a government hotel or facility) and measures to increase public awareness ('disseminating information related to COVID-19 to the public that is reliable and factually accurate').

Twenty-three NPIs in the CoronaNet dataset do not show statistical significance in any method, including several restrictions and regulations of government services (for example, for tourist sites, parks, public museums, telecommunications), hygiene measures for public areas and other measures that target very specific populations (for example, certain age groups, visa extensions).

Country-level approach. A sensitivity check of our results with respect to the removal of individual continents from the analysis also indicates substantial variations between world geographical regions in terms of NPI effectiveness (Supplementary Information). To further quantify how much the effectiveness of an NPI depends on the particular territory (country or US state) where it has been introduced, we measure the heterogeneity of NPI rankings in different territories through an entropic approach in the TF method (Methods). Figure 3 shows the normalized entropy of each NPI category versus its rank. A value of entropy close to zero implies that the corresponding NPI has a similar rank relative to all other NPIs in all territories: in other words, the effectiveness of the NPI does not depend on the specific country or state. On the other hand, a high value of the normalized entropy signals that the performance of each NPI depends largely on the geographical region.

The values of the normalized entropies for many NPIs are far from one, and are also below the corresponding values obtained through temporal reshuffling of NPIs in each country. The effectiveness of many NPIs therefore is, first, significant and, second,
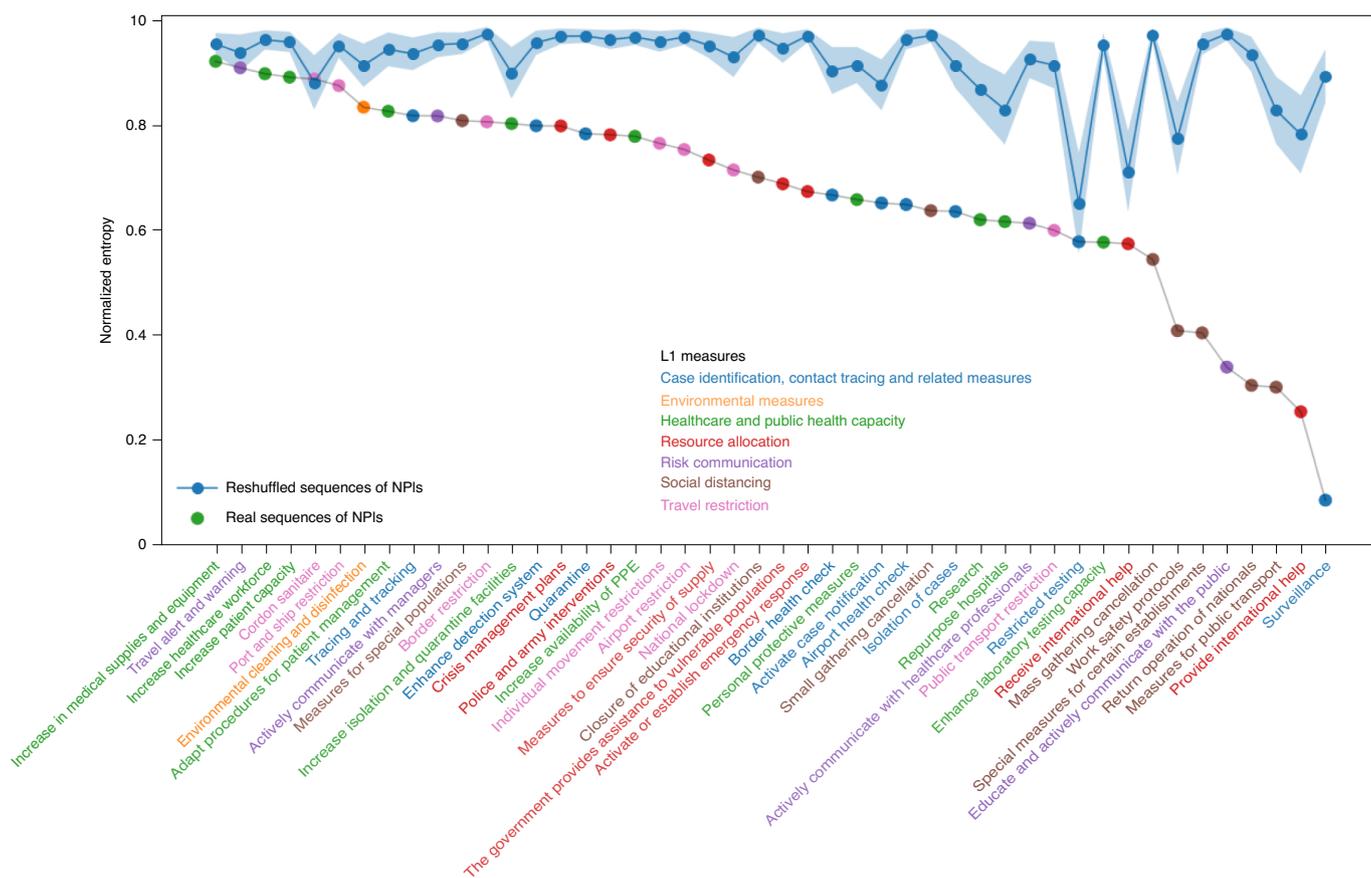
Fig. 3 | Normalized entropies versus rank for all NPIs at level L2. Each NPI is colour coded according to its theme of belonging (L1), as indicated in the legend. The blue curve represents the same information obtained from a reshuffled dataset of NPIs.

depends on the local context (combination of socio-economic features and NPIs already adopted) to varying degrees. In general, social distancing measures and travel restrictions show a high entropy (effectiveness varies considerably across countries) whereas case identification, contact tracing and healthcare measures show substantially less country dependence.

We further explore this interplay of NPIs with socio-economic factors by analysing the effects of demographic and socio-economic covariates, as well as indicators for governance and human and economic development in the CC method (Supplementary Information). While the effects of most indicators vary across different NPIs at rather moderate levels, we find a robust tendency that NPI effectiveness correlates negatively with indicator values for governance-related accountability and political stability (as quantified by World Governance Indicators provided by the World Bank).

Because the heterogeneity of the effectiveness of individual NPIs across countries points to a non-independence among different NPIs, the impact of a specific NPI cannot be evaluated in isolation. Since it is not possible in the real world to change the sequence of NPIs adopted, we resort to 'what-if' experiments to identify the most likely outcome of an artificial sequence of NPIs in each country. Within the TF approach, we selectively delete one NPI at a time from all sequences of interventions in all countries and compute the ensuing evolution of R_t compared to the actual case.

To quantify whether the effectiveness of a specific NPI depends on its epidemic age of implementation, we study artificial sequences of NPIs constructed by shifting the selected NPI to other days, keeping the other NPIs fixed. In this way, for each country and

each NPI, we obtain a curve of the most likely change in R_t versus the adoption time of the specific NPI.

Figure 4 shows an example of the results for a selection of NPIs (see Supplementary Information for a more extensive report on other NPIs). Each curve shows the average change in R_t versus the adoption time of the NPI, averaged over the countries where that NPI has been adopted. Figure 4a refers to the national lockdown (including stay-at-home order implemented in US states). Our results show a moderate effect of this NPI (low change in R_t) as compared to other, less drastic, measures. Figure 4b shows NPIs with the pattern 'the earlier, the better'. For those measures ('closure of educational institutions', 'small gatherings cancellation', 'airport restrictions' and many more shown in Supplementary Information), early adoption is always more beneficial. In Fig. 4c, 'enhancing testing capacity' and 'surveillance' exhibit a negative impact (that is, an increase) on R_t, presumably related to the fact that more testing allows for more cases to be identified. Finally, Fig. 4d, showing 'tracing and tracking' and 'activate case notification', demonstrates an initially negative effect that turns positive (that is, toward a reduction in R_t). Refer to Supplementary Information for a more comprehensive analysis of all NPIs.

Discussion

Our study dissects the entangled packages of NPIs[23] and quantifies their effectiveness. We validate our findings using three different datasets and four independent methods. Our findings suggest that no NPI acts as a silver bullet on the spread of COVID-19. Instead, we identify several decisive interventions that significantly
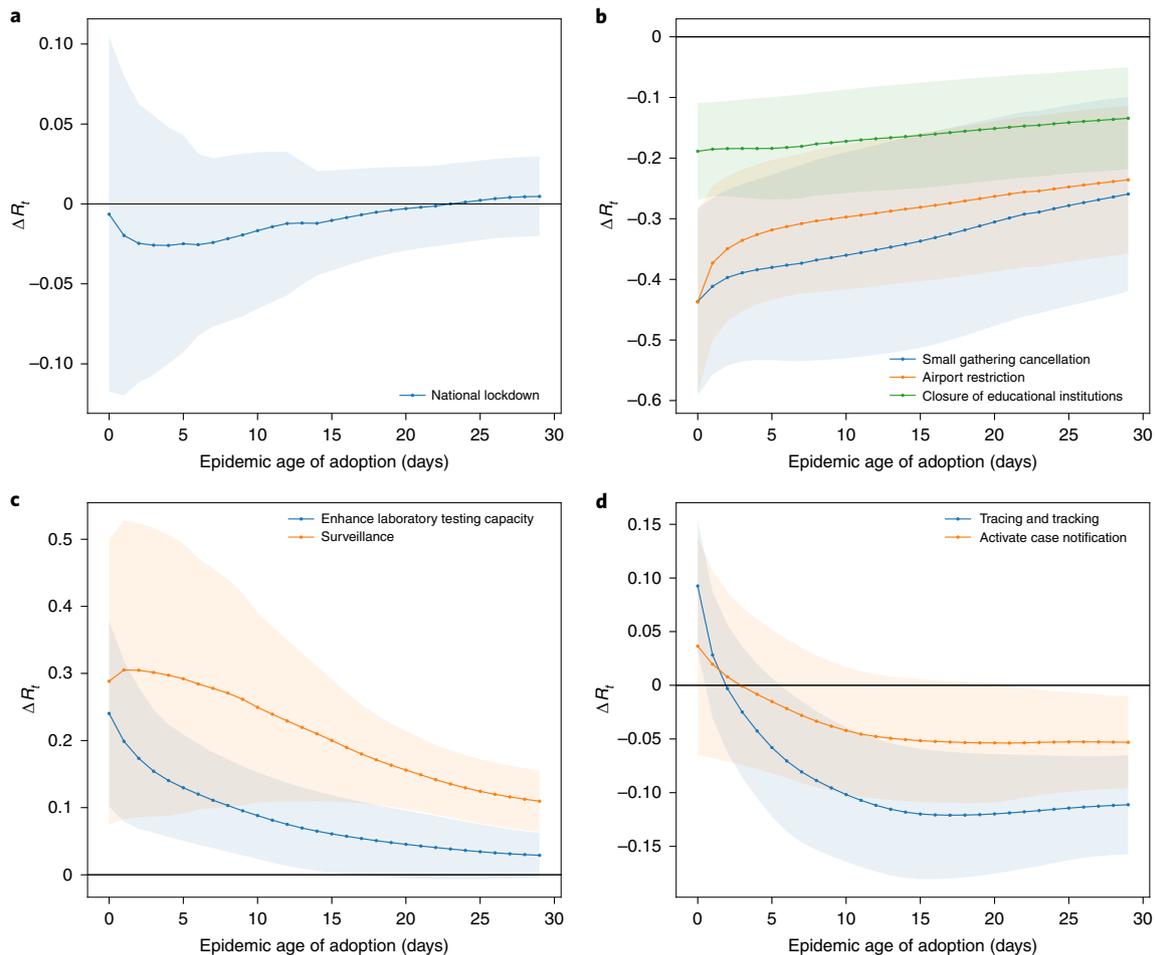
Fig. 4 | Change in R_t as a function of the adoption time of selected NPIs, averaged over countries where those NPIs had been adopted. a, National lockdown (including stay-at-home order in US states). **b,** A selection of three NPIs displaying 'the earlier the better' behaviour—that is, their impact is enhanced if implemented at earlier epidemic ages. **c,** Enhance laboratory testing capacity and Surveillance. **d,** Tracing and tracking and Activate case notification. Negative (positive) values indicate that the adoption of the NPI has reduced (increased) the value of R_t. Shaded areas denote s.d.

contribute to reducing R_t below one and that should therefore be considered as efficiently flattening the curve facing a potential second COVID-19 wave, or any similar future viral respiratory epidemics.

The most effective NPIs include curfews, lockdowns and closing and restricting places where people gather in smaller or large numbers for an extended period of time. This includes small gathering cancellations (closures of shops, restaurants, gatherings of 50 persons or fewer, mandatory home working and so on) and closure of educational institutions. While in previous studies, based on smaller numbers of countries, school closures had been attributed as having little effect on the spread of COVID-19 (refs. [19,20]), more recent evidence has been in favour of the importance of this NPI[28,29]; school closures in the United States have been found to reduce COVID-19 incidence and mortality by about 60% (ref. [28]). This result is also in line with a contact-tracing study from South Korea, which identified adolescents aged 10–19 years as more likely to spread the virus than adults and children in household settings[30]. Individual movement restrictions (including curfew, the prohibition of gatherings and movements for non-essential activities or measures segmenting the population) were also amongst the top-ranked measures.

However, such radical measures have adverse consequences. School closure interrupts learning and can lead to poor nutrition, stress and social isolation in children[31–33]. Home confinement has

strongly increased the rate of domestic violence in many countries, with a huge impact on women and children[34,35], while it has also limited the access to long-term care such as chemotherapy, with substantial impacts on patients' health and survival chance[36,37]. Governments may have to look towards less stringent measures, encompassing maximum effective prevention but enabling an acceptable balance between benefits and drawbacks[38].

Previous statistical studies on the effectiveness of lockdowns came to mixed conclusions. Whereas a relative reduction in R_t of 5% was estimated using a Bayesian hierarchical model[19], a Bayesian mechanistic model estimated a reduction of 80% (ref. [20]), although some questions have been raised regarding the latter work because of biases that overemphasize the importance of the most recent measure that had been implemented[24]. The susceptibility of other modelling approaches to biases resulting from the temporal sequence of NPI implementations remains to be explored. Our work tries to avoid such biases by combining multiple modelling approaches and points to a mild impact of lockdowns due to an overlap with effects of other measures adopted earlier and included in what is referred to as 'national (or full) lockdown'. Indeed, the national lockdown encompasses multiple NPIs (for example, closure of land, sea and air borders, closure of schools, non-essential shops and prohibition of gatherings and visiting nursing homes) that countries may have already adopted in parts. From this perspective, the relatively

attenuated impact of the national lockdown is explained as the little delta after other concurrent NPIs have been adopted. This conclusion does not rule out the effectiveness of an early national lockdown, but suggests that a suitable combination (sequence and time of implementation) of a smaller package of such measures can substitute for a full lockdown in terms of effectiveness, while reducing adverse impacts on society, the economy, the humanitarian response system and the environment[6,39–41].

Taken together, the social distancing and movement-restriction measures discussed above can therefore be seen as the 'nuclear option' of NPIs: highly effective but causing substantial collateral damages to society, the economy, trade and human rights[4,39].

We find strong support for the effectiveness of border restrictions. The role of travelling in the global spread of respiratory diseases proved central during the first SARS epidemic (2002–2003)[42], but travelling restrictions show a large impact on trade, economy and the humanitarian response system globally[41,43]. The effectiveness of social distancing and travel restrictions is also in line with results from other studies that used different statistical approaches, epidemiological metrics, geographic coverage and NPI classification[2,8–11,13,19,20].

We also find a number of highly effective NPIs that can be considered less costly. For instance, we find that risk-communication strategies feature prominently amongst consensus NPIs. This includes government actions intended to educate and actively communicate with the public. The effective messages include encouraging people to stay at home, promoting social distancing and workplace safety measures, encouraging the self-initiated isolation of people with symptoms, travel warnings and information campaigns (mostly via social media). All these measures are non-binding government advice, contrasting with the mandatory border restriction and social distancing measures that are often enforced by police or army interventions and sanctions. Surprisingly, communicating on the importance of social distancing has been only marginally less effective than imposing distancing measures by law. The publication of guidelines and work safety protocols to managers and healthcare professionals was also associated with a reduction in R_t, suggesting that communication efforts also need to be tailored toward key stakeholders. Communication strategies aim at empowering communities with correct information about COVID-19. Such measures can be of crucial importance in targeting specific demographic strata found to play a dominant role in driving the spread of COVID-19 (for example, communication strategies to target individuals aged <40 years[44]).

Government food assistance programmes and other financial supports for vulnerable populations have also turned out to be highly effective. Such measures are, therefore, not only impacting the socio-economic sphere[45] but also have a positive effect on public health. For instance, facilitating people's access to tests or allowing them to self-isolate without fear of losing their job or part of their salary may help in reducing R_t.

Some measures are ineffective in (almost) all methods and datasets—for example, environmental measures to disinfect and clean surfaces and objects in public and semi-public places. This finding is at odds with current recommendations of the WHO (World Health Organization) for environmental cleaning in non-healthcare settings[46], and calls for a closer examination of the effectiveness of such measures. However, environmental measures (for example, cleaning of shared surfaces, waste management, approval of a new disinfectant, increased ventilation) are seldom reported by governments or the media and are therefore not collected by NPI trackers, which could lead to an underestimation of their impact. These results call for a closer examination of the effectiveness of such measures. We also find no evidence for the effectiveness of social distancing measures in regard to public transport. While infections on buses and trains have been reported[47], our results

may suggest a limited contribution of such cases to the overall virus spread, as previously reported[48]. A heightened public risk awareness associated with commuting (for example, people being more likely to wear face masks) might contribute to this finding[49]. However, we should note that measures aiming at limiting engorgement or increasing distancing on public transport have been highly diverse (from complete cancellation of all public transport to increase in the frequency of traffic to reduce traveller density) and could therefore lead to widely varying effectiveness, also depending on the local context.

The effectiveness of individual NPIs is heavily influenced by governance (Supplementary Information) and local context, as evidenced by the results of the entropic approach. This local context includes the stage of the epidemic, socio-economic, cultural and political characteristics and other NPIs previously implemented. The fact that gross domestic product is overall positively correlated with NPI effectiveness whereas the governance indicator 'voice and accountability' is negatively correlated might be related to the successful mitigation of the initial phase of the epidemic of certain south-east Asian and Middle East countries showing authoritarian tendencies. Indeed, some south-east Asian government strategies heavily relied on the use of personal data and police sanctions whereas the Middle East countries included in our analysis reported low numbers of cases in March–April 2020.

By focusing on individual countries, the what-if experiments using artificial country-specific sequences of NPIs offer a way to quantify the importance of this local context with respect to measurement of effectiveness. Our main takeaway here is that the same NPI can have a drastically different impact if taken early or later, or in a different country.

It is interesting to comment on the impact that 'enhancing testing capacity' and 'tracing and tracking' would have had if adopted at different points in time. Enhancing testing capacity should display a short-term increase in R_t. Counter-intuitively, in countries testing close contacts, tracing and tracking, if they are effective, would have a similar effect on R_t because more cases will be found (although tracing and tracking would reduce R_t in countries that do not test contacts but rely on quarantine measures). For countries implementing these measures early, indeed, we find a short-term increase in R_t (when the number of cases was sufficiently small to enable tracing and testing of all contacts). However, countries implementing these NPIs later did not necessarily find more cases, as shown by the corresponding decrease in R_t. We focus on March and April 2020, a period in which many countries had a sudden surge in cases that overwhelmed their tracing and testing capacities, which rendered the corresponding NPIs ineffective.

Assessment of the effectiveness of NPIs is statistically challenging, because measures were typically implemented simultaneously and their impact might well depend on the particular implementation sequence. Some NPIs appear in almost all countries whereas in others only a few, meaning that we could miss some rare but effective measures due to a lack of statistical power. While some methods might be prone to overestimation of the effects from an NPI due to insufficient adjustments for confounding effects from other measures, other methods might underestimate the contribution of an NPI by assigning its impact to a highly correlated NPI. As a consequence, estimates of ΔR_t might vary substantially across different methods whereas agreement on the significance of individual NPIs is much more pronounced. The strength of our study, therefore, lies in the harmonization of these four independent methodological approaches combined with the usage of an extensive dataset on NPIs. This allows us to estimate the structural uncertainty of NPI effectiveness—that is, the uncertainty introduced by choosing a certain model structure likely to affect other modelling works that rely on a single method only. Moreover, whereas previous studies often subsumed a wide range of social distancing and travel restriction

measures under a single entity, our analysis contributes to a more fine-grained understanding of each NPI.

The CCCSL dataset features non-homogeneous data completeness across the different territories, and data collection could be biased by the data collector (native versus non-native) as well as by the information communicated by governments (see also ref. [23]). The WHO-PHSM and CoronaNet databases contain a broad geographic coverage whereas CCCSL focuses mostly on developed countries. Moreover, the coding system presents certain drawbacks, notably because some interventions could belong to more than one category but are recorded only once. Compliance with NPIs is crucial for their effectiveness, yet we assumed a comparable degree of compliance by each population. We tried to mitigate this issue by validating our findings on two external databases, even if these are subject to similar limitations. We did not perform a formal harmonization of all categories in the three NPI trackers, which limits our ability to perform full comparisons among the three datasets. Additionally, we neither took into account the stringency of NPI implementation nor the fact that not all methods were able to describe potential variations in NPI effectiveness over time, besides the dependency on the epidemic age of its adoption. The time window is limited to March–April 2020, where the structure of NPIs is highly correlated due to simultaneous implementation. Future research should consider expanding this window to include the period when many countries were easing policies, or maybe even strenghening them again after easing, as this would allow clearer differentiation of the correlated structure of NPIs because they tended to be released, and implemented again, one (or a few) at a time.

To compute R_t, we used time series of the number of confirmed COVID-19 cases[50]. This approach is likely to over-represent patients with severe symptoms and may be biased by variations in testing and reporting policies among countries. Although we assume a constant serial interval (average timespan between primary and secondary infection), this number shows considerable variation in the literature[51] and depends on measures such as social distancing and self-isolation.

In conclusion, here we present the outcome of an extensive analysis on the impact of 6,068 individual NPIs on the R_t of COVID-19 in 79 territories worldwide. Our analysis relies on the combination of three large and fine-grained datasets on NPIs and the use of four independent statistical modelling approaches.

The emerging picture reveals that no one-size-fits-all solution exists, and no single NPI can decrease R_t below one. Instead, in the absence of a vaccine or efficient antiviral medication, a resurgence of COVID-19 cases can be stopped only by a suitable combination of NPIs, each tailored to the specific country and its epidemic age. These measures must be enacted in the optimal combination and sequence to be maximally effective against the spread of SARS-CoV-2 and thereby enable more rapid reopening.

We showed that the most effective measures include closing and restricting most places where people gather in smaller or larger numbers for extended periods of time (businesses, bars, schools and so on). However, we also find several highly effective measures that are less intrusive. These include land border restrictions, governmental support to vulnerable populations and risk-communication strategies. We strongly recommend that governments and other stakeholders first consider the adoption of such NPIs, tailored to the local context, should infection numbers surge (or surge a second time), before choosing the most intrusive options. Less drastic measures may also foster better compliance from the population.

Notably, the simultaneous consideration of many distinct NPI categories allows us to move beyond the simple evaluation of individual classes of NPIs to assess, instead, the collective impact of specific sequences of interventions. The ensemble of these results calls for a strong effort to simulate what-if scenarios at the country level for planning the most probable effectiveness of future NPIs, and,

thanks to the possibility of going down to the level of individual countries and country-specific circumstances, our approach is the first contribution toward this end.

Methods

Data. *NPI data.* We use the publicly available CCCSL dataset on NPIs[23], in which NPIs are categorized using a four-level hierarchical coding scheme. L1 defines the theme of the NPI: 'case identification, contact tracing and related measures', 'environmental measures', 'healthcare and public health capacity', 'resource allocation', 'returning to normal life', 'risk communication', 'social distancing' and 'travel restriction'. Each L1 (theme) is composed of several categories (L2 of the coding scheme) that contain subcategories (L3), which are further subdivided into group codes (L4). The dataset covers 56 countries; data for the United States are available at the state level (24 states), making a total of 79 territories. In this analysis, we use a static version of the CCCSL, retrieved on 17 August 2020, presenting 6,068 NPIs. A glossary of the codes, with a detailed description of each category and its subcategories, is provided on GitHub. For each country, we use the data until the day for which the measures have been reliably updated. NPIs that have been implemented in fewer than five territories are not considered, leading to a final total of 4,780 NPIs of 46 different L2 categories for use in the analyses.

Second, we use the CoronaNet COVID-19 Government Response Event Dataset (v.1.0)[27] that contains 31,532 interventions and covers 247 territories (countries and US states) (data extracted on 17 August 2020). For our analysis, we map their columns 'type' and 'type_sub_cat' onto L1 and L2, respectively. Definitions for the entire 116 L2 categories can be found on the GitHub page of the project.

Using the same criterion as for the CCCSL, we obtain a final total of 18,919 NPIs of 107 different categories.

Third, we use the WHO-PHSM dataset[26], which merges and harmonizes the following datasets: ACAPS[41], Oxford COVID-19 Government Response Tracker[52], the Global Public Health Intelligence Network (GPHIN) of Public Health Agency of Canada (Ottawa, Canada), the CCCSL[23], the United States Centers for Disease Control and Prevention and HIT-COVID[53]. The WHO-PHSM dataset contains 24,077 interventions and covers 264 territories (countries and US states; data extracted on 17 August 2020). Their encoding scheme has a heterogeneous coding depth and, for our analysis, we map 'who_category' onto L1 and either take 'who_subcategory' or a combination of 'who_subcategory' and 'who_measure' as L2. This results in 40 measure categories. A glossary is available at: https://www.who.int/emergencies/diseases/novel-coronavirus-2019/phsm.

The CoronaNet and WHO-PHSM datasets also provide information on the stringency of the implementation of a given NPI, which we did not use in the current study.

COVID-19 case data. To estimate R_t and growth rates of the number of COVID-19 cases, we use time series of the number of confirmed COVID-19 cases in the 79 territories considered[50]. To control for weekly fluctuations, we smooth the time series by computing the rolling average using a Gaussian window with a standard deviation of 2 days, truncated at a maximum window size of 15 days.

Regression techniques. We apply four different statistical approaches to quantify the impact of a NPI, M, on the reduction in R_t (Supplementary Information).

CC. Case-control analysis considers each single category (L2) or subcategory (L3) M separately and evaluates in a matched comparison the difference, ΔR_t, in R_t between all countries that implemented M (cases) and those that did not (controls) during the observation window. The matching is done on epidemic age and the time of implementation of any response. The comparison is made via a linear regression model adjusting for (1) epidemic age (days after the country has reached 30 confirmed cases), (2) the value of R_t before M takes effect, (3) total population, (4) population density, (5) the total number of NPIs implemented and (6) the number of NPIs implemented in the same category as M. With this design, we investigate the time delay of τ days between implementation of M and observation of ΔR_t, as well as additional country-based covariates that quantify other dimensions of governance and human and economic development. Estimates for R_t are averaged over delays between 1 and 28 days.

Step function Lasso regression. In this approach we assume that, without any intervention, the reproduction factor is constant and deviations from this constant result from a delayed onset by τ days of each NPI on L2 (categories) of the hierarchical dataset. We use a Lasso regularization approach combined with a meta parameter search to select a reduced set of NPIs that best describe the observed ΔR_t. Estimates for the changes in ΔR_t attributable to NPI M are obtained from country-wise cross-validation.

RF regression. We perform a RF regression, where the NPIs implemented in a country are used as predictors for R_t, time-shifted τ days into the future. Here, τ accounts for the time delay between implementation and onset of the effect of a given NPI. Similar to the Lasso regression, the assumption underlying the RF approach is that, without changes in interventions, the value of R_t in a territory

remains constant. However, contrary to the two methods described above, RF represents a nonlinear model, meaning that the effects of individual NPIs on R_t do not need to add up linearly. The importance of a NPI is defined as the decline in predictive performance of the RF on unseen data if the data concerning that NPI are replaced by noise, also called permutation importance.

Transformer modelling. Transformers[54] have been demonstrated as models suitable for dynamic discrete element processes such as textual sequences, due to their ability to recall past events. Here we extended the transformer architecture to approach the continuous case of epidemic data by removing the probabilistic output layer with a linear combination of transformer output, whose input is identical to that for RF regression, along with the values of R_t. The best-performing network (least mean-squared error in country-wise cross-validation) is identified as a transformer encoder with four hidden layers of 128 neurons, an embedding size of 128, eight heads, one output described by a linear output layer and 47 inputs (corresponding to each category and R_t). To quantify the impact of measure M on R_t, we use the trained transformer as a predictive model and compare simulations without any measure (reference) to those where one measure is presented at a time to assess ΔR_t. To reduce the effects of overfitting and multiplicity of local minima, we report results from an ensemble of transformers trained to similar precision levels.

Estimation of R_t. We use the R package EpiEstim[55] with a sliding time window of 7 days to estimate the time series of R_t for every country. We choose an uncertain serial interval following a probability distribution with a mean of 4.46 days and a standard deviation of 2.63 days[56].

Ranking of NPIs. For each of the methods (CC, Lasso regression and TF), we rank the NPI categories in descending order according to their impact—that is, the estimated degree to which they lower R_t or their feature importance (RF). To compare rankings, we count how many of the 46 NPIs considered are classified as belonging to the top x ranked measures in all methods, and test the null hypothesis that this overlap has been obtained from completely independent rankings. The P value is then given by the complementary cumulative distribution function for a binomial experiment with 46 trials and success probability $(x/46)$[4]. We report the median P value obtained over all $x \leq 10$ to ensure that the results are not dependent on where we impose the cut-off for the classes.

Co-implementation network. If there is a statistical tendency that a country implementing NPI i also implements NPI j later in time, we draw a direct link from i to j. Nodes are placed on the y axis according to the average epidemic age at which the corresponding NPI is implemented; they are grouped on the x axis by their L1 theme. Node colours correspond to themes. The effectiveness scores for all NPIs are re-scaled between zero and one for each method; node size is proportional to the re-scaled scores, averaged over all methods.

Entropic country-level approach. Each territory can be characterized by its socio-economic conditions and the unique temporal sequence of NPIs adopted. To quantify the NPI effect, we measure the heterogeneity of the overall rank of a NPI amongst the countries that have taken that NPI. To compare countries that have implemented different numbers of NPIs, we consider the normalized rankings where the ranking position is divided by the number of elements in the ranking list (that is, the number of NPIs taken in a specific country). We then bin the interval [0, 1] of the normalized rankings into ten sub-intervals and compute for each NPI the entropy of the distribution of occurrences of that NPI in the different normalized rankings per country:

$$S(\text{NPI}) = -\frac{1}{\log(10)} \sum_i P_i \log(P_i), \tag{1}$$

where P_i is the probability that the NPI considered appeared in the ith bin in the normalized rankings of all countries. To assess the confidence of these entropic values, results are compared with expectations from a temporal reshuffling of the data. For each country, we keep the same NPIs adopted but reshuffle the time stamps of their adoption.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability
The CCCSL dataset can be downloaded from http://covid19-interventions. com/. The CoronaNet data can be found at https://www.coronanet-project. org/. The WHO-PHSM dataset is available at https://www.who.int/emergencies/ diseases/novel-coronavirus-2019/phsm. Snapshots of the datasets used in our study are available in the following github repository: https://github.com/ complexity-science-hub/ranking_npis.

Code availability
Custom code for the analysis is available in the following github repository: https:// github.com/complexity-science-hub/ranking_npis.

References
1. Qualls, N. L. et al. Community mitigation guidelines to prevent pandemic influenza – United States, 2017. *MMWR Recomm. Rep.* **66**, 1–34 (2017).
2. Tian, H. et al. An investigation of transmission control measures during the first 50 days of the COVID-19 epidemic in China. *Science* **368**, 638–642 (2020).
3. Chen, S. et al. COVID-19 control in China during mass population movements at New Year. *Lancet* **395**, 764–766 (2020).
4. Lee, K., Worsnop, C. Z., Grépin, K. A. & Kamradt-Scott, A. Global coordination on cross-border travel and trade measures crucial to COVID-19 response. *Lancet* **395**, 1593–1595 (2020).
5. Chakraborty, I. & Maity, P. Covid-19 outbreak: migration, effects on society, global environment and prevention. *Sci. Total Environ.* **728**, 138882 (2020).
6. Pfefferbaum, B. & North, C. S. Mental health and the COVID-19 pandemic. *N. Eng. J. Med.* **383**, 510–512.
7. *COVID-19 dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University of Medicine* (Johns Hopkins University of Medicine, accessed 4 June 2020); https://coronavirus.jhu.edu/map.html.
8. Chinazzi, M. et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* **368**, 395–400 (2020).
9. Arenas, A., Cota, W., Granell, C. & Steinegger, B. Derivation of the effective reproduction number R for COVID-19 in relation to mobility restrictions and confinement. Preprint at *medRxiv* https://doi.org/10.1101/2020.04. 06.20054320 (2020).
10. Wang, J., Tang, K., Feng, K. & Lv, W. When is the COVID-19 pandemic over? Evidence from the stay-at-home policy execution in 106 Chinese cities. Preprint at *SSRN* https://doi.org/10.2139/ssrn.3561491 (2020).
11. Soucy, J.-P. R. et al. Estimating effects of physical distancing on the COVID-19 pandemic using an urban mobility index. Preprint at *medRxiv* https://doi.org/10.1101/2020.04.05.20054288 (2020).
12. Anderson, S. C. et al. Estimating the impact of Covid-19 control measures using a Bayesian model of physical distancing. Preprint at *medRxiv* https://doi.org/10.1101/2020.04.17.20070086 (2020).
13. Teslya, A. et al. Impact of self-imposed prevention measures and short-term government intervention on mitigating and delaying a COVID-19 epidemic. *PLoS Med.* https://doi.org/10.1371/journal.pmed.1003166 (2020).
14. Kraemer, M. U. et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **497**, 493–497 (2020).
15. Prem, K. & Liu, Y. et al. The effect of control strategies to reduce social mixing on outcomes of the COVID-19 epidemic in Wuhan, China: a modelling study. *Lancet Public Health* **5**, e261–e270 (2020).
16. Gatto, M. et al. Spread and dynamics of the COVID-19 epidemic in Italy: effects of emergency containment measures. *Proc. Natl Acad. Sci. USA* **117**, 10484–10491 (2020).
17. Lorch, L. et al. A spatiotemporal epidemic model to quantify the effects of contact tracing, testing, and containment. Preprint at *arXiv* https://arxiv.org/ abs/2004.07641 (2020).
18. Dehning, J. & Zierenberg, J. et al. Inferring change points in the spread of COVID-19 reveals the effectiveness of interventions. *Science* **369**, eabb9789 (2020).
19. Banholzer, N. et al. Impact of non-pharmaceutical interventions on documented cases of COVID-19. Preprint at *medRxiv* https://doi.org/10.1101/2020.04. 16.20062141 (2020).
20. Flaxman, S. et al. Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* **584**, 257–261 (2020).
21. Hsiang, S. et al. The effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Nature* **584**, 262–267 (2020).
22. Nachega, J., Seydi, M. & Zumla, A. The late arrival of coronavirus disease 2019 (Covid-19) in Africa: mitigating pan-continental spread. *Clin. Infect. Dis.* **71**, 875–878 (2020).
23. Desvars-Larrive, A. et al. A structured open dataset of government interventions in response to COVID-19. *Sci. Data* **7**, 285 (2020).
24. Bryant, P. & Elofsson, A. The limits of estimating COVID-19 intervention effects using Bayesian models. Preprint at *medRxiv* https://doi.org/10.1101/ 2020.08.14.20175240 (2020).
25. *Protecting People and Economies: Integrated Policy Responses to COVID-19* (World Bank, 2020); https://openknowledge.worldbank.org/handle/ 10986/33770
26. Tracking Public Health and Social Measures: A Global Dataset (World Health Organization, 2020); https://www.who.int/emergencies/diseases/novel- coronavirus-2019/phsm
27. Cheng, C., Barceló, J., Hartnett, A. S., Kubinec, R. & Messerschmidt, L. COVID-19 government response event dataset (CoronaNet v.1.0). *Nat. Hum. Behav.* **4**, 756–768 (2020).
28. Auger, K. A. et al. Association between statewide school closure and COVID-19 incidence and mortality in the US. *JAMA* **324**, 859–870 (2020).

29. Liu, Y. et al. The impact of non-pharmaceutical interventions on SARS-CoV-2 transmission across 130 countries and territories. Preprint at *medRxiv* https://doi.org/10.1101/2020.08.11.20172643 (2020).

30. Park, Y., Choe, Y. et al. Contact tracing during coronavirus disease outbreak. *Emerg. Infect. Dis.* **26**, 2465–2468(2020).

31. Adverse Consequences of School Closures (UNESCO, 2020); https://en.unesco.org/covid19/educationresponse/consequences

32. Education and COVID-19: Focusing on the Long-term Impact of School Closures (OECD, 2020); https://www.oecd.org/coronavirus/policy-responses/education-and-covid-19-focusing-on-the-long-term-impact-of-school-closures-2cea926e/

33. Orben, A., Tomova, L. & Blakemore, S.-J. The effects of social deprivation on adolescent development and mental health. *Lancet Child Adolesc. Health* **4**, 634–640 (2020).

34. Taub, A. A new covid-19 crisis: domestic abuse rises worldwide. *The New York Times* https://www.nytimes.com/2020/04/06/world/coronavirus-domestic-violence.html (6 April 2020).

35. Abramian, J. The Covid-19 pandemic has escalated domestic violence worldwide. *Forbes* https://www.forbes.com/sites/jackieabramian/2020/07/22/the-covid-19-pandemic-has-escalated- global-domestic-violence/#57366498173e (22 July 2020).

36. Tsamakis, K. et al. Oncology during the COVID-19 pandemic: challenges, dilemmas and the psychosocial impact on cancer patients (review). *Oncol. Lett.* **20**, 441–447 (2020).

37. Raymond, E., Thieblemont, C., Alran, S. & Faivre, S. Impact of the COVID-19 outbreak on the management of patients with cancer. *Target. Oncol.* **15**, 249–259 (2020).

38. Couzin-Frankel, J., Vogel, G. & Weiland, M. School openings across globe suggest ways to keep coronavirus at bay, despite outbreaks. *Science* https://www.sciencemag.org/news/2020/07/school-openings-across-globe-suggest-ways-keep-coronavirus-bay-despite-outbreaks# (2020).

39. Vardoulakis, S., Sheel, M., Lal, A. & Gray, D. Covid-19 environmental transmission and preventive public health measures. *Aust. N. Z. J. Public Health* **44**, 333–335 (2020).

40. Saadat, S., Rawtani, D. & Hussain, C. M. Environmental perspective of Covid-19. *Sci. Total Environ.* **728**, 138870 (2020).

41. Covid-19 Government Measures Dataset (ACAPS, 2020); https://www.acaps.org/covid19-government-measures-dataset

42. Brockmann, D. & Helbing, D. The hidden geometry of complex, network-driven contagion phenomena. *Science* **342**, 1337–1342 (2013).

43. Guan, D. et al. Global supply-chain effects of Covid-19 control measures. *Nat. Hum. Behav.* **4**, 577–587 (2020).

44. Malmgren, J., Guo, B. & Kaplan, H. G. Covid-19 confirmed case incidence age shift to young persons aged 0–19 and 20–39 years over time: Washington State March–April 2020. Preprint at *medRxiv* https://doi.org/10.1101/2020.05.21.20109389 (2020).

45. Gentilini, U., Almenfi, M., Orton, I. & Dale, P. *Social Protection and Jobs Responses to COVID-19* (World Bank, 2020); https://openknowledge.worldbank.org/handle/10986/33635

46. Cleaning and Disinfection of Environmental Surfaces in the Context of COVID-19 (World Health Organization, 2020); https://www.who.int/publications/i/item/cleaning-and-disinfection-of-environmental- surfaces-inthe-context-of-covid-19

47. Shen, J. et al. Prevention and control of COVID-19 in public transportation: experience from China. *Environ. Pollut.* **266**, 115291 (2020).

48. Islam, N. et al. Physical distancing interventions and incidence of coronavirus disease 2019: natural experiment in 149 countries. *BMJ* **370**, m2743 (2020).

49. Liu, X. & Zhang, S. Covid-19: face masks and human-to-human transmission. *Influenza Other Respir. Viruses* **14**, 472–473 (2020).

50. *2019 Novel Coronavirus COVID-19 (2019-nCoV) Data Repository by Johns Hopkins CSSE* (Johns Hopkins University of Medicine, 2020); https://github.com/CSSEGISandData/COVID-19

51. Griffin, J. et al. A rapid review of available evidence on the serial interval and generation time of COVID-19. Preprint at *medRxiv* https://doi.org/10.1101/2020.05.08.20095075 (2020).

52. Hale, T., Webster, S., Petherick, A., Phillips, T. & Kira, B. *Oxford COVID-19 Government Response Tracker* (Blavatnik School of Government & University of Oxford, 2020); https://www.bsg.ox.ac.uk/research/research-projects/coronavirus-government-response-tracker

53. Zheng, Q. et al. HIT-COVID, a global database tracking public health interventions to COVID-19. *Sci. Data* **7**, 286 (2020).

54. Vaswani, A. et al. in *Advances in Neural Information Processing Systems 30* (eds Guyon, I. et al.) 5998–6008 (Curran Associates, 2017).

55. Cori, A., Ferguson, N. M., Fraser, C. & Cauchemez, S. A new framework and software to estimate time-varying reproduction numbers during epidemics. *Am. J. Epidemiol.* **178**, 1505–1512 (2013).

56. Valka, F. & Schuler, C. Estimation and interactive visualization of the time-varying reproduction number R_t and the time-delay from infection to estimation. Preprint at *medRxiv* https://doi.org/10.1101/2020.09.19.20197970 (2020).

Acknowledgements

Author contributions

N.H., L.G., A.L., V.L. and P.K. conceived and performed the analyses. V.L., S.T. and P.K. supervised the study. E.D. contributed additional tools. N.H., L.G., A.L., A.D.-L., B.P. and P.K. wrote the first draft of the paper. A.D.-L. supervised data collection on NPIs. All authors discussed the results and contributed to revision of the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41562-020-01009-0.

Supplementary information is available for this paper at https://doi.org/10.1038/s41562-020-01009-0.

Correspondence and requests for materials should be addressed to P.K.

Peer review information Peer review reports are available. Primary handling editor: Stavroula Kousta.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Score, CCCSL

L1 Theme	L2 Category	Normalized Score
Social distancing	Small gathering cancellation	~0.83
Social distancing	Closure of educational institutions	~0.73
Travel restriction	Border restriction	~0.55
Social distancing	Mass gathering cancellation	~0.52
Healthcare and public health capacity	Increase availability of PPE	~0.50
Risk communication	Educate and actively communicate with the public	~0.47
Travel restriction	Individual movement restrictions	~0.45
Resource allocation	The government provides assistance to vulnerable populations	~0.43
Risk communication	Actively communicate with managers	~0.42
Travel restriction	Airport restriction	~0.40
Social distancing	Measures for special populations	~0.38
Healthcare and public health capacity	Increase healthcare workforce	~0.37
Resource allocation	Crisis management plans	~0.34
Case identification, contact tracing and related measures	Quarantine	~0.33
Resource allocation	Activate or establish emergency response	~0.34
Travel restriction	National lockdown	~0.30
Case identification, contact tracing and related measures	Enhance detection system	~0.31
Healthcare and public health capacity	Increase in medical supplies and equipment	~0.30
Resource allocation	Police and army interventions	~0.33
Healthcare and public health capacity	Increase patient capacity	~0.30
Risk communication	Travel alert and warning	~0.28
Healthcare and public health capacity	Adapt procedures for patient management	~0.22
Social distancing	Special measures for certain establishments	~0.21
Travel restriction	Public transport restriction	~0.20
Resource allocation	Measures to ensure security of supply	~0.18
Healthcare and public health capacity	Research	~0.18
Risk communication	Actively communicate with healthcare professionals	~0.18
Healthcare and public health capacity	Personal protective measures	~0.17
Case identification, contact tracing and related measures	Tracing and tracking	~0.16
Case identification, contact tracing and related measures	Border health check	~0.15
Travel restriction	Cordon sanitaire	~0.14
Social distancing	Return operation of nationals	~0.12
Resource allocation	Provide international help	~0.12
Travel restriction	Port and ship restriction	~0.12
Social distancing	Work safety protocols	~0.11
Case identification, contact tracing and related measures	Isolation of cases	~0.10
Case identification, contact tracing and related measures	Restricted testing	~0.10
Case identification, contact tracing and related measures	Activate case notification	~0.09
Case identification, contact tracing and related measures	Surveillance	~0.09
Healthcare and public health capacity	Repurpose hospitals	~0.07
Environmental measures	Environmental cleaning and disinfection	~0.06
Case identification, contact tracing and related measures	Airport health check	~0.06
Social distancing	Measures for public transport	~0.04
Healthcare and public health capacity	Increase isolation and quarantine facilities	~0.04

Extended Data Fig. 1 | Main results for the CCCSL dataset. Normalised scores (relative effect within a method) of the NPI categories in CCCSL, averaged over the four different approaches.

Extended Data Fig. 2 | Main results for the CoronaNet dataset. Normalised scores (relative effect within a method) of the NPI categories in CoronaNet, averaged over the four different approaches. Full names of the abbreviated L2 categories can be looked up in SI; Supplementary Table 3.

Score, WHOPHSM

Normalized Score

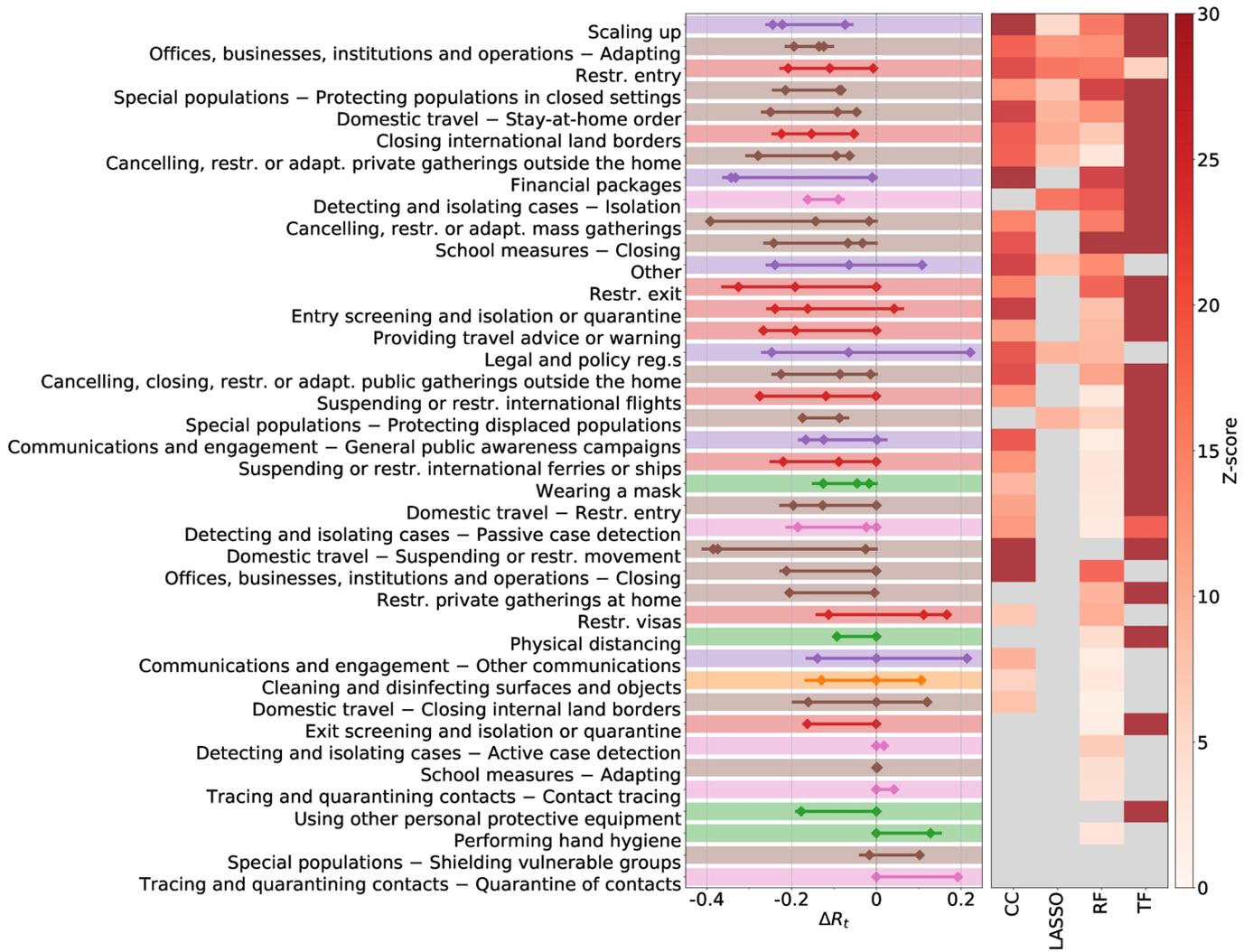| L1 Theme	L2 Category	Normalized Score
Other measures	Financial packages	~0.68
Other measures	Scaling up	~0.66
International travel measures	Restricting entry	~0.62
Social and physical distancing measures	Special populations -- Protecting populations in [...]	~0.50
Social and physical distancing measures	Domestic travel -- Stay-at-home order	~0.50
Social and physical distancing measures	Offices, businesses, institutions and operations -- Adapting	~0.48
Surveillance and response measures	Detecting and isolating cases -- Isolation	~0.44
Social and physical distancing measures	School measures -- Closing	~0.42
International travel measures	Restricting exit	~0.42
Other measures	Other	~0.43
International travel measures	Entry screening and isolation or quarantine	~0.40
Social and physical distancing measures	Offices, businesses, institutions and operations -- Closing	~0.39
Social and physical distancing measures	Gatherings, businesses and services -- Cancelling, [...]	~0.38
Other measures	Legal and policy regulations	~0.38
Social and physical distancing measures	Domestic travel -- Suspending or restricting movement	~0.38
International travel measures	Closing international land borders	~0.40
Social and physical distancing measures	Gatherings, businesses and services -- Cancelling, [...]	~0.38
International travel measures	Providing travel advice or warning	~0.40
Social and physical distancing measures	Special populations -- Protecting displaced populations	~0.37
International travel measures	Suspending or restricting international flights	~0.38
Social and physical distancing measures	Gatherings, businesses and services -- Cancelling, [...]	~0.35
Other measures	Communications and engagement -- General public [...]	~0.27
International travel measures	Suspending or restricting international ferries or ships	~0.32
Social and physical distancing measures	Gatherings, businesses and services -- Restricting [...]	~0.30
Individual measures	Wearing a mask	~0.30
Surveillance and response measures	Detecting and isolating cases -- Passive case detection	~0.27
Social and physical distancing measures	Domestic travel -- Restricting entry	~0.27
International travel measures	Restricting visas	~0.26
Individual measures	Physical distancing	~0.24
Other measures	Communications and engagement -- Other communications	~0.22
Environmental measures	Cleaning and disinfecting surfaces and objects	~0.20
Social and physical distancing measures	Domestic travel -- Closing internal land borders	~0.20
International travel measures	Exit screening and isolation or quarantine	~0.20
Surveillance and response measures	Detecting and isolating cases -- Active case detection	~0.19
Surveillance and response measures	Tracing and quarantining contacts -- Contact tracing	~0.19
Individual measures	Using other personal protective equipment	~0.18
Social and physical distancing measures	School measures -- Adapting	~0.19
Individual measures	Performing hand hygiene	~0.18

Extended Data Fig. 3 | Main results for the WHO-PHSM dataset. Normalised scores (relative effect within a method) of the NPI categories in WHO-PHSM, averaged over the four different approaches. Full names of the abbreviated L2 categories can be looked up in SI; Supplementary Table 4.

Extended Data Fig. 4 | Measure effectiveness in the WHO-PHSM dataset. Analogue to Fig. 1 of the main text if the analysis is done on the WHO-PHSM dataset. Full names of the abbreviated L2 categories can be looked up in SI; Supplementary Table 4.
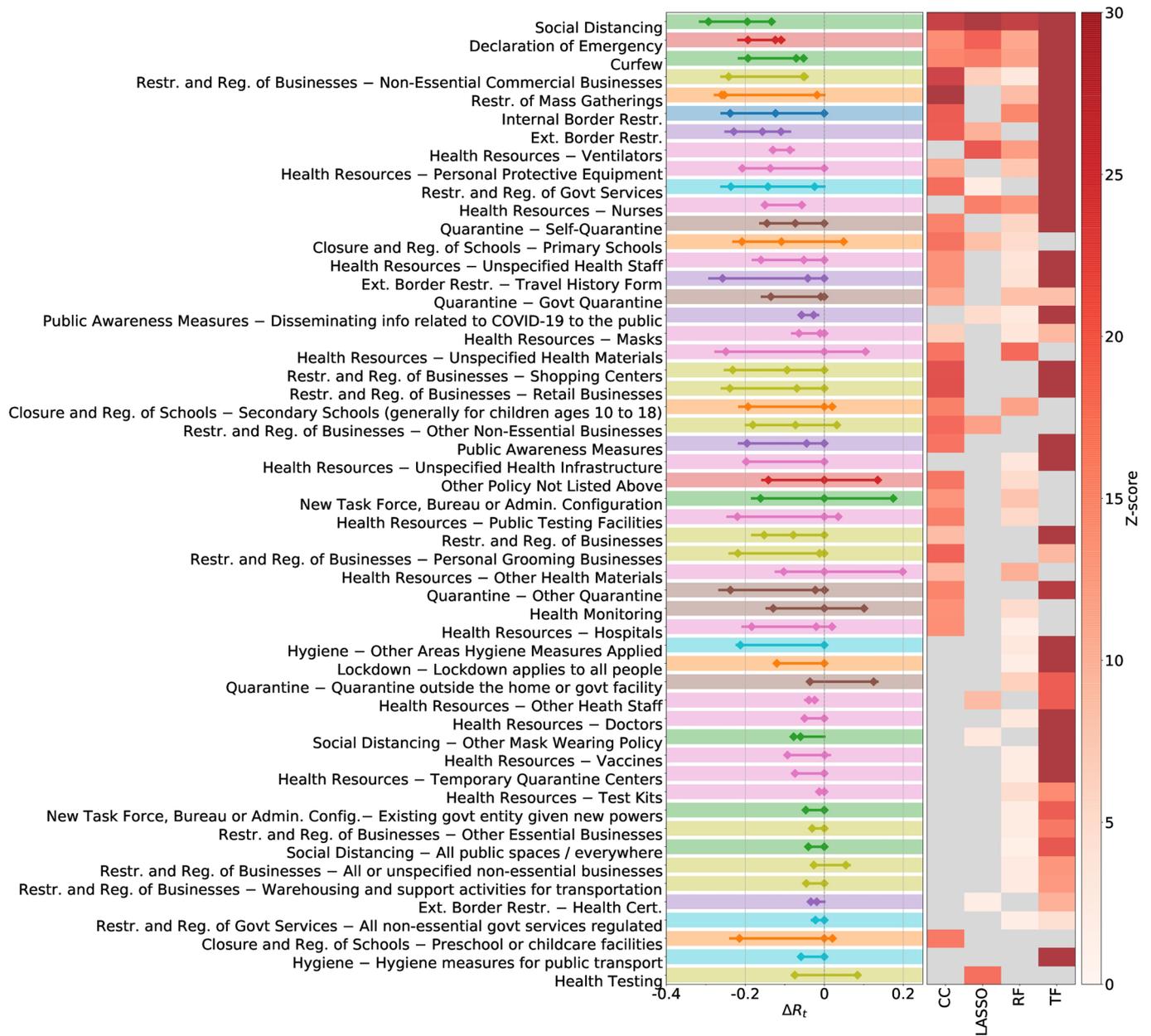
Extended Data Fig. 5 | Measure effectiveness in the CoronaNet dataset (part 1). Analogue to Fig. 1 of the main text if the analysis is done on the CoronaNat dataset (continued in Extended Data Fig. 6). Full names of the abbreviated L2 categories can be looked up in SI; Supplementary Table 3.
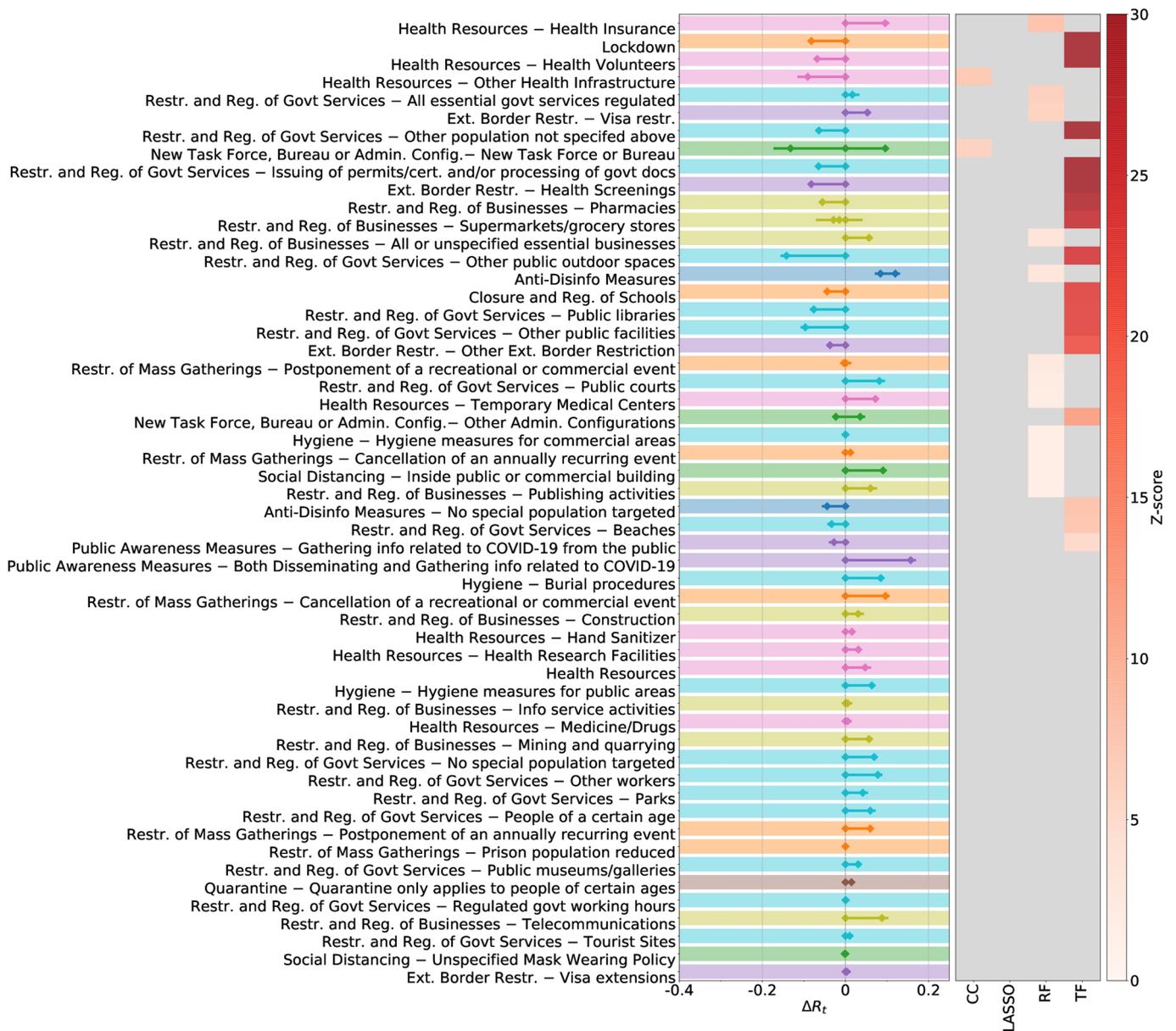
Extended Data Fig. 6 | Measure effectiveness in the WHO-PHSM dataset (part 2). Analogue to Fig. 1 of the main text if the analysis is done on the CoronaNet dataset (continued from Extended Data Fig. 5). Full names of the abbreviated L2 categories can be looked up in SI; Supplementary Table 3.

nature research

Corresponding author(s): Klimek, Peter

Last updated by author(s): Oct 21, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☐	☒	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data and code is available on our github repository, as linked in the text. The analysis was done in Python v3.8 and MatLab R2018a.
Data analysis	Data and code is available on our github repository, as linked in the text. The analysis was done in Python v3.8 and MatLab R2018a.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The NPI datasets are all available for the public online. The versions of these datasets we used can be found on our github repository.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size for the NPIs and countries analysed were determined by the numbers of available records in the datasets.
Data exclusions	We excluded NPIs if they were implemented (observed) less than 10 times.
Replication	We used four independent modelling approaches and three different datasets to replicate our findings.
Randomization	Not applicable.
Blinding	Not applicable.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒ ☐	Antibodies
☒ ☐	Eukaryotic cell lines
☒ ☐	Palaeontology and archaeology
☒ ☐	Animals and other organisms
☒ ☐	Human research participants
☒ ☐	Clinical data
☒ ☐	Dual use research of concern

Methods

n/a	Involved in the study
☒ ☐	ChIP-seq
☒ ☐	Flow cytometry
☒ ☐	MRI-based neuroimaging