nature portfolio

https://doi.org/10.1038/s41562-020-01009-0

Supplementary information

Ranking the effectiveness of worldwide COVID-19 government interventions

In the format provided by the authors and unedited

Supplementary Information:
Ranking the effectiveness of worldwide COVID-19 government interventions

1 Supplementary Methods

Case-control analysis. We perform a case-control regression analysis to quantify the impact of implementing an NPI measure on the effective reproduction number, R_t. The central idea is to compare all countries that have implemented a certain measure with all countries that have not implemented this measure at the same stage of the epidemic while adjusting for several country-specific covariates through regression analysis. These covariates include timing (time-span between the day on which more than 30 cases were confirmed and the day the measure was implemented), baseline R_t (reproduction number before the measure was implemented), size and population density, as well as covariates for how many other measures have already been implemented. It is assumed that implementing a NPI on day t will have measurable impacts on R_t at day $t + \tau$. We also study how various country-specific indicators for economic and human development as well as governance impact on these results.

Exposure variable. We consider each NPI implemented in more than ten countries. We include measures published by [1] on two different resolution levels (L2 and L3) separately. Let T_M be the day on which a country implemented measure M. The covariates and outcome variables are measured relative to this point in time. As an exposure variable, we use a dummy variable, X, encoding whether a country has implemented the measure during the observation window or not.

Covariates. We include the following covariates in all analyses. First, a country's epidemic age when measure M was implemented, $A(M)$, defined as the number of days between the implementation of the measure, T_M, and the first day with more than 30 confirmed cases, denoted by T_0, giving $A(M) = T_M - T_0$. Second, the baseline effective reproduction number, R_t^{BL}, is taken at day $T_M + \tau$. Third and fourth, we include for each country the logarithms of its total population P and population density D [2]. Furthermore, we include the number $N^{All}(T_M)$ of all implemented L2 measures and the number $N^{L2}(T_M)$ of all measures (L2 or L3) from the same categories as M that have been implemented before T_M. These covariates are supposed to capture whether a country introduces the intervention under consideration early or late in relation to its epidemic age and the number of measures already taken.

Outcome variable. As a dependent variable in the regression, we consider the change in effective reproduction number after implementation of the measure, $\Delta R_t(\tau) = R_t^M(\tau) - R_t^{BL}$, where R_t^M is the value of R_t taken on day $T_M + \tau$. We report the average of $\Delta R_t(\tau)$ taken over time lags $0 < \tau \leq 28$.

Matching. As cases we include all countries implementing a given measure at a time $A(M^c) > \tau$. As controls we consider all other countries that have not implemented this measure within the observation window but that implemented any other intervention not more than one day sooner or later than its matched case. These matching criteria ensure that we do not introduce a bias by comparing countries that implemented any response with countries that did nothing at all and maybe had the epidemics already under control.

Additional country variables. We consider variables that capture economic and human development, as well as different dimensions of governance. Economic development is measured by the country's per capita GDP adjusted for power purchasing parity (PPP) [3]. Human development is quantified by the human development index (HDI), an indicator provided by the United Nations Development Programme taking life expectancy, education, and per capita income into account [4]. Finally, we consider six different dimensions of governance as quantified by the World Governance Indicators (WGI) [5]. These indicators include Voice & Accountability (free media, the extent to which citizens participate in the government), Political Stability (including the absence of violence), Government Effectiveness (quality of public services), Regulatory Quality (the ability of government to implement sound policies), Rule of Law (the extent to which citizens abide by the rules of society) and Control of Corruption.

Statistical analysis and multiple testing. A linear regression model of the form,

$$\Delta R_t \sim R_t^{BL} + A(M) + P + D + N^{All}(T_M) + N^{L2}(T_M) + X \quad , \tag{S1}$$

is evaluated for all NPIs meeting the inclusion criteria. We evaluate the above model for all measures and permissible delays τ. We adjust for multiple testing by controlling the false discovery rate at level 0.0001 via the Benjamini-Hochberg procedure. We investigate the impact of the additional country variables by evaluating models as in Eq. (S1) with an additional term for the considered indicator.

Lasso regression. We assume that without any implemented NPIs, the $R(t)$ is constant, and deviations are caused by a time-delayed onset of the effects of each NPI on L2. This hypothesis is an oversimplification of the reality, but it allows to estimate this time-delay itself and to quantify magnitudes of impacts for each NPI. A similar approach has been reported, for instance, in [6], although the authors use a smaller list of NPIs.

Formally, this model for a single territory c is captured in the expression

$$\mathbf{R}_c^{\text{model}}(\boldsymbol{\beta}, \tau) = \big(\beta_{\text{avg}} + \beta_c\big)\mathbf{1} + \sum_{\text{NPIs } M} \beta_M \boldsymbol{\Theta}_{A(M,c)+\tau} \ . \tag{S2}$$

Here, bold symbols indicate vectors containing discrete time-series, starting at 30 confirmed cases for each country. The vector $\boldsymbol{\Theta}_t$ models a discrete version of the Heaviside function with $\boldsymbol{\Theta}_t = \big(0, \ldots, 0, 1, \ldots, 1\big)$, with the step from 0 to 1 at position t. In Eq. (S2), $A(M, c)$ indicates the day when a NPI M is implemented in territory c, and τ is the offset in days to account for the time-delay until the NPIs affect the case numbers. The first term in Eq. (S2) is a constant composed of the average effects β_{avg} and potential territory-specific effects β_c. For inference, we concatenate all territories into one large vector. Thus, in total we need to estimate 126/326/265 coefficients $\boldsymbol{\beta} = \big(\beta_{\text{avg}}, \beta_{c_1}, \ldots, \beta_{M_1} \ldots\big)$ (1 constant, 79/218/224 territories and 46/107/40 NPIs for CCCSL/CORONANET/WHOPHSM) in this model.

Regression is carried out by minimizing the target function

$$\min_{\boldsymbol{\beta}} \left\{ \left\| \mathbf{R}^{\mathrm{model}}(\boldsymbol{\beta}, \tau) - \mathbf{R}^{\mathrm{data}} \right\|_2^2 + \alpha \left\| \boldsymbol{\beta} \right\|_1 \right\} \ , \tag{S3}$$

to obtain values for $\boldsymbol{\beta}$. The first term is the Residual Sum of Squares (RSS) of the model with respect to actual observation data. The second term penalises too large values of the coefficients $\boldsymbol{\beta}$, with α a penalty parameter that indicates the weight of this penalization in the optimization procedure. Here, $\left\|\cdot\right\|_2$ and $\left\|\cdot\right\|_1$ denote the 2-norm (euclidean distance) and 1-norm (sum of absolute values), respectively. The 1-norm for the penalizing term is characteristic for lasso regression [7], which acts as feature selection by estimating multiple coefficient values $\boldsymbol{\beta}$ as zero. Formally, this additional penalty term can be shown to be equivalent to assuming a Laplace-distribution prior on all values in $\boldsymbol{\beta}$. Thus, the tuning parameter α effectively balances the trade-off between two objectives: the goodness of fit to the data (RSS) with the complexity of the model (second term) [7,8].

Cross-validation and meta-parameter search. This model includes two meta-parameters (τ, α), which are estimated by cross-validation to obtain a minimal RSS: At first, countries are randomly assigned to one of 10 groups. Then, each of these country groups is dropped from the vectors $\mathbf{R}$, and coefficients $\boldsymbol{\beta}$ are estimated via the minimization in Eq. (S3). These coefficients $\boldsymbol{\beta}$ from the training group are used to compute an R^2_{Test} value of the model on the dropped countries to test how good the model can predict previously unseen observations. As the different country groups can contain a different number of observations, we compute the overall R^2_{Test} for a given set of (τ, α) as

$$R^2_{\mathrm{Test}}(\tau, \alpha) = 1 - \sum_{\mathrm{groups}\ g} \mathrm{RSS}_g(\tau, \alpha) / \sum_{\mathrm{groups}\ g} \mathrm{Var}(\mathbf{R}_g) \ , \tag{S4}$$

which weighs the individual coefficients of determination r^2 for each test with the variance in the reproduction number $\mathbf{R}_g$ within the test group. This whole procedure is carried out on a grid of possible (τ, α) values, to find a set of meta-parameters, where the model Eq. (S2) can best describe the data. Supplementary Figure 2 depicts the resulting values of R^2_{Test} for a sufficient large range of meta-parameters. As the overall curve for R^2_{Test} is relatively flat, we repeat the cross-validation 30 times with different assignments to the 10 country groups, which reduces the overall noise. The final R^2_{Test} is then computed using the same weighting procedure in Eq. (S4) for the set of all 20 repetitions and each of the 10 territory groups within a repetition. We find that a time-offset of $\tau = 11/11/15$ days and a penalty parameter $\alpha = 0.005623/0.00100/0.004217$ are the best parameters to describe the observations with this model for the CCCSL/CORONANET/WHOPHSM, see Supplementary Figure 2.

Final coefficient estimation. We estimate the ranges of each NPI effectiveness using a territory-wise cross-validation, i.e., by reducing the sizes of test groups to one. For each estimation leaving out one territory, we compute the coefficients β_M for each NPI M and β_c for all territories c. We identify these coefficients β_M and β_c as the change ΔR_t. Shown intervals contain 95% of all observed values, with the center indicating the median. From the overall 126/355/305 coefficients $\boldsymbol{\beta}$, the feature selection aspect of lasso regression finds 23/19/23 relevant NPI coefficients and 1/1/23 relevant territory coefficients, while other coefficients are estimated as zero (for CCCSL/CORONANET/WHOPHSM) Note that this does not indicate that these NPIs are useless to reduce the spread of the virus, but rather that the algorithm can explain all observations with a smaller number of coefficients. The number of selected coefficients is highly sensitive to the value of the penalty parameter, however, the more impactful NPIs are consistently present with similar coefficients (see Supplementary Figure 2 CFI).

Random forest regression. We use random forest regression [9] as a third method to assess the impact of the implemented measures on the spreading of COVID-19, measured in terms of the effective reproduction number R_t.

We represent for each country and each day of the observation period the NPIs which have been implemented in that country until that day in the form of a binary vector. The analysis is performed on the NPI categories (L2). The binary data on implemented NPIs in a given country on day t is regressed on the value of R_t in that country τ days later, $R_{t+\tau}$. The time shift τ accounts for the time delay between infection and case confirmation. We vary τ between $\tau = 0$ and $\tau = 20$.

Each random forest consists of 500 decision trees with maximum depth d. At each split of a node in one of the trees in the random forest, m randomly selected NPI categories out of the total number of 46 categories are considered. We employ bootstrapping, meaning that each decision tree is fitted on a different random subset containing 75% of the rows of the predictor matrix.

To implement the random forest regression, we use the RandomForestRegressor class from the python library scikit-learn [10].

Cross validation. For each $0 \leq \tau \leq 20$, we use 10-fold country-wise cross validation to determine the optimal values of the maximum tree depth d and percentage of considered features m. We vary the parameters d and m in the range $1 \leq d \leq 15$ and $1 \leq m \leq 100$. For each combination of d and m, we randomly split the set of all territories into a training set and a test set. The random forest is trained on the training set data; subsequently, we measure the performance of this random forest in predicting the time series of $R_{t+\tau}$ for the countries in the test set. As a performance metric, we quantify the difference between the predicted time series of $R_{t+\tau}$ and the observed one by using the coefficient of determination r^2. We repeat the same procedure 10 times with different random splits of the set of territories into training and test set. Then we take the mean of the coefficient of determination over the 10 splits to obtain the average out-of-sample performance of the random forest for this combination of d and m. The heatmap in Supplementary Figure 3 shows the dependence of the performance of the random forest on unseen data depending on the parameters d and m for fixed time shift $\tau = 10$.

Feature importance. To quantify the importance of a NPI, we measure the loss of predictive performance of the random forest if the information carried by the given NPI is replaced by noise. This measure of feature importance is also known as permutation importance. If the loss in performance is high for a given NPI, then we conclude that knowledge of the implementation status of this NPI is important for predicting $R_{t+\tau}$.

Specifically, for each time shift $0 \leq \tau \leq 20$ and each NPI M, we measure the reduction in performance of the random forest in predicting $R_{t+\tau}$ for unseen data when the values in the column corresponding to M are randomly shuffled. The maximum tree depth d and the percentage m of features considered are set to their optimal values as determined in the cross-validation step. To obtain sharp estimates for the permutation importance of the different NPIs we use repeated 10-fold country-wise cross validation with 100 repetitions. Different NPIs attain their maximum value for different values of τ.

Transformers modelling. The temporal and dynamic nature of epidemic propagation finds a suitable tool to model such a temporal evolution in Transformers [11]. This neural network can recall past events presented at the input by leveraging its innate ability to take into account the previous information. The intrinsic Transformer architecture ensures this ability for all the temporal data is represented and considered at once without recurrent connections as in Recurrent Neural Networks. At a generic time t, to approximate the value of the effective reproduction number at time $t+1$, R_{t+1}, we use a Transformer whose input is the daily representation of the adopted measures in a given country in binary form, similar to the encoding performed in Random Forest, along with the value of R_t measured in the same day. The best performing network has been identified as having four hidden layers of 128 neurons, an embedding size of 128, 8 heads, one output described by a linear output layer, and 54 inputs ($measures \oplus R_t$, where $\oplus$ is the concatenation operator). NPIs are represented through a theta function over time, whose step position corresponds to the day when the specific measure has been adopted in the selected country. For the neural network training, 10 out of 78 territories were chosen as the validation set, and the best neural network was found when the mean square error between the values of the validation set and the corresponding network outputs reached the minimum amount. Different Transformers training can lead to slightly different predictions. Although the main trends of forecast maintain a substantial similarity across the various training, the corresponding time evolution may change when some containment measures are perturbed. This issue is possibly due to a large number of local minima affecting the neural network's energetic panorama. To address this problem and to provide an adequate assessment of the importance of the measures, we built an ensemble of trained Transformers with a comparable level of training precision.

Transformer ranking assessment. This approach has been chosen to be consistent with the other methods implemented in this paper, in which the impact of the NPIs is related to the comparison between countries having or not having implemented a measure. Therefore the comparison is performed between the normal prediction of the Transformer after the application of a given NPI and the corresponding prediction when the same measure is removed. The difference between the predictions quantifies the impact of that NPI on the overall behaviour of R_t of the selected country. The final relative importance of the containment measure $meas_i$ is then given by $mean_{TF}\left[\sum_t (R^{TF}(t) - R^{TF}_{meas_i}(t))/T\right]$, where $mean_{TF}(\cdot)$ is the mean operator over the RNN ensemble, $R^{TF}(t)$ is the reproduction number predicted by the TFth network, $R^{TF}_{meas_i}(t)$ is the corresponding temporal prediction when the measure $meas_i$ is removed, and T is the period over which the assessment is performed.

Transformer model at country level. Implementing a Transformer as a general model for describing the dependence on the measures of R_t allows, at the same time, to evaluate the impact of the measures on a single country. Each country is identified by a specific temporal sequence of applied measures and a unique R_t evolution over time. Then, we can interrogate the Transformer to assess the importance of every single measure for that particular context by removing the measure under investigation from the specific sequence as in the general knockout case and building an associated ranking of NPIs for each country. For example, Supplementary Figure 4 shows the impact on R_t of the NPI "Small gathering cancellation" in Italy when such a measure is removed from the temporal sequence of adopted NPIs. This approach allows us to assess the country-specific NPI ranking. By comparing these ranking across all countries, we measure the specificity of the impact of its measures.

What-If experiments with Transformers. As mentioned above, Transformers offer the possibility to perform what-if experiments, i.e., exploring different scenarios corresponding to time-series of events different from the actual ones. This approach opens the possibility to examine scenarios corresponding to different combinations of NPIs. In the specific case of the Non-Pharmaceutical Interventions to mitigate the spread of COVID-19, starting from the actual sequence of NPIs adopted by one particular country, we tested the efficacy of artificial sequences by removing specific NPIs or shifting their adoption to other days. In this framework, one can compare the actual evolution obtained through the synthetic sequence of actions to suitable reference sequences. More specifically, we adopted the Transformers to assess what would have happened if a given NPI had been adopted on a different day with respect to the actual day of adoption. To this end, we adopted the following procedure:

- for each country and each NPI, we first compute a knockout evolution of the system by letting the Transformer simulate the evolution of the system once the specific NPI has been removed.
- we create a synthetic sequence by keeping the sequence of NPIs of the country untouched except for the specific NPI that is positioned in a generic day t_i. For that sequence, we compute the evolution of the system as given by the Transformer, i.e., we calculate the time evolution of R_t for the specific synthetic sequence.
- We repeat the above operation for a generic choice of the day of adoption, t_i, of the specific NPI.
- For each country and each NPI, we compute the average difference between the synthetic evolution of R_t with the specific NPI positioned at t_i and the knockout evolution over the 30 days following t_i. The average here is performed over several realisations of the Transformers.
- The outcome is a series of curves that, for each NPI, report the variation of R_t, ΔR_t, averaged over all the countries that adopted that NPI and over several realisations of the Transformers.

Supplementary Figure 5 shows the effect of the selected NPI "Small gathering cancellation" in Italy if it had been adopted on different days, compared to the case where the same NPI is absent (knockout evolution). The reduction of R_t is visible, while the overall impact decreases as one shifts the day of adoption ahead. The Transformer evolution, when the selected NPI is adopted on any other day, takes into account the effect of all the other NPIs adopted in Italy and kept fixed throughout the simulation.

Supplementary Figure 6 reports the evolution of ΔR_t for a selection of NPIs that display a "the earlier, the better" behaviour, that is, whose ability to reduce R_t tends to decrease with the epidemic age of their adoption.

Obtaining results from the Transformer We trained 20 different Transformers after selecting the best out of 100 training procedures to build a representative ensemble of networks to address multiple local minima issues. All the Transformers reached similar values of test loss with a relative deviation of less than 5%. The NPIs impact evaluation is performed by averaging the difference between the prediction and the prediction when a given measure is removed. The area under such a difference normalized by the temporal amplitude represents the mean variation of $\Delta R(t)$.

2 Supplementary Results

Results from the Case-control Analysis on country covariate effects. The impact of different country variables, including demographic and governance indicators, as well as measures for human and economic development, are shown in Supplementary Figure 7. This heatmap presents the average standardised coefficient values from all regression models for a given NPI category (L2) that yield a significant impact on $R(t)$ for various delays (the average is taken over the delays). The standardised coefficient values are the coefficient values (estimated change in R_t) divided by its standard error (SE); the t-Statistic. The significant NPIs show effect sizes in reducing R_t of about 5 standard errors. Note that no statements concerning causal relationships can be made based on these findings; we report correlations from a cross-sectional analysis.

A number of NPIs appear to be less effective in countries with a high population density and low GDP [12]. Human development, as quantified by the Human Development Indicator, has no or only very mild impacts on measure effectiveness. The governance indicators Political Stability and Voice & Accountability are negatively correlated with the effectiveness of most measures. Most of the remaining governance indicators show either no or moderately negative correlations. There are, however, exceptions for some interventions that show consistently positive correlations with government effectiveness, regulatory quality, rule of law and control of corruption, in particular measures to increase the availability of medical supplies and PPE. It is unlikely that these results are confounded by poor reporting in less developed countries, as such biases would affect R_t values from before and after measure implementation similarly and hence cancel each other out. Although there is no clear pattern concerning the influence of the total number of NPIs already implemented, there seems to be a trend regarding the number of NPIs of the same category. As a general tendency, having already implemented measures from a given category makes additional NPIs from the same category more effective, hinting at a relationship where one measure (e.g., closure of restaurants) amplifies the effectiveness of other related measures (e.g., home office).

Results from individual methods and datasets. Supplementary Figures 8–19 give the rankings of the different NPIs in descending order according to their ΔR values or feature importance for the three analyzed datasets. The prediction of R_t from the LASSO, RF and Transformer models are shown in Supplementary Figure 20 for selected countries.

Robustness check and validation We check the robustness of the results obtained against removal of the Americas (North and South America), Asia+Oceania, and Europe+Africa in the CCCSL dataset — see Supplementary Figures 21–23. Note that the only African countries in the data set are Ghana, Senegal and Mauritius, and the single Oceanian country is New Zealand. Supplementary Table S6 shows the number of consensus measures (i.e. NPIs with a significant effect in all methods), as well as the expected number of consensus measures one would expect under the assumption that the significant measures would be distributed independently for the different methods.

Two external datasets further corroborate our findings. These datasets also contain dedicated entries for policies related to mask wearing, which appear to show moderate reductions in R_t when compared to the other measures discussed above. We further confirmed the estimated impact of all consensus measures using two external datasets presenting a broader geographical coverage (and therefore counting more individual NPIs). This large analysis is, to date, unique, and proves the robustness of our results. It also shows that, although the NPI trackers have been built independently, sometimes for different purposes, and present different semantics and structure (which is a limitation to harmonizing results), their analysis provides convergent results with four methods, and the use of a smaller number of countries (56 in the CCCSL versus ~ 200 in the CoronaNet and WHO-PHSM datasets) does not qualitatively affect the outcome. This finding is of importance when analysis of government policies needs to be conducted in emergency, to save computational time.

3 Supplementary Discussion

Social distancing. Bans of small gatherings (gatherings of 50 persons or less) and the closure of educational institutions have a more substantial effect on R_t (but are also more intrusive to our daily lives) than the prohibition of mass gatherings, measures targeting special populations (e.g., elderly, vulnerable populations, hospitalised patients, prisoners or more exposed non-healthcare professionals) or adaptive measures for certain establishments (e.g., places of worship, administrative institutions, entertainment venues, nursing homes). While in earlier studies based on smaller numbers of countries, school closures had been attributed only a little effect on the spread of COVID-19 [6,13], more recent evidence has been in favour of the effectiveness of this NPI [14,15]. This is also in line with a contact tracing study from South Korea which identified adolescents aged 10–19 as the biggest spreaders in household settings [16]. Social distancing measures are less effective in countries with a high population density and a high degree of citizen participation in the government, as well as freedom of expression or free media (WGI Voice & Accountability). These results indicate by no means a causal relationship but they might be confounded by the fact that some south-east Asian countries with authoritarian tendencies managed to contain the virus particularly well in the early phases of the pandemic. The country-level analysis confirms that these NPIs have a particularly high entropy, meaning that their effectiveness varies indeed substantially across countries. An exception to that are the measures for public transport and work safety protocols, where the latter mostly refers to mandatory guidelines for, e.g., physical barriers or fever checks at workplaces. These two social distancing measures have a low effectiveness rank (little significance across the methods) and low entropy, meaning that they had no impact on R_t

consistently across most countries.

Healthcare and public health capacity. An increase in the availability of personal protective equipment (PPE) to the healthcare workforce, together with measures aiming to reduce the number of non-COVID-19 or non-critical COVID-19 patients in medical centres and hospitals (by promoting self-isolation of mildly symptomatic patients, setting up health hotlines, etc.) are also essential building blocks of successful containment strategies. All of these measures combine high effectiveness of early implementation and low entropy, meaning that they are similarly effective in most countries. Actions aiming to enhance the health system are critical. They are the primary response to patients and have no (or few) negative repercussions on individual rights of liberty (exception on the travel restriction for healthcare workers imposed by several countries). Our results demonstrate that government support to the health system needs to be a priority during a health crisis in order to reduce mortality [17]. In line with our result "the earlier, the better", we argue that those actions must be taken early enough to prepare for a surge in healthcare demand . Compared to other interventions, increased medical supplies and availability of PPE show substantially stronger positive correlations with several governance indicators including government effectiveness and control of corruption. Indeed, there are increased news reports currently on scandals related to government procurement of PPE [18–20].

Travel restrictions. Different types of travel restrictions also show significant effects, in particular border restrictions (e.g., border closure, border controls), individual movement restrictions (e.g., curfews, the prohibition of non-essential activities) and cordons sanitaires (containment zones).

The high effectiveness of border restrictions is driven by European countries (its impact on R_t turns insignificant in two of our methods after removing all European countries); most likely for geographic reasons. This finding is in line with a high entropy score of border, airport, port and ship as well as individual movement restrictions.

Effectiveness of ultimate measures such as stay-at-home orders or lockdowns is still controversial. Recent studies suggest that a national lockdown reduces R_t by an average of 5% [6] to 80% [13], whereas other interventions seem to reduce the virus spread by 5% [13] to 30% [6]. In some countries or territories, the effect of a lockdown decided in the late stage of the epidemic may not be more effective than previously implemented bans on gatherings [6,13,21]. Our analysis highlights the importance of early national lockdowns by showing how the relative effectiveness of that measure correlates with the epidemic age of its adoption. However, the reduced effectiveness of lockdowns at higher epidemic age, as observed in Fig. 4, does not necessarily imply that taking this NPI late is useless.

Risk communication. In terms of risk communication, we find that pro-active communication with stakeholders from the private sector (e.g., business owners or chief executive officers) to promote voluntary safety protocols in enterprises, businesses, event organization, government administrations, etc., shows a significant effect in each of the four analyses, mainly when implemented early. Three out of four approaches also indicate a substantial impact of public health communication strategies (i.e., non-binding NPIs) encouraging citizen engagement and empowering them with information.

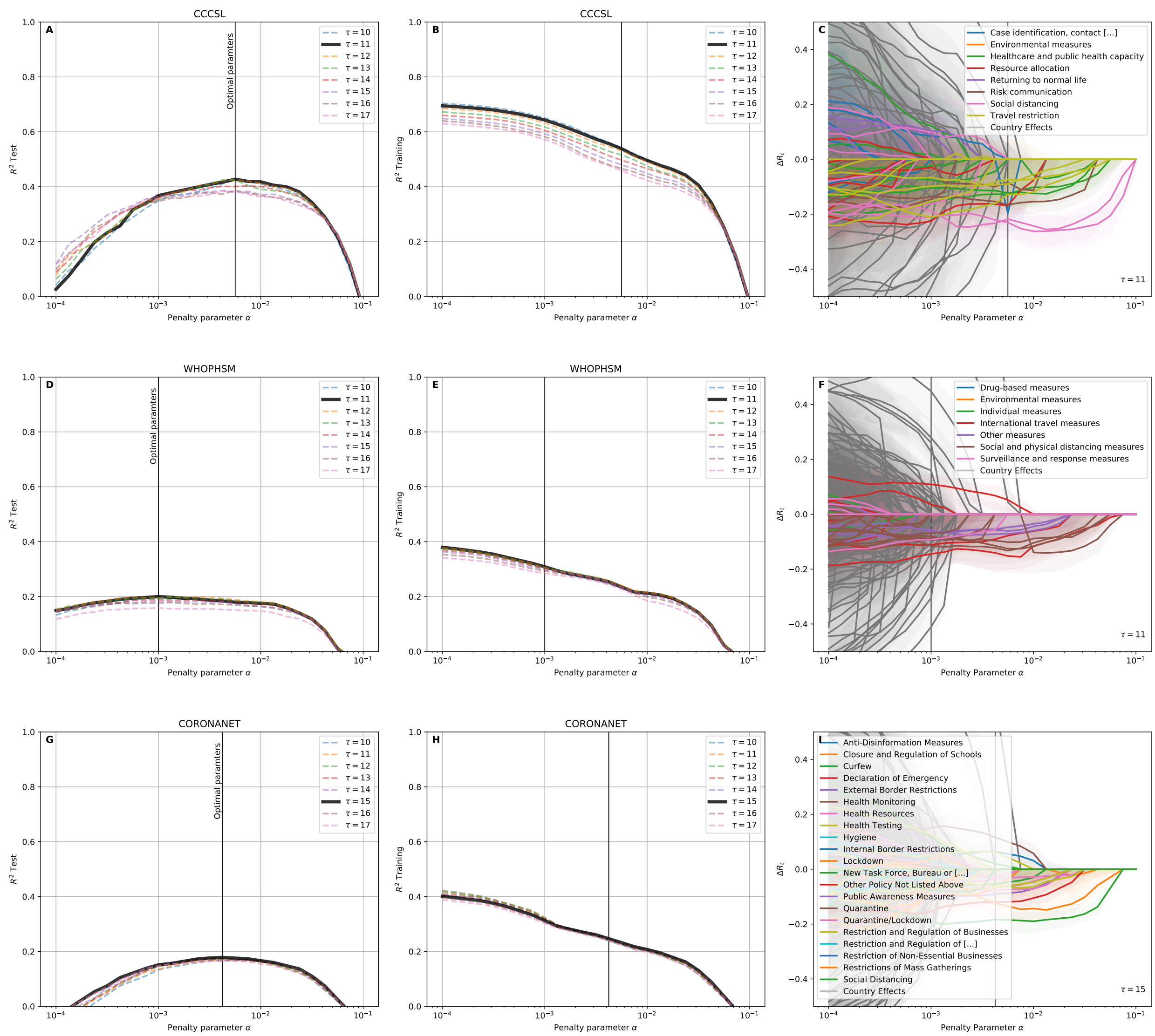
Resource allocation. Measures for resource allocation show limited impacts on R_t in our analysis (e.g., police and army interventions being insignificant in all studies) with relatively high entropy, meaning that country-level effects are important. Surprisingly, the implementation of crisis management plans turns out to be highly effective, except for the Americas. After removing countries from North and South America from the analyses, all four of our methods agree on significant effects of crisis management plans with an ΔR_t of down to -0.3, suggesting a lack of effective crisis plans in American territories. For instance, US states had to focus on providing health insurance and economic stimulus as well as facilitating administrative procedures, while European countries could develop their plans on top of a stronger socio-economic basis [22,23]. Crisis management plans are also more effective in countries with a non-participatory government, meaning that countries with increasingly authoritarian practices might be at an advantage at implementing such policies, as can be seen in the swift response of Singapore [24].

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

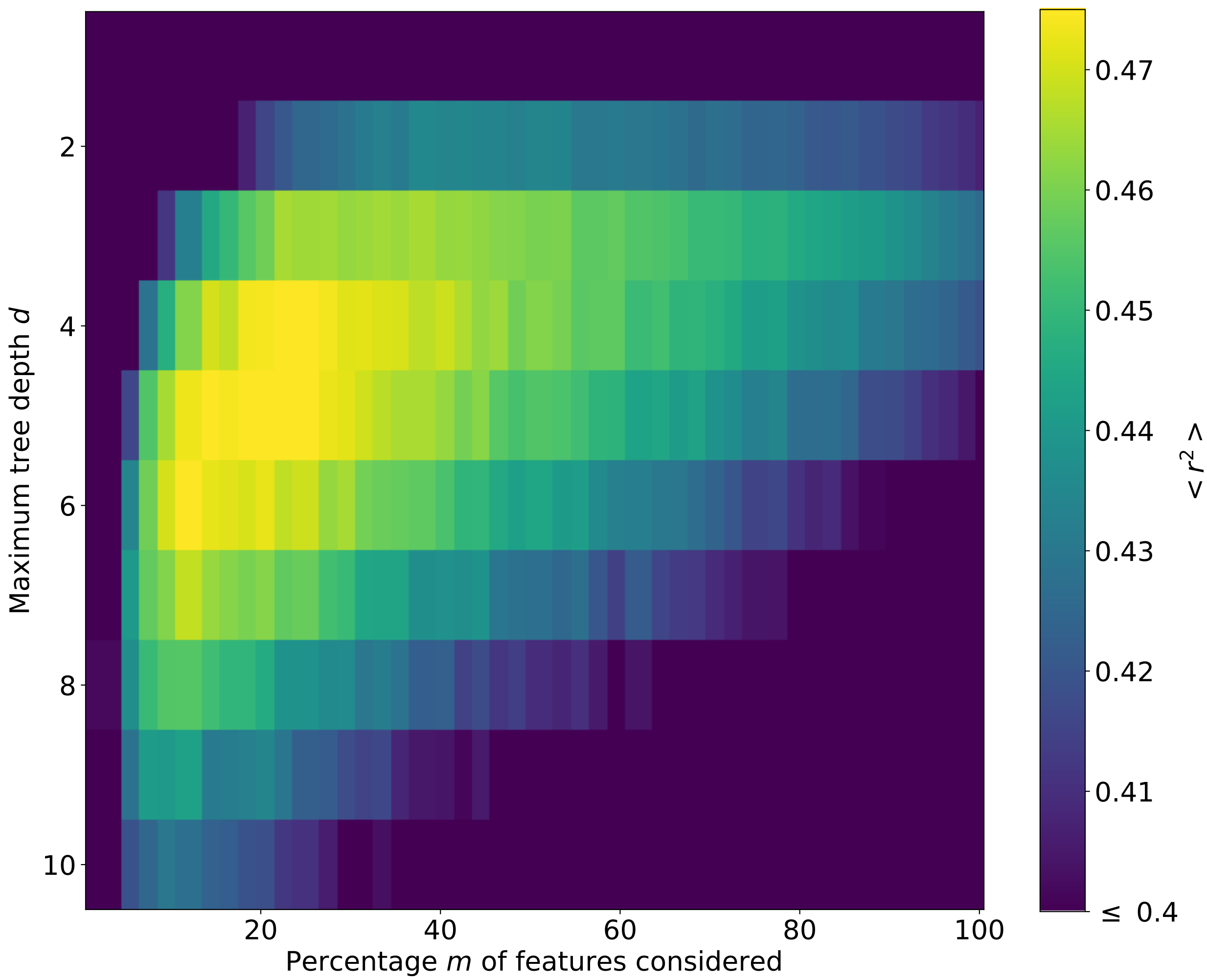
4 Supplementary Figures

L1 themes	CC	LASSO	RF	TF
Social distancing	**1**	**1**	**1**	**1**
Travel restriction	**2**	**2**	**2**	**2**
Healthcare and public health capacity	5	4	4	**3**
Risk communication	6	**3**	**3**	5
Resource allocation	4	6	5	4
Case identification, contact tracing and related measures	**3**	5	6	6
Environmental measures	7	7	7	7

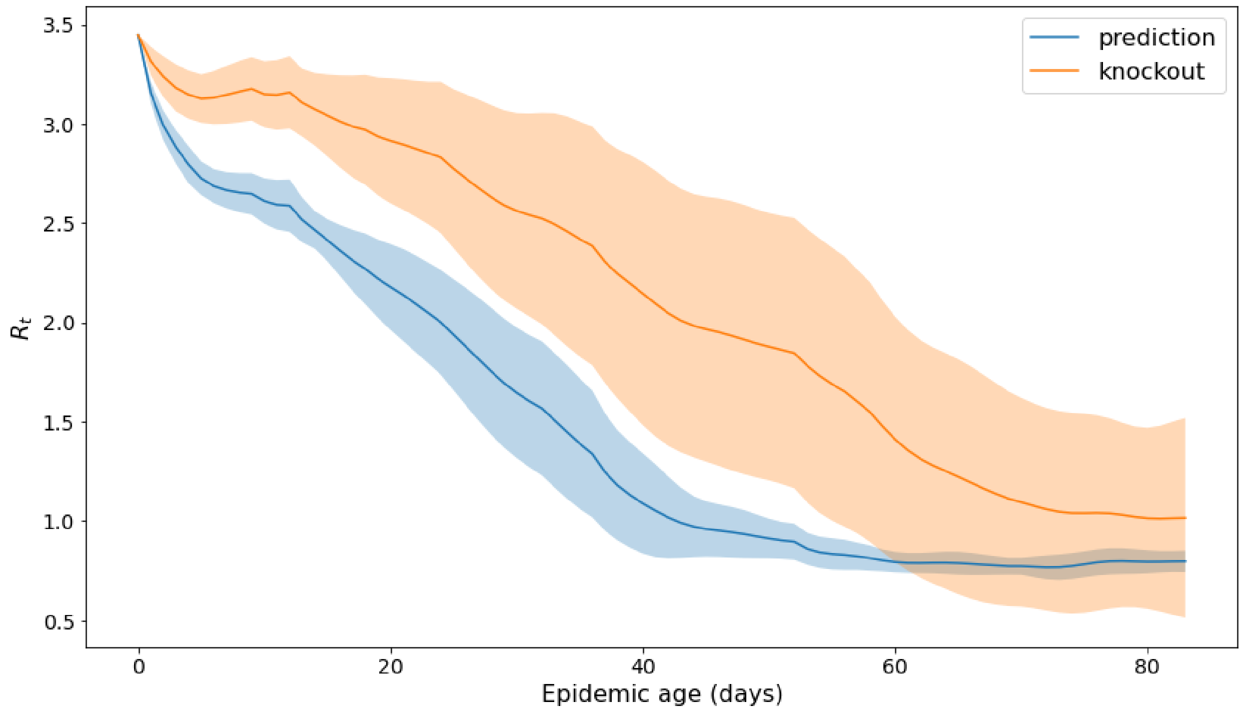
Supplementary Figure 1: Comparison of effectiveness rankings on the coarsest hierarchical level for the case-control analysis (CC), LASSO regression (LASSO), random forest regression (RF), and transformer analysis (TF). To obtain a ranking of the eight different themes (L1) of NPIs, we sum the impacts of the 3 highest ranked categories of each theme and then rank the themes according to this cumulative impact. All methods indicate that NPIs belonging to the themes of social distancing, travel restrictions as well as healthcare and public health capacity lead to the most significant reductions of R_t. Environmental measures (e.g., cleansing public places) are ranked least effective in each approach.



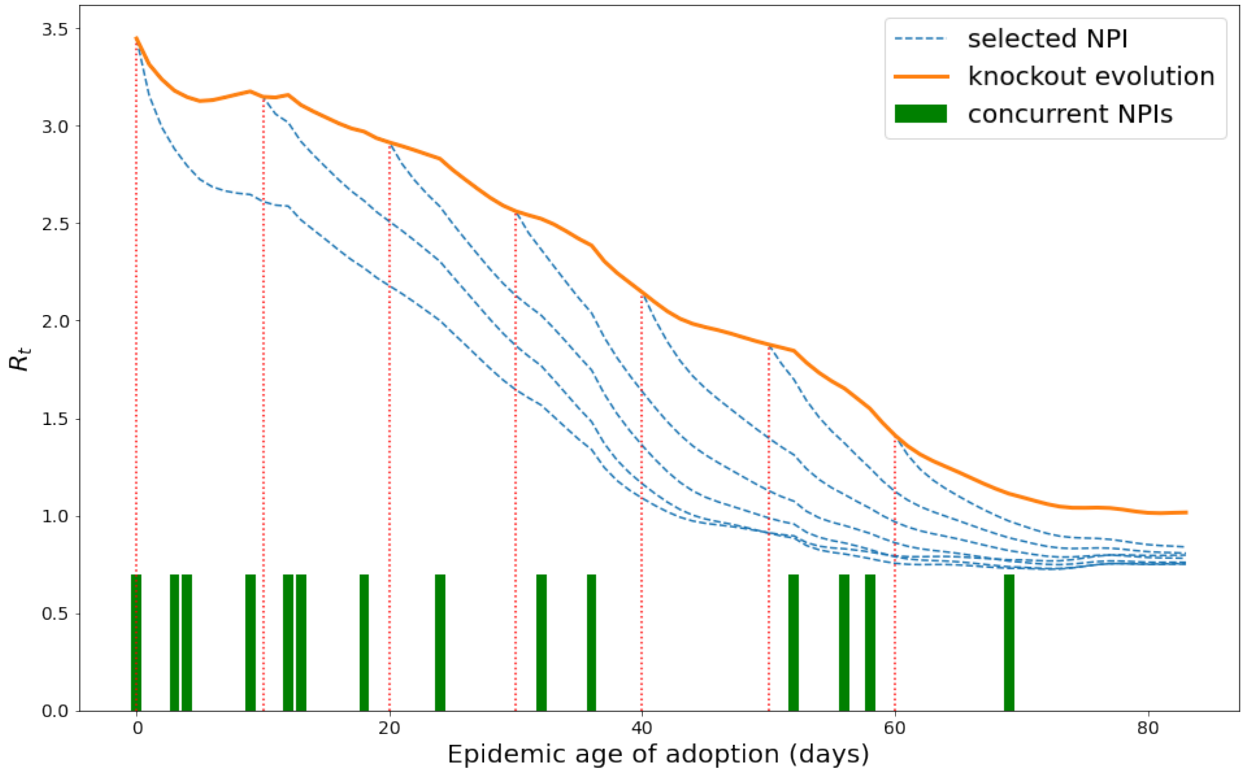
Supplementary Figure 2: Cross-validation results for LASSO regression. (A,B,D,E,G,H) R^2 values for the test and training data, averaged over 30 repetitions of assigning countries into 10 groups for each of the three data sets (CCCSL/CORONANET/WHOPHSM). With smaller penalty parameters α, R^2_{Training} increases as the algorithm uses more NPIs to explain the training data. This overfitting, however, reduces the ability to fit the previously unseen test data. In contrast, for large penalty parameters α, not enough NPIs are used to explain the data, and the method cannot explain either training or test data. The optimal meta parameters are found as $(\tau, \alpha)_{\text{CCCSL}} = (11, 0.005623)$, $(\tau, \alpha)_{\text{CORONANET}} = (15, 0.004217)$ and $(\tau, \alpha)_{\text{WHOPHSM}} = (11, 0.00100)$ (C,F,I) Effects of all NPIs and territories vary when changing the penalty parameter α. However, most of the effects of the significant NPIs stay roughly constant, only the number of significant NPIs increases.



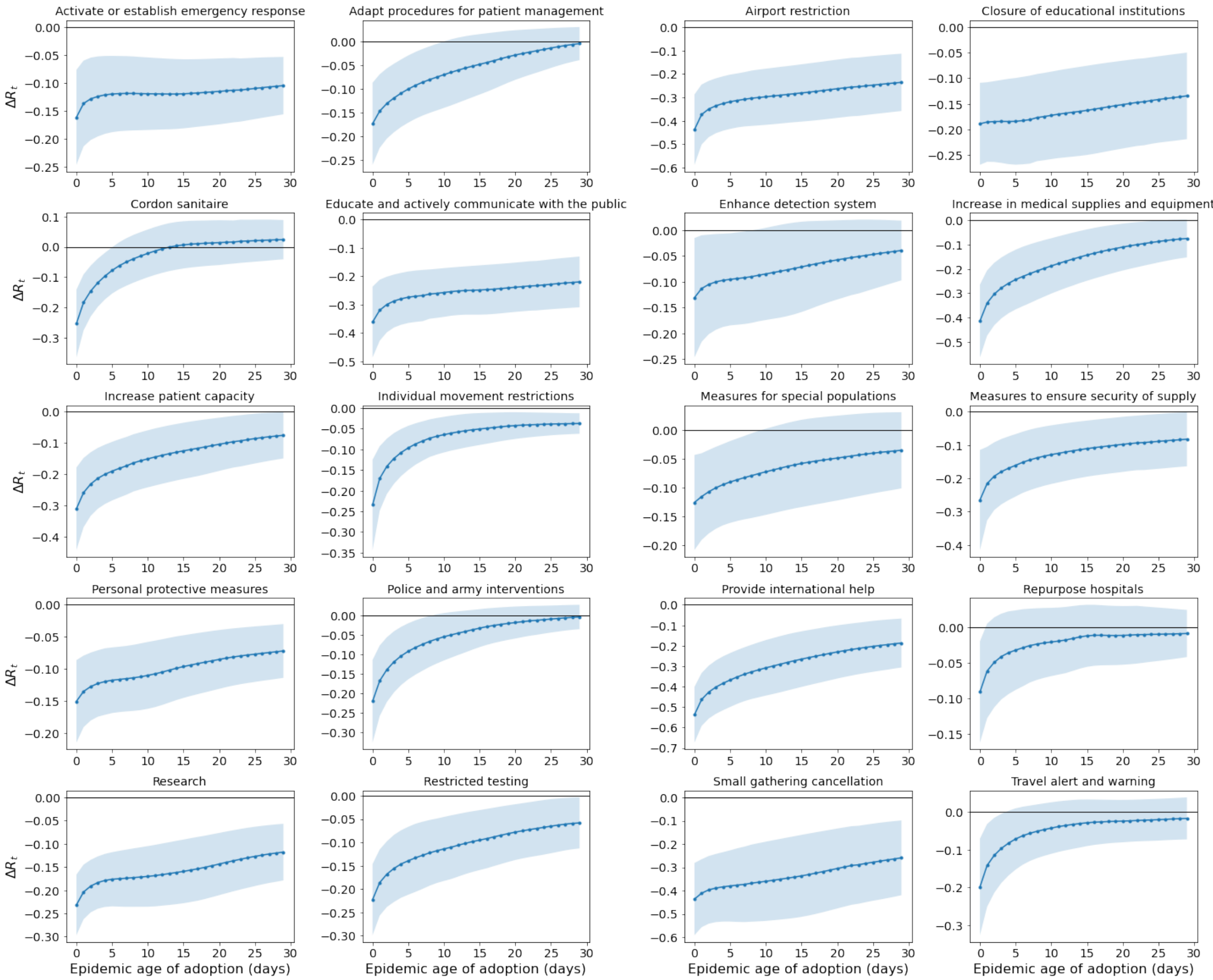
Supplementary Figure 3: Heatmap of the predictive performance of the random forest depending on the maximum tree depth d and the percentage of features m considered at each split for time shift $\tau = 10$. The predictive performance is measured in terms of the coefficient of determination (r^2) on unseen data, averaged over 10 random splits of the territories into training and test sets.



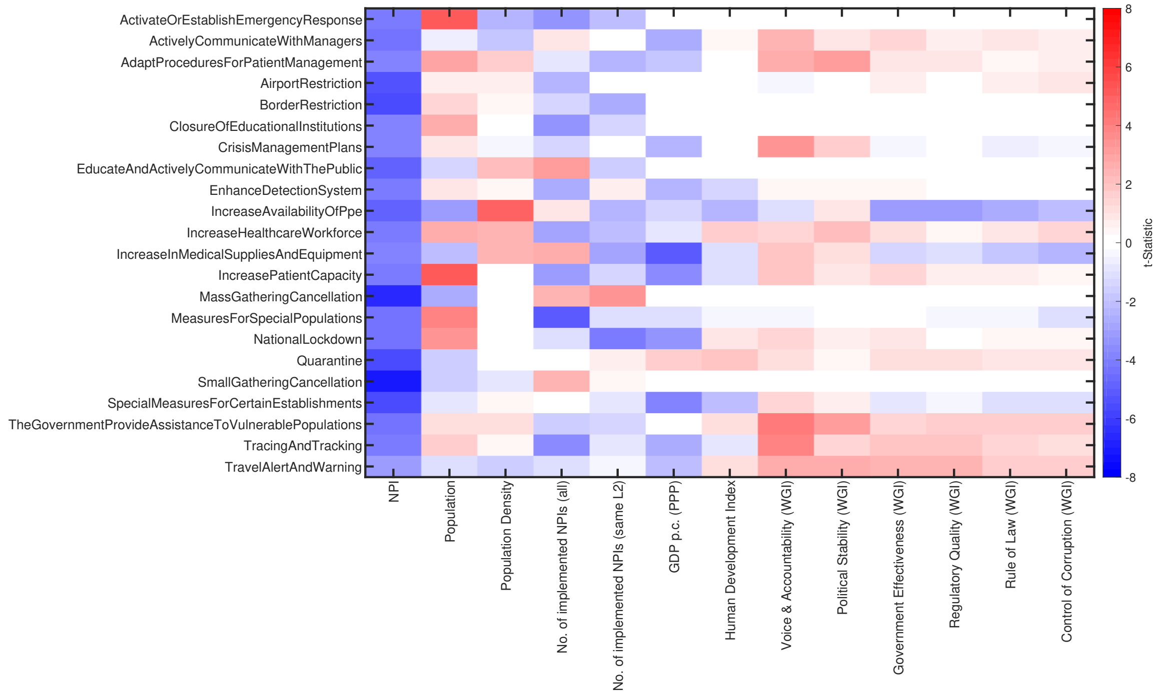
Supplementary Figure 4: Example of the country-specific impact of an NPI. The plot shows the R_t increase when the NPI "Small gathering cancellation" is removed from the temporal measures sequence in Italy. Shaded areas are the standard deviations of the Transformer prediction averaged on the networks ensemble.



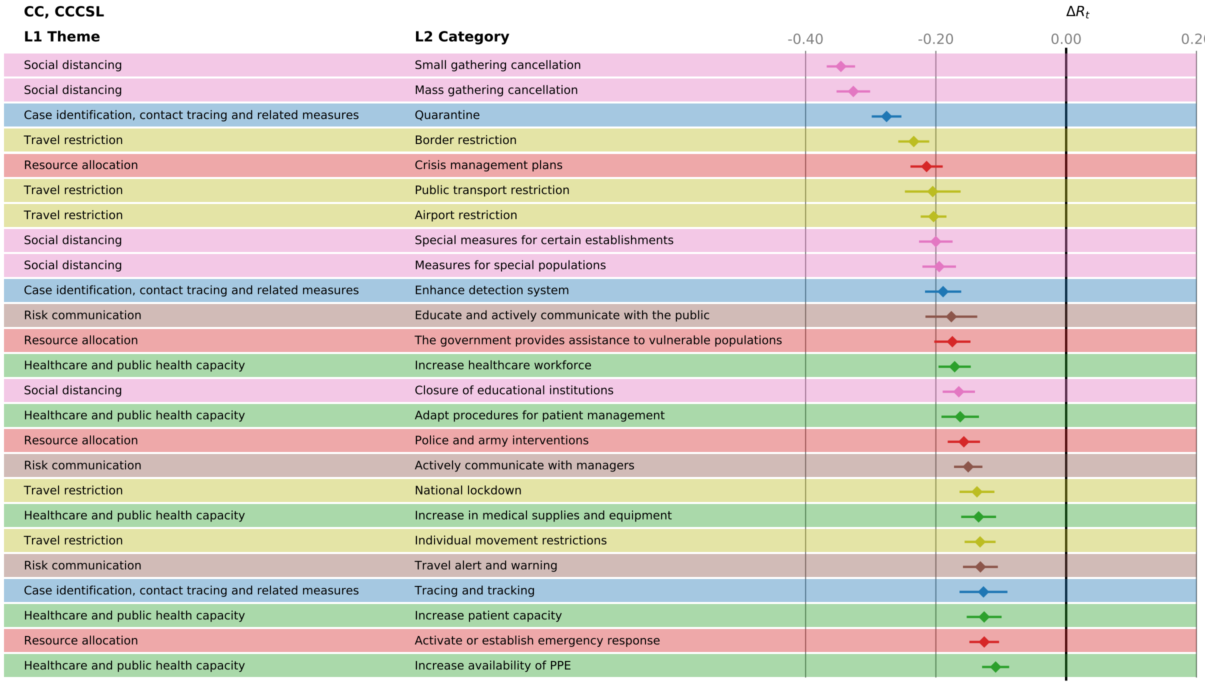
Supplementary Figure 5: Example illustrating the what-if experiment described in the text. Here we consider the NPI "Small gathering cancellation" in Italy and we simulate what would happen if it had been taken at epidemic ages 0, 10, 20, 30, 40, 50, 60. We then compare this evolution with an evolution obtained through a knowknout of the same NPI (knockout evolution). The other concurrent NPIs and their days of adoption are kept fixed throughout the simulation (green rectangles at the bottom).



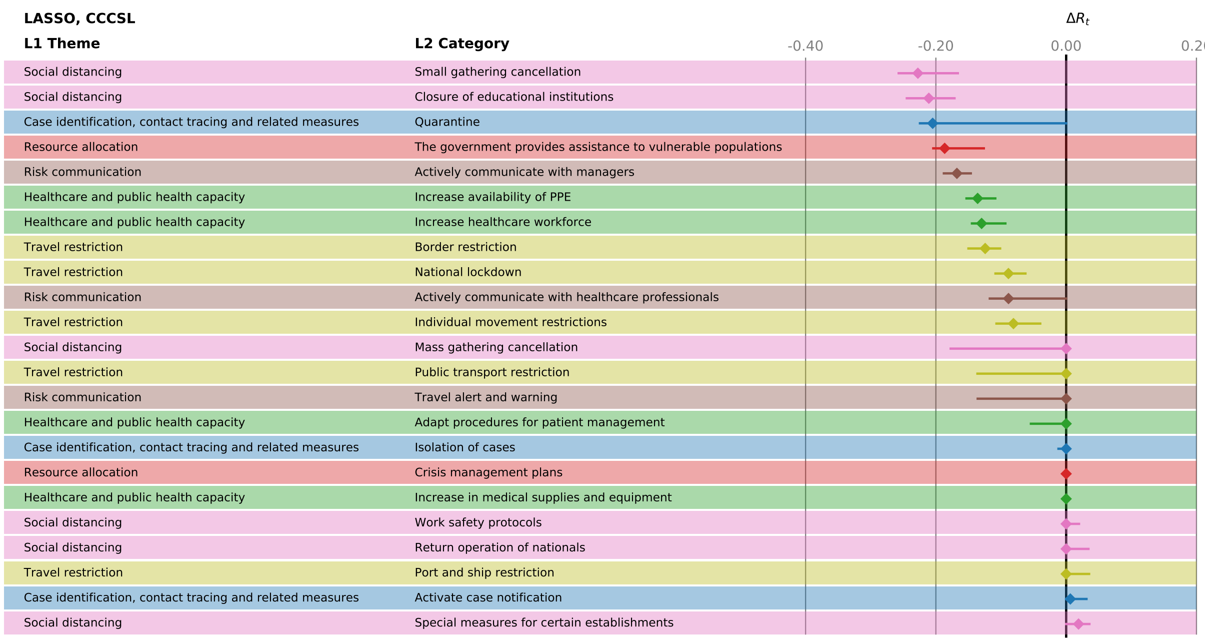
Supplementary Figure 6: Outcome of the what-if experiment performed with the Transformers. We report, for several NPIs, the behaviour of the averaged variation of R_t, ΔR_t (see text for the definition), when the NPI is adopted at a generic epidemic age compared with an evolution in which that NPI has been knocked-out. Negative (Positive) values of ΔR_t indicate a decrease (increase) of R_t.



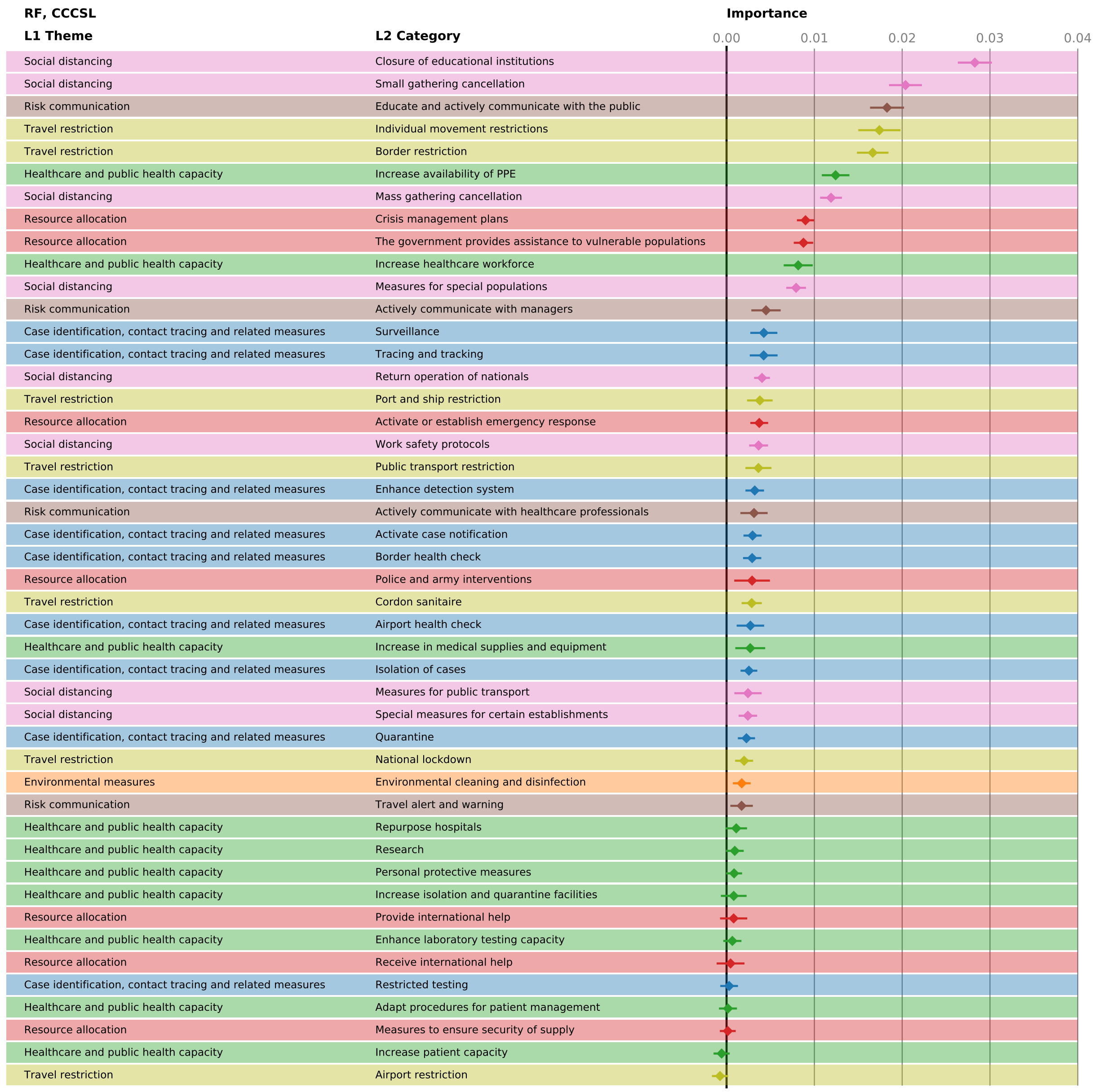
Supplementary Figure 7: Impacts on $R(t)$ of country variables on measure effectiveness. The heatmap gives the average effect size (t-statistic) for a given NPI category (L2) (rows) and a country variable (columns). Blue (red) color indicates that the variable is positively (negatively) correlated with measure effectiveness.



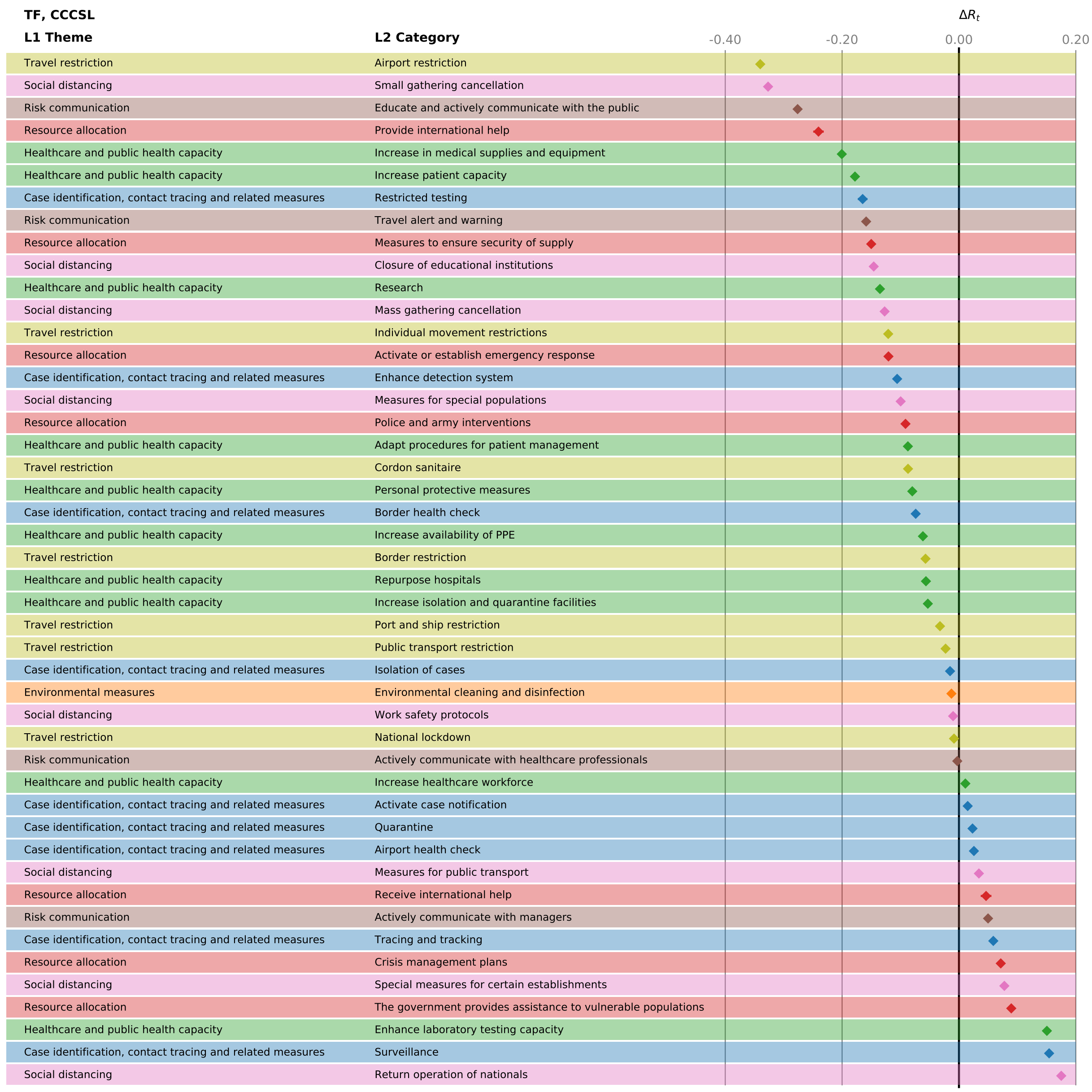
Supplementary Figure 8: Effectiveness, ΔR_t of the different NPIs in the CC analysis for CCCSL. The horizontal bars mark the 95% confidence intervals.



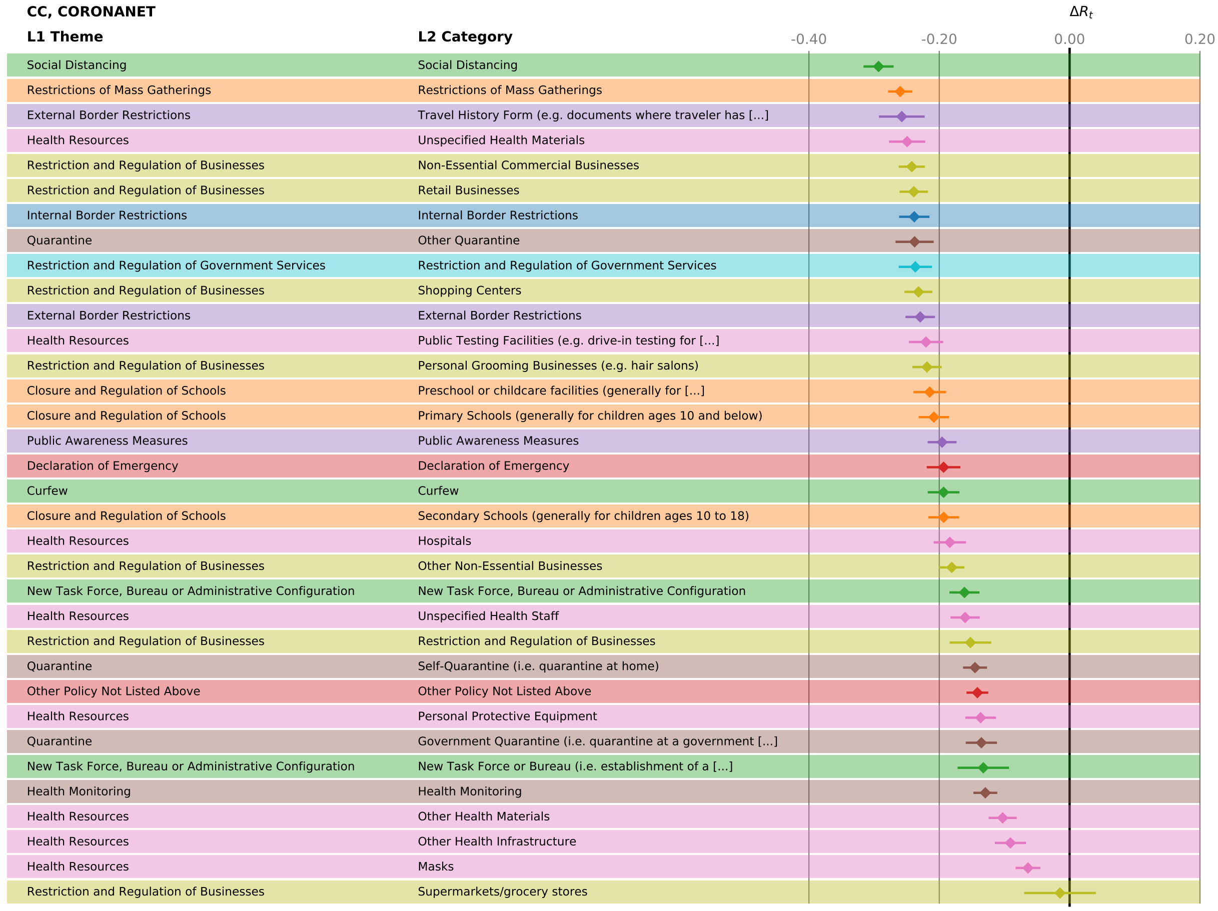
Supplementary Figure 9: Effectiveness, ΔR_t of the different NPIs in the LASSO analysis for CCCSL. The horizontal bars mark the 95% confidence intervals.



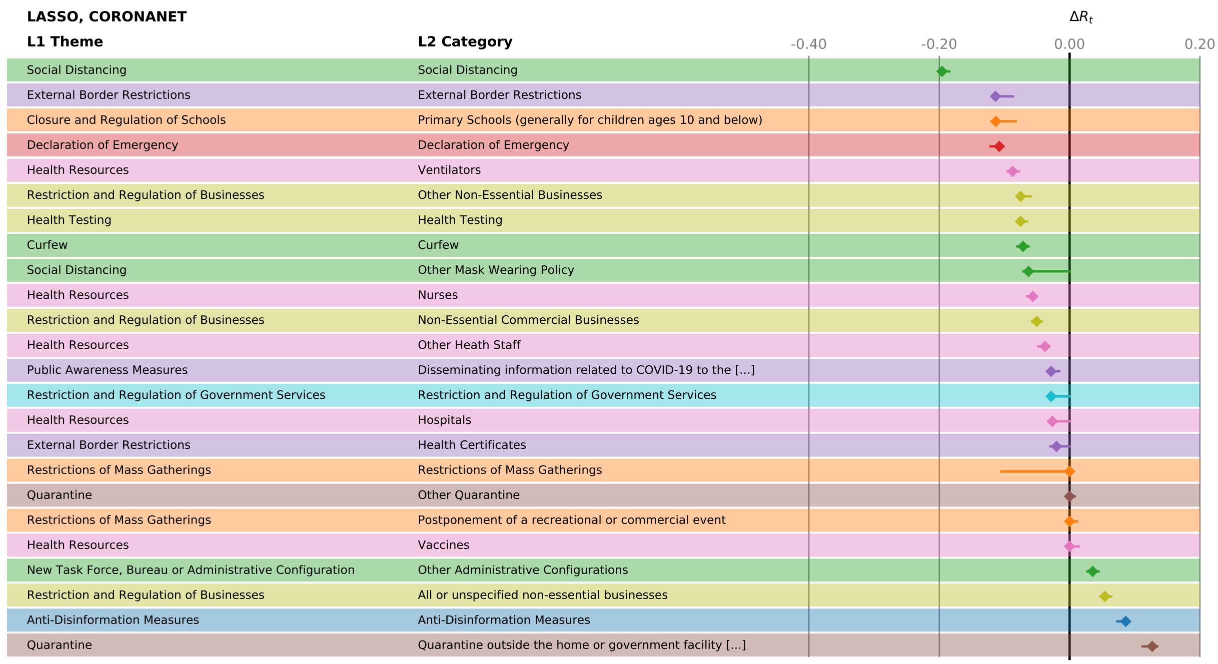
Supplementary Figure 10: Feature importance of the different NPIs in the random forest model for CCCSL. The horizontal bars mark the 95% confidence intervals.



Supplementary Figure 11: Effectiveness, ΔR_t of the different NPIs in the Transformer analysis for CCCSL. The bars marking the 95% confidence intervals are too small to be visible.

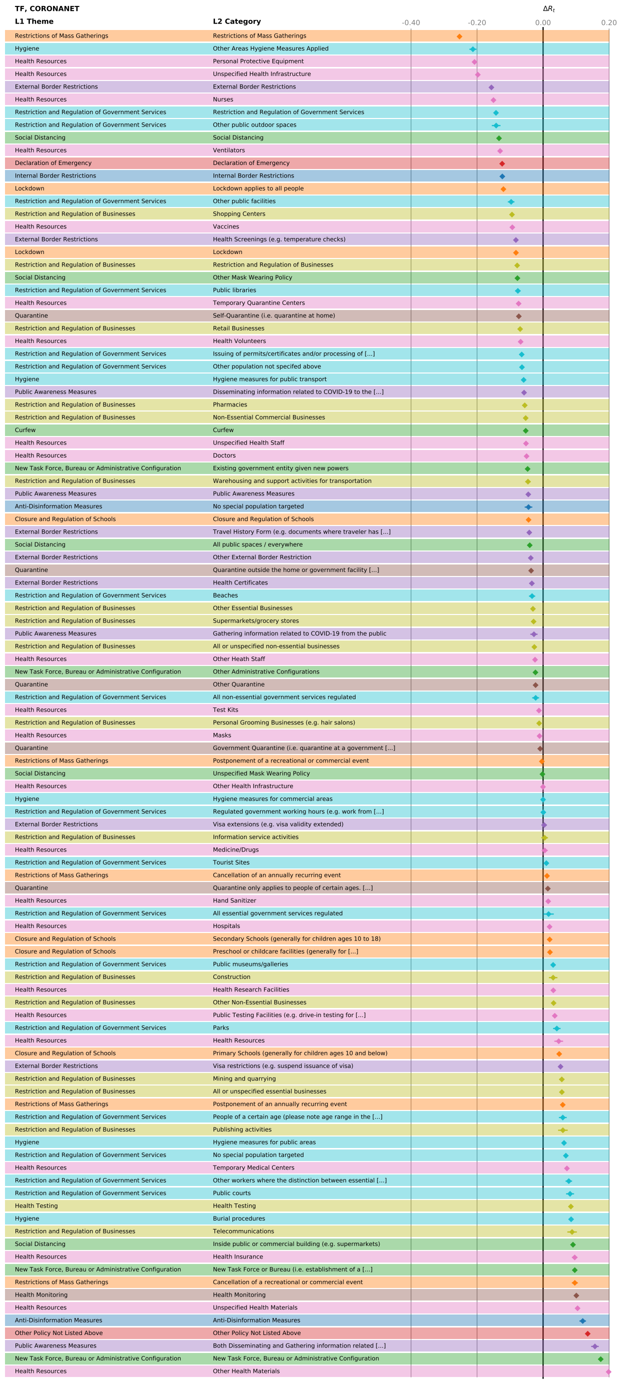


Supplementary Figure 12: Effectiveness, ΔR_t of the different NPIs in the CC analysis for CORONANET. The horizontal bars mark the 95% confidence intervals.

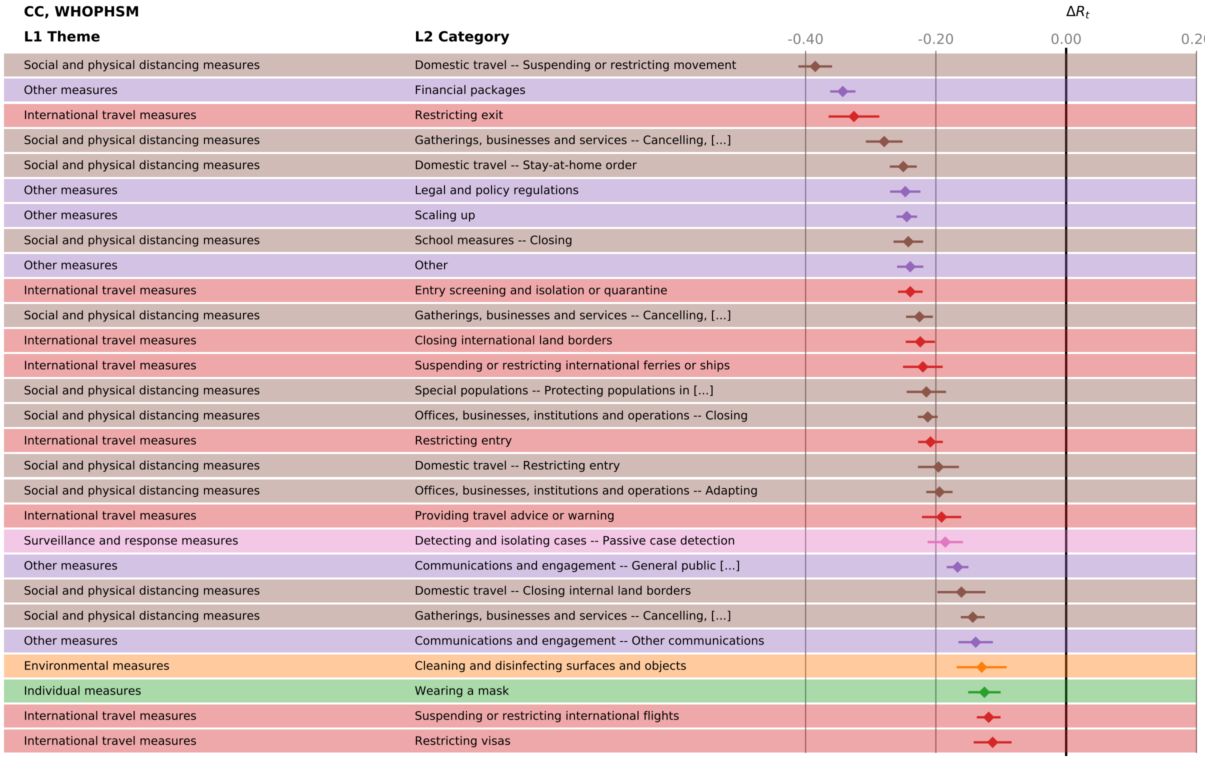


Supplementary Figure 13: Effectiveness, ΔR_t of the different NPIs in the LASSO analysis for CORONANET. The horizontal bars mark the 95% confidence intervals.

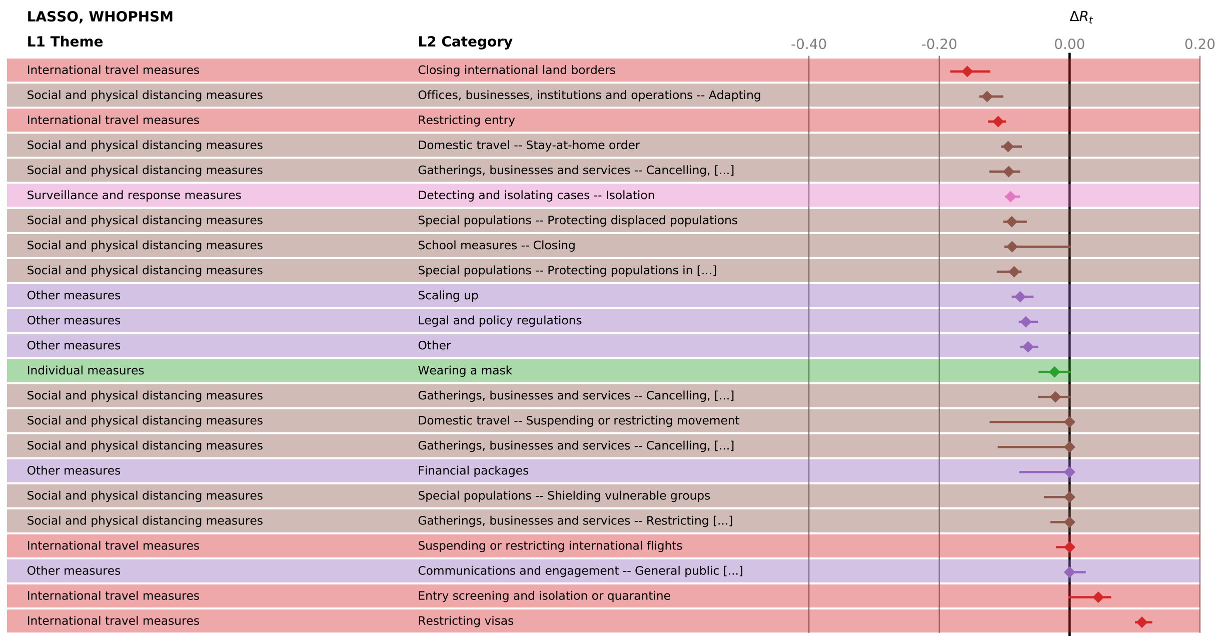
Supplementary Figure 14: Feature importance of the different NPIs in the random forest model for CORONANET. The horizontal bars mark the 95% confidence intervals.



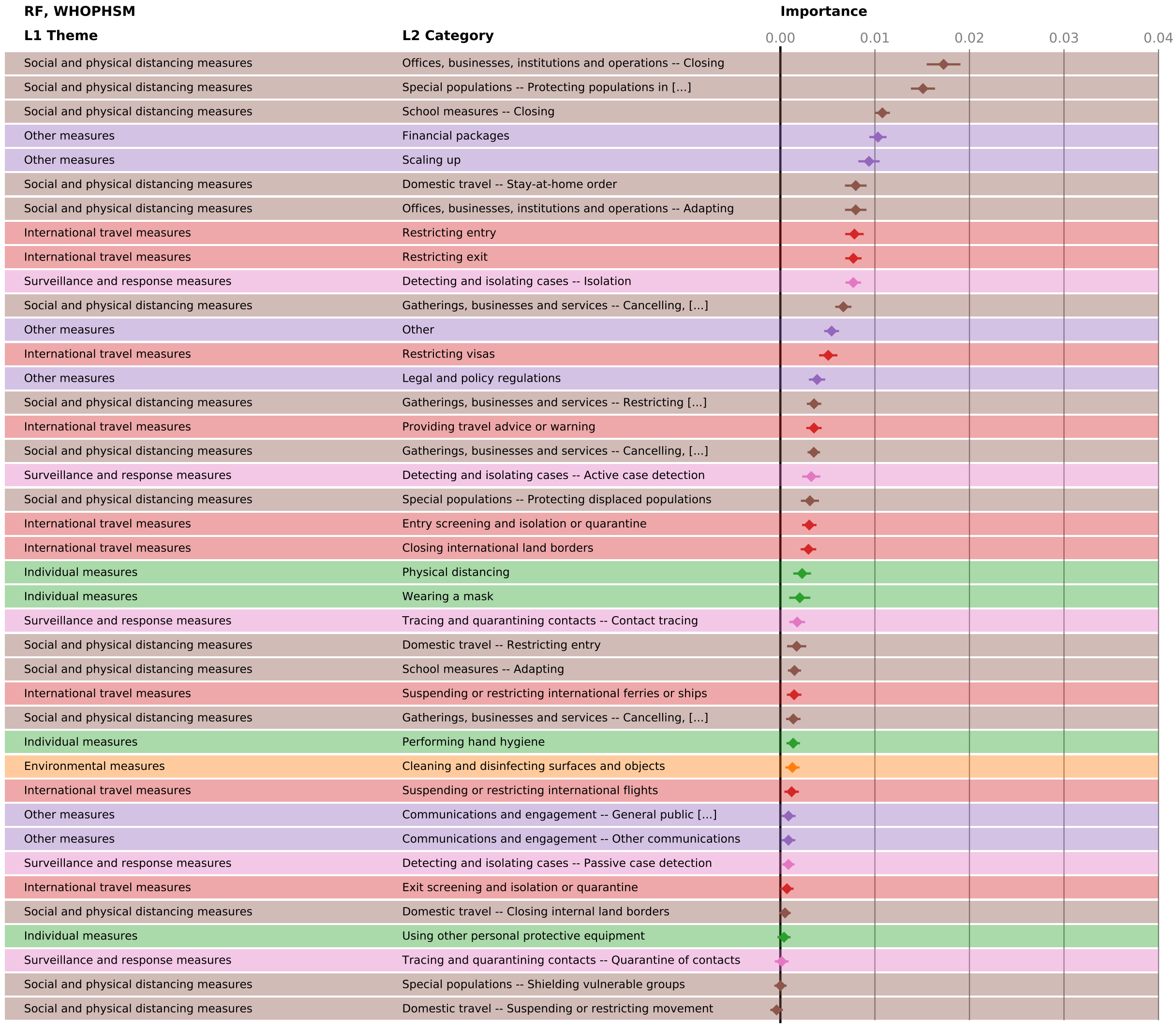
Supplementary Figure 15: Effectiveness, ΔR_t of the different NPIs in the Transformer analysis for CORONANET. The bars marking the 95% confidence intervals are too small to be visible.



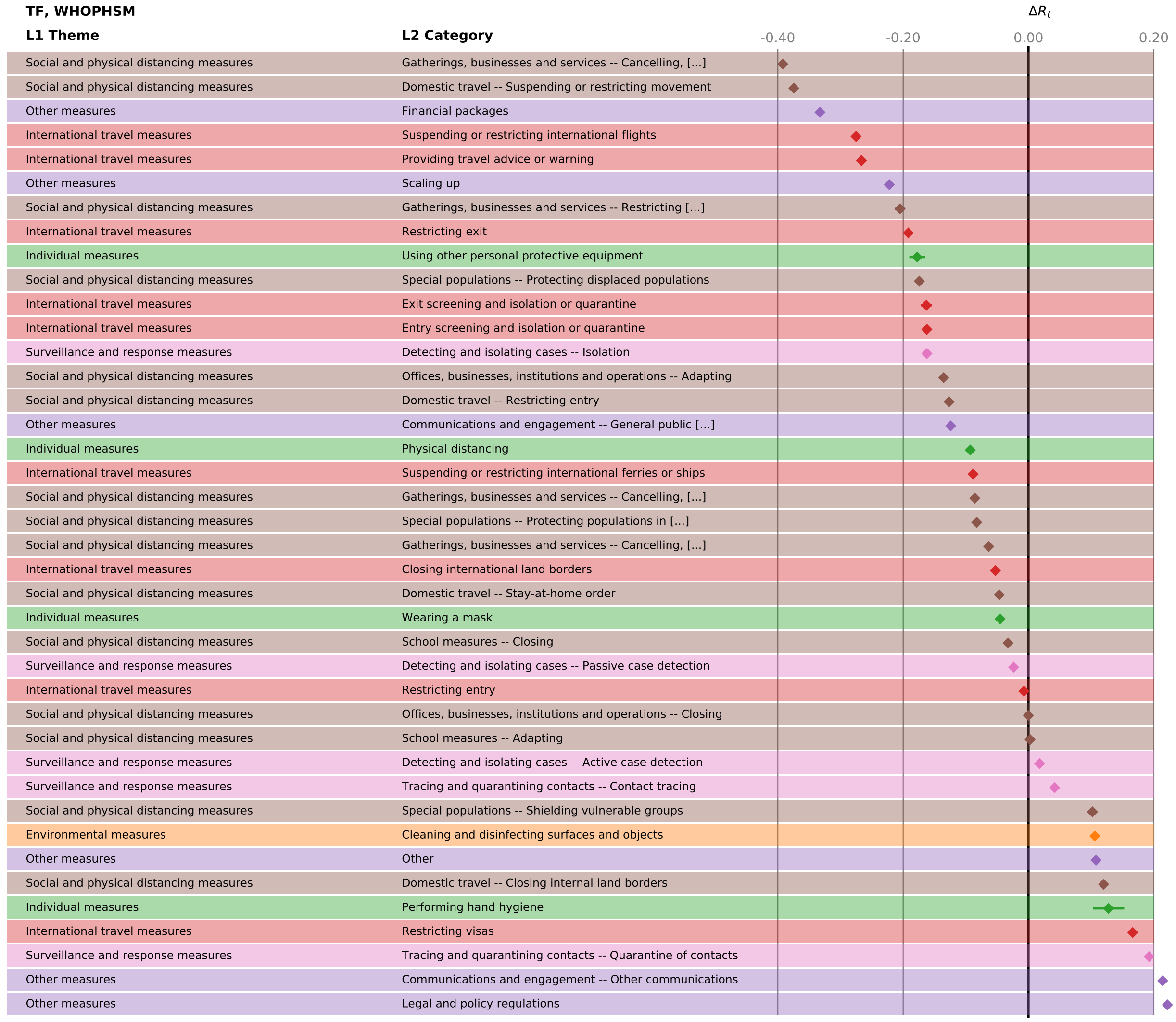
Supplementary Figure 16: Effectiveness, ΔR_t of the different NPIs in the CC analysis for WHOPHSM. The horizontal bars mark the 95% confidence intervals.



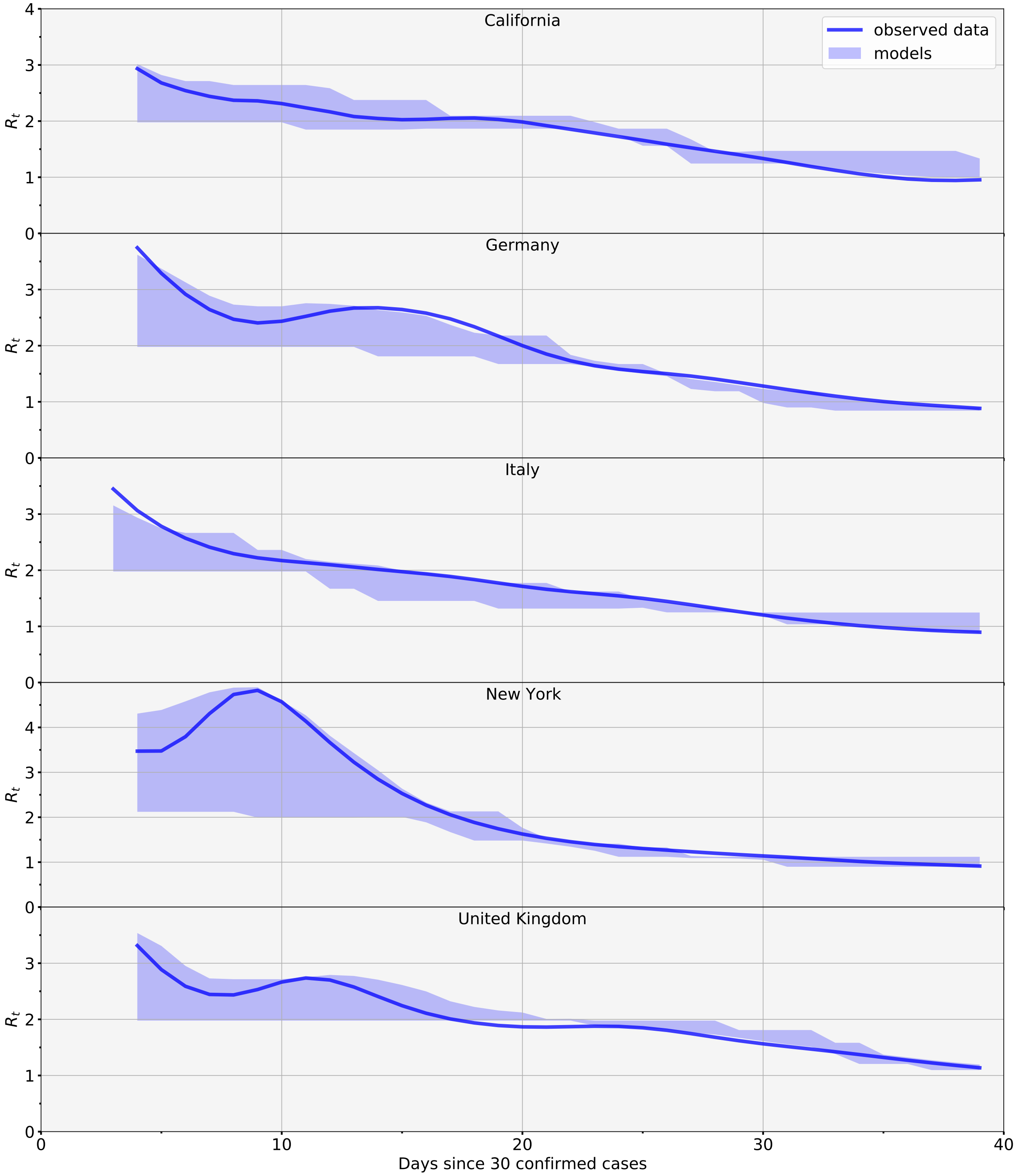
Supplementary Figure 17: Effectiveness, ΔR_t of the different NPIs in the LASSO analysis for WHOPHSM. The horizontal bars mark the 95% confidence intervals.



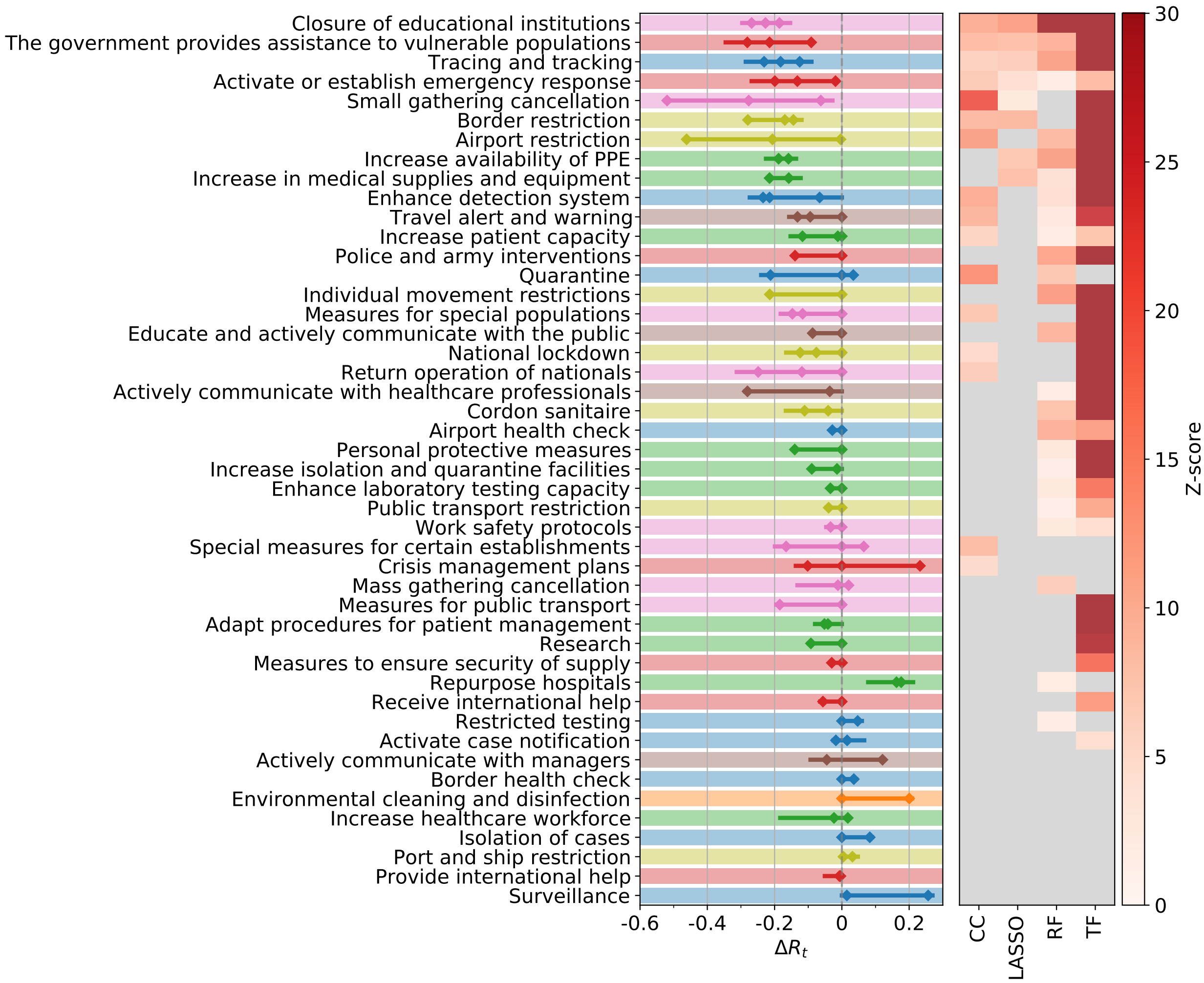
Supplementary Figure 18: Feature importance of the different NPIs in the random forest model for WHOPHSM. The horizontal bars mark the 95% confidence intervals.



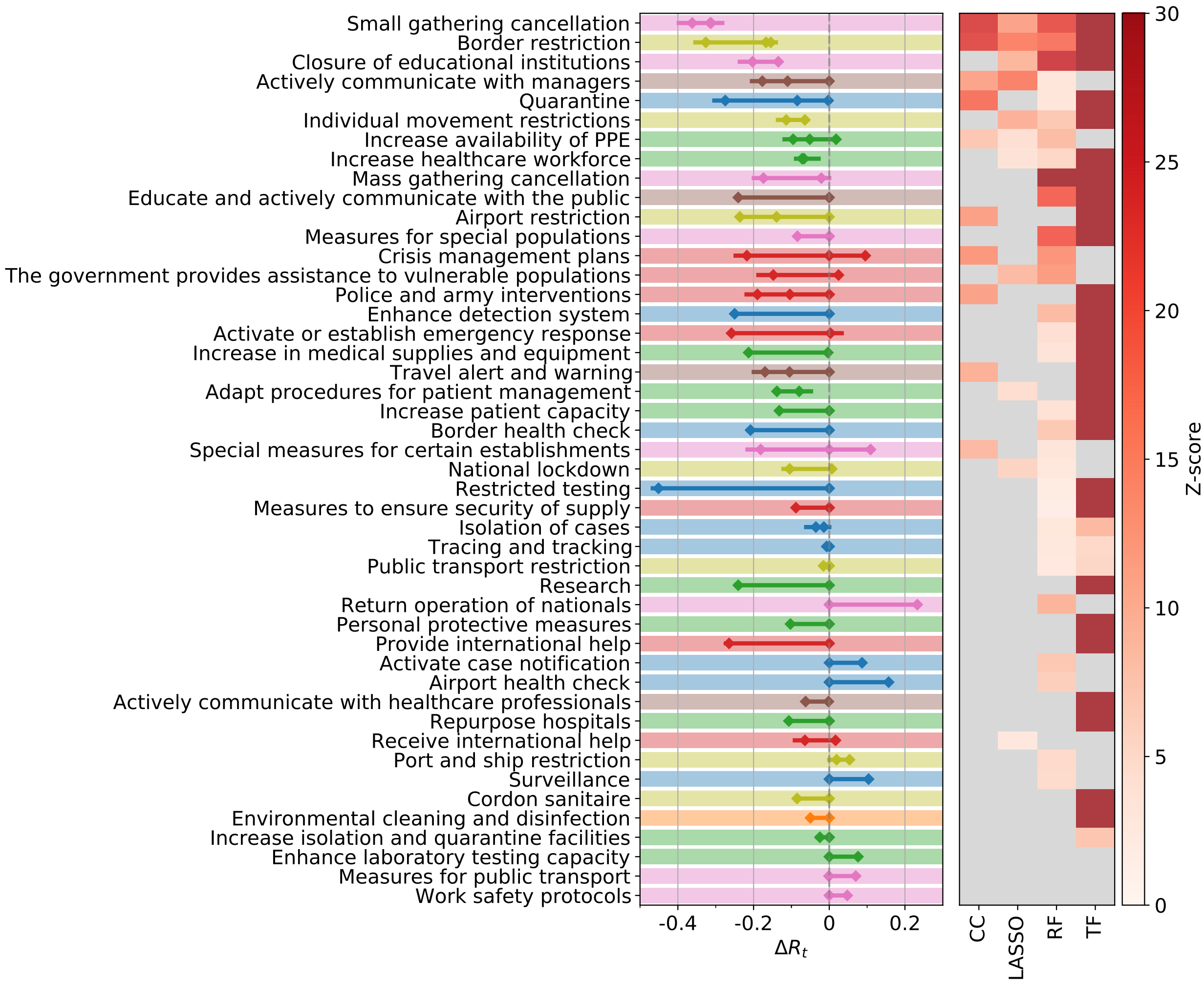
Supplementary Figure 19: Effectiveness, ΔR_t of the different NPIs in the Transformer analysis for WHOPHSM. The bars marking the 95% confidence intervals are too small to be visible.



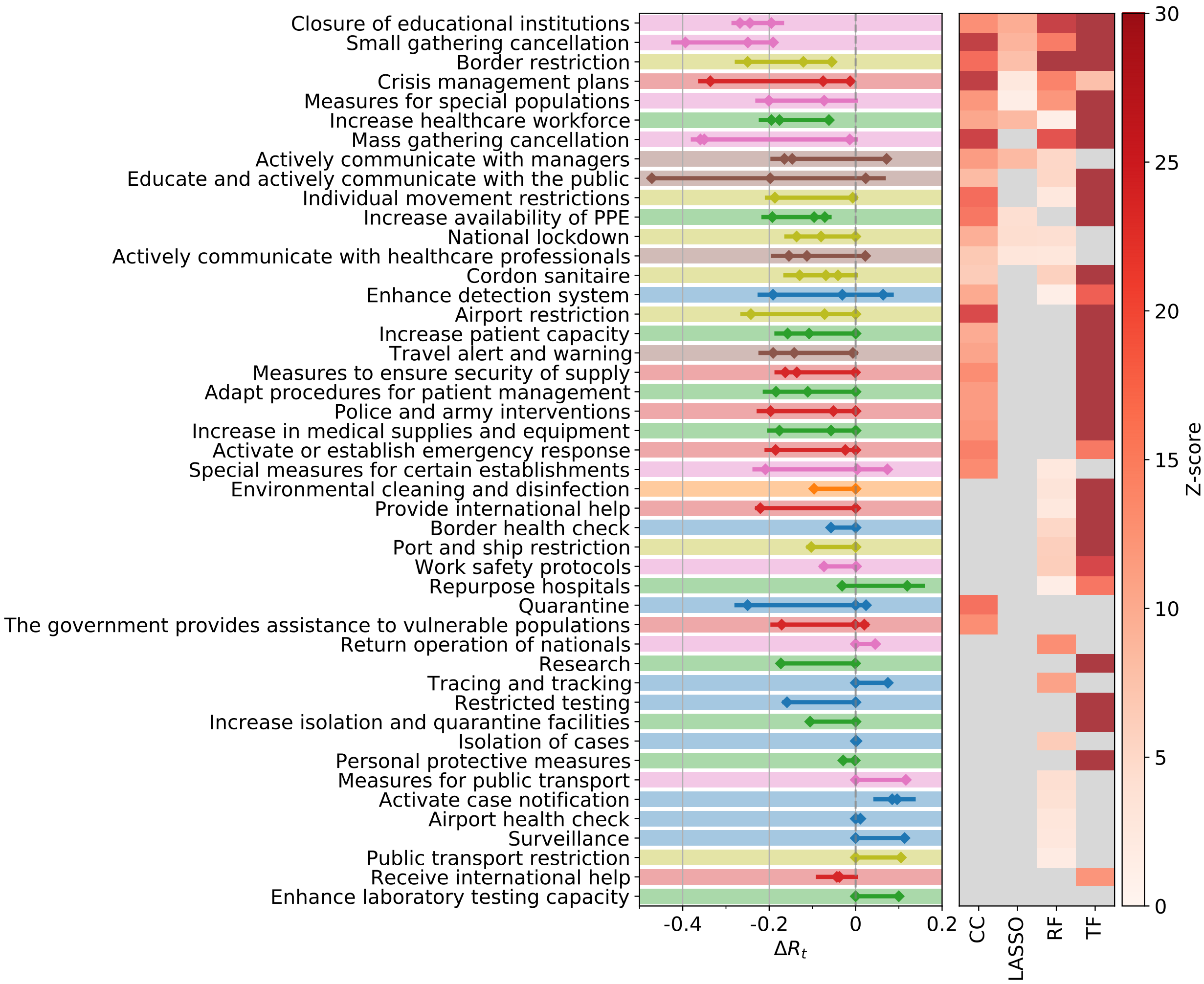
Supplementary Figure 20: Observed values for R_t (solid lines), together with the range of values for R_t as predicted from the different methods (shaded regions).



Supplementary Figure 21: Analogue to Fig. 1 of the main text if the analysis is done on all countries except those in Europe and Africa.



Supplementary Figure 22: Analogue to Fig. 1 of the main text if the analysis is done on all countries except those in Asia and Oceania.



Supplementary Figure 23: Analogue to Fig. 1 of the main text if the analysis is done on all countries except those in North, Central or South America.

5 Supplementary Tables

L2 category	L3 subcategories
Small gathering cancellation	Complete prohibition of gathering; Closure of restaurants/bars/cafes; Closure of non-essential shops; Limit up to 5 persons; Implement part-time work; Mandatory home office; Limit up to 10 persons; Restriction on private and familial events; Limit up to 30 persons; Limit up to 2 persons; Remote Psychotherapy Consultation; Limit up to 6 persons; Limit up to 20 persons; Closure of short-term accommodation; Limit up to 1 persons; Closure of student dormitories; Reduce close physical contact in workplaces; Mandatory 2m distance in public spaces; Limit up to 3 persons; Limit up to 50 persons; Limit up to 4 persons; Limit up to 25 persons; Non-critical court operations suspended
Closure of educational institutions	Complete closure of kindergartens; Complete closure of primary and secondary schools; Complete closure of universities; Reduction of excursions, out-of-house events; Complete closure of all educational institutions; Cancellation of exams; Partial closure of primary and secondary schools; Partial closure of universities; Complete closure of secondary schools; Extracurriculars cancelled; Partial closure of educational institutions; Rules for exams; Partial closure of kindergartens; Restrictions on exams; Closure of adult educational schools; Complete closure of scientific institutions; Complete closure of research institutes

L2 category	L3 subcategories
Border restriction	Land borders closed; Land borders closed; Entry ban to people from high-risk areas other than China; Land border controls; Entry ban to non-citizens; Conditional entry of persons from neighboring countries; Conditional entry of citizens; Entry ban to non EU citizens; Temporary reduction of service; Entry ban to people from China; Entry ban to refugees; Entry ban for symptomatic people and case contacts; Travel ban to high-risk areas; Total entry ban; Close land border to prevent virus spread; Entry ban to infected persons; Entry ban to people with a travel history to China; Ban on passenger transport from China; Force departure of Chinese nationals; Border control; Suspension of passenger railway transports crossing the country border; Restriction of freight transport; Ban on road passenger transport from high risk areas
Increase availability of PPE	PPE for healthcare professionals; Face masks; PPE (not specified); PPE other than face masks; Prohibition of export of protective personal equipment; Hand sanitizers; Increase domestic production of PPE
Individual movement restrictions	Movements for non-essential activities forbidden; Curfew; Non-essential travels abroad/out-of-state forbidden; Segmentation of the population; Partial restriction on movements; Prohibition of moving out the municipality of residence; Restrictions on the movements of children
National lockdown	for 21 days; For 2 weeks; Stay-at-home Order; Safer-at-home Order

Supplementary Table 1: Subcategories (L3) belonging to the eight consensus categories (L2) identified as significant by all four methods.

L3 category	L2 category	ΔR_t (SE)	p-value
Warning against travel to and return from high risk areas	**Travel alert and warning**	$\mathbf{-0.14(1)}$	$\mathbf{< 10^{-7}}$
Encourage stay at home	**Educate and actively communicate with the public**	$\mathbf{-0.14(1)}$	$\mathbf{< 10^{-4}}$
Promote social distancing measures	**Educate and actively communicate with the public**	$\mathbf{-0.20(2)}$	$\mathbf{< 10^{-4}}$
Promote workplace safety measures	**Educate and actively communicate with the public**	$\mathbf{-0.19(2)}$	$\mathbf{< 10^{-4}}$
Promote self-initiated isolation of people with mild respiratory symptoms	**Educate and actively communicate with the public**	$\mathbf{-0.19(2)}$	$\mathbf{0.0001}$
Information campaign	**Educate and actively communicate with the public**	$\mathbf{-0.13(2)}$	$\mathbf{0.0003}$
Respiratory etiquette	Educate and actively communicate with the public	$-0.10(1)$	0.0005
Answer to questions	Educate and actively communicate with the public	$-0.077(1)$	0.0007
Call for return of nationals living abroad	Educate and actively communicate with the public	$-0.25(2)$	0.0009
Recommendations for work safety protocols	Actively communicate with managers	$-0.13(2)$	0.0013
Encourage self-initiated quarantine	Educate and actively communicate with the public	$-0.19(2)$	0.005
Communication targets protection of vulnerable populations	Educate and actively communicate with the public	$-0.14(2)$	0.007
Guidelines	Actively communicate with managers	$-0.14(2)$	0.02
Information about travels	Educate and actively communicate with the public	$-0.17(3)$	0.02
Direct government communication	Educate and actively communicate with the public	$-0.072(2)$	0.02
Encourage hand hygiene	Educate and actively communicate with the public	$-0.084(2)$	0.03
Direct advice to vulnerable populations	Educate and actively communicate with the public	$-0.14(3)$	0.04
Foster community assistance	Educate and actively communicate with the public	$-0.14(2)$	0.04

Supplementary Table 2: Results of the CC analysis for risk communication NPIs on level L3. For each measure we give the change in R_t with the SE in brackets and its p-value. Measures with significant effects after a multiple testing correction are highlighted in bold.

CoronaNet	
L1 Theme	L2 Category
Anti-Disinformation Measures	Anti-Disinformation Measures
Anti-Disinformation Measures	No special population targeted
Closure and Regulation of Schools	Closure and Regulation of Schools
Closure and Regulation of Schools	Preschool or childcare facilities (generally for children ages 5 and below)
Closure and Regulation of Schools	Primary Schools (generally for children ages 10 and below)
Closure and Regulation of Schools	Secondary Schools (generally for children ages 10 to 18)
Curfew	Curfew
Declaration of Emergency	Declaration of Emergency
External Border Restrictions	External Border Restrictions
External Border Restrictions	Health Certificates
External Border Restrictions	Health Screenings (e.g. temperature checks)
External Border Restrictions	Other External Border Restriction
External Border Restrictions	Travel History Form (e.g. documents where traveler has recently been)
External Border Restrictions	Visa extensions (e.g. visa validity extended)
External Border Restrictions	Visa restrictions (e.g. suspend issuance of visa)
Health Monitoring	Health Monitoring
Health Resources	Doctors
Health Resources	Hand Sanitizer
Health Resources	Health Insurance
Health Resources	Health Research Facilities

CoronaNet	
L1 Theme	L2 Category
Health Resources	Health Resources
Health Resources	Health Volunteers
Health Resources	Hospitals
Health Resources	Masks
Health Resources	Medicine/Drugs
Health Resources	Nurses
Health Resources	Other Health Infrastructure
Health Resources	Other Health Materials
Health Resources	Other Heath Staff
Health Resources	Personal Protective Equipment
Health Resources	Public Testing Facilities (e.g. drive-in testing for COVID-19)
Health Resources	Temporary Medical Centers
Health Resources	Temporary Quarantine Centers
Health Resources	Test Kits
Health Resources	Unspecified Health Infrastructure
Health Resources	Unspecified Health Materials
Health Resources	Unspecified Health Staff
Health Resources	Vaccines
Health Resources	Ventilators
Health Testing	Health Testing
Hygiene	Burial procedures
Hygiene	Hygiene measures for commercial areas

CoronaNet L1 Theme	L2 Category
Hygiene	Hygiene measures for public areas
Hygiene	Hygiene measures for public transport
Hygiene	Other Areas Hygiene Measures Applied
Internal Border Restrictions	Internal Border Restrictions
Lockdown	Lockdown
Lockdown	Lockdown applies to all people
New Task Force, Bureau or Administrative Configuration	Existing government entity given new powers
New Task Force, Bureau or Administrative Configuration	New Task Force or Bureau (i.e. establishment of a temporary body)
New Task Force, Bureau or Administrative Configuration	New Task Force, Bureau or Administrative Configuration
New Task Force, Bureau or Administrative Configuration	Other Administrative Configurations
Other Policy Not Listed Above	Other Policy Not Listed Above
Public Awareness Measures	Both Disseminating and Gathering information related to COVID-19
Public Awareness Measures	Disseminating information related to COVID-19 to the public that is reliable and factually accurate
Public Awareness Measures	Gathering information related to COVID-19 from the public
Public Awareness Measures	Public Awareness Measures

CoronaNet L1 Theme	L2 Category
Quarantine	Government Quarantine (i.e. quarantine at a government hotel or facility)
Quarantine	Other Quarantine
Quarantine	Quarantine only applies to people of certain ages. Please note the age restrictions in the text box.
Quarantine	Quarantine outside the home or government facility (i.e. quarantine in a hotel)
Quarantine	Self-Quarantine (i.e. quarantine at home)
Restriction and Regulation of Businesses	All or unspecified essential businesses
Restriction and Regulation of Businesses	All or unspecified non-essential businesses
Restriction and Regulation of Businesses	Construction
Restriction and Regulation of Businesses	Information service activities
Restriction and Regulation of Businesses	Mining and quarrying
Restriction and Regulation of Businesses	Non-Essential Commercial Businesses
Restriction and Regulation of Businesses	Other Essential Businesses
Restriction and Regulation of Businesses	Other Non-Essential Businesses
Restriction and Regulation of Businesses	Personal Grooming Businesses (e.g. hair salons)
Restriction and Regulation of Businesses	Pharmacies
Restriction and Regulation of Businesses	Publishing activities
Restriction and Regulation of Businesses	Restriction and Regulation of Businesses
Restriction and Regulation of Businesses	Retail Businesses
Restriction and Regulation of Businesses	Shopping Centers
Restriction and Regulation of Businesses	Supermarkets/grocery stores

CoronaNet L1 Theme	L2 Category
Restriction and Regulation of Businesses	Telecommunications
Restriction and Regulation of Businesses	Warehousing and support activities for transportation
Restriction and Regulation of Government Services	All essential government services regulated
Restriction and Regulation of Government Services	All non-essential government services regulated
Restriction and Regulation of Government Services	Beaches
Restriction and Regulation of Government Services	Issuing of permits/certificates and/or processing of government documents
Restriction and Regulation of Government Services	No special population targeted
Restriction and Regulation of Government Services	Other population not specifed above
Restriction and Regulation of Government Services	Other public facilities
Restriction and Regulation of Government Services	Other public outdoor spaces
Restriction and Regulation of Government Services	Other workers where the distinction between essential or non-essential is not explictely made (please note the occupation in the text entry)
Restriction and Regulation of Government Services	Parks

CoronaNet L1 Theme	L2 Category
Restriction and Regulation of Government Services	People of a certain age (please note age range in the text entry)
Restriction and Regulation of Government Services	Public courts
Restriction and Regulation of Government Services	Public libraries
Restriction and Regulation of Government Services	Public museums/galleries
Restriction and Regulation of Government Services	Regulated government working hours (e.g. work from home policies for government workers)
Restriction and Regulation of Government Services	Restriction and Regulation of Government Services
Restriction and Regulation of Government Services	Tourist Sites
Restrictions of Mass Gatherings	Cancellation of a recreational or commercial event
Restrictions of Mass Gatherings	Cancellation of an annually recurring event
Restrictions of Mass Gatherings	Postponement of a recreational or commercial event
Restrictions of Mass Gatherings	Postponement of an annually recurring event
Restrictions of Mass Gatherings	Prison population reduced (e.g. early release of prisoners)
Restrictions of Mass Gatherings	Restrictions of Mass Gatherings
Social Distancing	All public spaces / everywhere
Social Distancing	Inside public or commercial building (e.g. supermarkets)
Social Distancing	Other Mask Wearing Policy

CoronaNet	
L1 Theme	L2 Category
Social Distancing	Social Distancing
Social Distancing	Unspecified Mask Wearing Policy

Supplementary Table 3: Full table of NPI Themes and Categories in CoronaNet without abbreviations.

WHOPHSM	
L1 Theme	L2 Category
Environmental measures	Cleaning and disinfecting surfaces and objects
Individual measures	Performing hand hygiene
Individual measures	Physical distancing
Individual measures	Using other personal protective equipment
Individual measures	Wearing a mask
International travel measures	Closing international land borders
International travel measures	Entry screening and isolation or quarantine
International travel measures	Exit screening and isolation or quarantine
International travel measures	Providing travel advice or warning
International travel measures	Restricting entry
International travel measures	Restricting exit
International travel measures	Restricting visas
International travel measures	Suspending or restricting international ferries or ships
International travel measures	Suspending or restricting international flights

WHOPHSM	
L1 Theme	L2 Category
Other measures	Communications and engagement – General public awareness campaigns
Other measures	Communications and engagement – Other communications
Other measures	Financial packages
Other measures	Legal and policy regulations
Other measures	Other
Other measures	Scaling up
Social and physical distancing measures	Domestic travel – Closing internal land borders
Social and physical distancing measures	Domestic travel – Restricting entry
Social and physical distancing measures	Domestic travel – Stay-at-home order
Social and physical distancing measures	Domestic travel – Suspending or restricting movement
Social and physical distancing measures	Gatherings, businesses and services – Cancelling, closing, restricting or adapting public gatherings outside the home
Social and physical distancing measures	Gatherings, businesses and services – Cancelling, restricting or adapting mass gatherings
Social and physical distancing measures	Gatherings, businesses and services – Cancelling, restricting or adapting private gatherings outside the home
Social and physical distancing measures	Gatherings, businesses and services – Restricting private gatherings at home
Social and physical distancing measures	Offices, businesses, institutions and operations – Adapting
Social and physical distancing measures	Offices, businesses, institutions and operations – Closing
Social and physical distancing measures	School measures – Adapting

WHOPHSM	
L1 Theme	L2 Category
Social and physical distancing measures	School measures – Closing
Social and physical distancing measures	Special populations – Protecting displaced populations
Social and physical distancing measures	Special populations – Protecting populations in closed settings
Social and physical distancing measures	Special populations – Shielding vulnerable groups
Surveillance and response measures	Detecting and isolating cases – Active case detection
Surveillance and response measures	Detecting and isolating cases – Isolation
Surveillance and response measures	Detecting and isolating cases – Passive case detection
Surveillance and response measures	Tracing and quarantining contacts – Contact tracing
Surveillance and response measures	Tracing and quarantining contacts – Quarantine of contacts

Supplementary Table 4: Full table of NPI Themes and Categories in WHOPHSM without abbreviations.

CCCSL	
L1 Theme	L2 Category
Case identification, contact tracing and related measures	Activate case notification
Case identification, contact tracing and related measures	Airport health check
Case identification, contact tracing and related measures	Border health check

CCCSL	
L1 Theme	L2 Category
Case identification, contact tracing and related measures	Enhance detection system
Case identification, contact tracing and related measures	Isolation of cases
Case identification, contact tracing and related measures	Quarantine
Case identification, contact tracing and related measures	Restricted testing
Case identification, contact tracing and related measures	Surveillance
Case identification, contact tracing and related measures	Tracing and tracking
Environmental measures	Environmental cleaning and disinfection
Healthcare and public health capacity	Adapt procedures for patient management
Healthcare and public health capacity	Enhance laboratory testing capacity
Healthcare and public health capacity	Increase availability of PPE
Healthcare and public health capacity	Increase healthcare workforce
Healthcare and public health capacity	Increase in medical supplies and equipment
Healthcare and public health capacity	Increase isolation and quarantine facilities
Healthcare and public health capacity	Increase patient capacity
Healthcare and public health capacity	Personal protective measures
Healthcare and public health capacity	Repurpose hospitals
Healthcare and public health capacity	Research

CCCSL	
L1 Theme	L2 Category
Resource allocation	Activate or establish emergency response
Resource allocation	Crisis management plans
Resource allocation	Measures to ensure security of supply
Resource allocation	Police and army interventions
Resource allocation	Provide international help
Resource allocation	Receive international help
Resource allocation	The government provide assistance to vulnerable populations
Risk communication	Actively communicate with healthcare professionals
Risk communication	Actively communicate with managers
Risk communication	Educate and actively communicate with the public
Risk communication	Travel alert and warning
Social distancing	Closure of educational institutions
Social distancing	Mass gathering cancellation
Social distancing	Measures for public transport
Social distancing	Measures for special populations
Social distancing	Return operation of nationals
Social distancing	Small gathering cancellation
Social distancing	Special measures for certain establishments
Social distancing	Work safety protocols
Travel restriction	Airport restriction
Travel restriction	Border restriction
Travel restriction	Cordon sanitaire

CCCSL	
L1 Theme	L2 Category
Travel restriction	Individual movement restrictions
Travel restriction	National lockdown
Travel restriction	Port and ship restriction
Travel restriction	Public transport restriction

Supplementary Table 5: Full table of NPI Themes and Categories in CCCSL without abbreviations.

Continents	Number of consensus measures	Expectation value
Europe and Africa	4	1.05
Asia and Oceania	2	1.05
Americas	6	2.40

Supplementary Table 6: Number of consensus NPIs after removal of the indicated continents, compared to the expected number of consensus NPIs if the significance of the NPIs in the different methods was statistically independent.

6 References

1. Desvars-Larrive, A. *et al.* A structured open dataset of government interventions in response to COVID-19. *Scientific Data* **7**, 285 (2020). URL `https://doi.org/10.1038/s41597-020-00609-9`.

2. World population prospects 2019. `https://population.un.org/wpp/DataQuery/`. Accessed: 2020-04-24.

3. World bank international comparison program database. https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD. Accessed: 2020-04-24.

4. Conceicao, P. *et al.* *Human Development Report 2019* (United Nations Development Programme, 2019).

5. World governance indicators. https://info.worldbank.org/governance/wgi/. Accessed: 2020-04-24.

6. Banholzer, N. *et al.* Estimating the impact of non-pharmaceutical interventions on documented infections with COVID-19: A cross-country analysis. *medRxiv* (2020).

7. Tibshirani, R. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)* **58**, 267–288 (1996).

8. Taylor, J. & Tibshirani, R. J. Statistical learning and selective inference. *Proceedings of the National Academy of Sciences* **112**, 7629–7634 (2015).

9. Hastie, T., Tibshirani, R. & Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (Springer, 2009).

10. Pedregosa, F. *et al.* Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research* **12**, 2825–2830 (2011).

11. Vaswani, A. *et al.* Attention is all you need. In Guyon, I. *et al.* (eds.) *Advances in Neural Information Processing Systems 30*, 5998–6008 (Curran Associates, Inc., 2017). URL http://papers.nips.cc/paper/7181-attention-is-all-you-need.pdf.

12. Islam, N. *et al.* Physical distancing interventions and incidence of coronavirus disease 2019: natural experiment in 149 countries. *BMJ* **370** (2020). URL https://www.bmj.com/content/370/bmj.m2743. https://www.bmj.com/content/370/bmj.m2743.full.pdf.

13. Flaxman, S. *et al.* Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe. *Nature* (2020). URL https://doi.org/10.1038/s41586-020-2405-7.

14. Auger, K. A. *et al.* Association between Statewide School Closure and COVID-19 Incidence and Mortality in the US. *JAMA - Journal of the American Medical Association* **45229** (2020).

15. Liu, Y. *et al.* The impact of non-pharmaceutical interventions on SARS-CoV-2 transmission across 130 countries and territories. *medRxiv* (2020).

16. Park, Y., Choe, Y. *et al.* Contact tracing during coronavirus disease outbreak. *Emerg Infect Dis.* (2020).

17. Adams, J. G. & Walls, R. M. Supporting the Health Care Workforce During the COVID-19 Global Epidemic. *JAMA* **323**, 1439–1440 (2020). URL `https://doi.org/10.1001/jama.2020.3972`. `https://jamanetwork.com/journals/jama/articlepdf/2763136/jama_adams_2020_vp_200048.pdf`.

18. Bloomberg. Slovene economy minister survives ouster bid over PPE scandal (2020). URL `https://www.bloomberg.com/news/articles/2020-06-11/slovene-economy-minister-survives-ouster-bid-over-ppe-scandal`. Accessed: 2020-07-02.

19. India Today. Whiff of a scandal (2020). URL `https://www.indiatoday.in/india-today-insight/story/whiff-of-a-scandal-1688053-2020-06-11`. Accessed: 2020-07-02.

20. The Telegraph. PPE: Government counted each glove as single item to reach one billion total, investigation shows (2020). URL `https://www.telegraph.co.uk/politics/2020/04/28/ppe-government-counted-glove-single-item-reach-one-billion-total/`. Accessed: 2020-07-02.

21. Post, R. A. J., Regis, M., Zhan, Z. & van den Heuvel, E. R. How did governmental interventions affect the spread of COVID-19 in european countries? *medRxiv* (2020).

22. Busisness Insider. 14 countries that are paying their workers during quarantine (2020). URL `https://www.businessinsider.com/countries-offering-direct-payments-or-basic-income-in-corona-crisis-2020-4?r=DE&IR=T`. Accessed: 2020-07-03.

23. BBC News. Coronavirus bailouts: Which country has the most generous deal? (2020). URL `https://www.bbc.com/news/business-52450958`. Accessed: 2020-07-03.

24. Wong, J. E., Leo, Y. S. & Tan, C. C. COVID-19 in Singapore — current experience: critical global issues that require attention and action. *Jama* **323**, 1243–1244 (2020).