

# The methodologies to assess the effects of non-pharmaceutical interventions during COVID-19: a systematic review

Nicolas Banholzer<sup>1\*‡</sup>, Adrian Lison<sup>2\*‡</sup>, Dennis Özcelik<sup>3</sup>, Tanja Stadler<sup>2</sup>, Stefan Feuerriegel<sup>1,4</sup>, and Werner Vach<sup>5,6</sup>

<sup>1</sup>ETH Zurich, Department of Management, Technology, and Economics, Zurich, Switzerland

<sup>2</sup>ETH Zurich, Department of Biosystems Science and Engineering, Zurich, Switzerland

<sup>3</sup>ETH Zurich, Chemistry | Biology | Pharmacy Information Center, Zurich, Switzerland

<sup>4</sup>LMU Munich, LMU Munich School of Management, Munich, Germany

<sup>5</sup>Basel Academy for Quality and Research in Medicine, Basel, Switzerland

<sup>6</sup>University of Basel, Department of Environmental Sciences, Basel, Switzerland

\*Corresponding authors: [nbanholzer@ethz.ch](mailto:nbanholzer@ethz.ch), [adrian.lison@bsse.ethz.ch](mailto:adrian.lison@bsse.ethz.ch)

‡These authors contributed equally to this work.

## Abstract

Non-pharmaceutical interventions, such as school closures and stay-at-home orders, have been implemented around the world to control the spread of SARS-CoV-2. Their effects on health-related outcomes have been the subject of numerous empirical studies. However, these studies show fairly large variation among methodologies in use, reflecting the absence of an established methodological framework. On the one hand, variation in methodologies may be desirable to assess the robustness of results; on the other hand, a lack of common standards can impede comparability among studies. To establish a comprehensive overview over the methodologies in use, we conducted a systematic review of studies assessing the effects of non-pharmaceutical interventions on health-related outcomes between January 1, 2020 and January 12, 2021 (n=248). We identified substantial variation in methodologies with respect to study setting, outcome, intervention, methodological approach, and effect assessment. On this basis, we point to shortcomings of existing studies and make recommendations for the design of future studies.

**Keywords:** non-pharmaceutical interventions, social distancing measures, control measures, COVID-19, systematic review, methodology review

## 28 **1 Introduction**

29 In response to the COVID-19 pandemic, countries around the world have implemented non-  
30 pharmaceutical interventions. These include a variety of public health measures implemented  
31 by governments with the intention of controlling, preventing, and mitigating transmission, e.g.  
32 school closures, stay-at-home orders, and mandates for compulsory wearing of masks in public  
33 places<sup>1-4</sup>. The widespread use of these interventions has raised interest in empirically studying their  
34 effects on health-related outcomes reflecting disease dynamics, e.g. the number of new cases or  
35 infection rates<sup>5-10</sup>. Such studies can play an important role in informing the discussion about the  
36 effectiveness of interventions. In particular, insights from the COVID-19 pandemic may contribute  
37 to an evidence-based public-health response in subsequent COVID-19 waves or future pandemics.

38 Accordingly, a plethora of studies assessing the effects of non-pharmaceutical interventions during  
39 the COVID-19 pandemic have been published. Their findings have been summarized by several  
40 meta-analyses<sup>11-15</sup>; nonetheless, each meta-analysis considered a different subset of studies. We  
41 argue that the latter is due to substantial variation in the methodologies used to conduct empirical  
42 studies on the effects of non-pharmaceutical interventions. The resulting lack of similarity constrains  
43 meta-analyses to a comparably small and specific subset of the overall evidence.

44 There are different reasons to expect variation in methodologies in the studies on the effectiveness  
45 of non-pharmaceutical interventions for controlling a pandemic. One possibility is the lack of empirical  
46 data before the COVID-19 pandemic, so that early studies have been largely theoretical<sup>16</sup>. Empirically  
47 assessing the effects of non-pharmaceutical interventions is therefore a relatively new subject, and  
48 corresponding studies do not build on an established scientific framework. Another possibility is  
49 that empirical assessments have been approached with different methods and domain knowledge by  
50 researchers from various fields, e.g. computational biology, infectious disease epidemiology, public  
51 health, economics, and statistical modeling.

52 Variation in methodologies can be manifold. Different study settings, outcomes, interventions,  
53 methodological approaches, and ways to assess effects may be used. On the one hand, such variation  
54 may be desired as it allows to assess the robustness of results against individual assumptions and  
55 methodologies. On the other hand, variation in methodologies can impede comparability among  
56 studies, which is necessary to arrive at conclusive evidence regarding the effects of non-pharmaceutical

57 interventions.

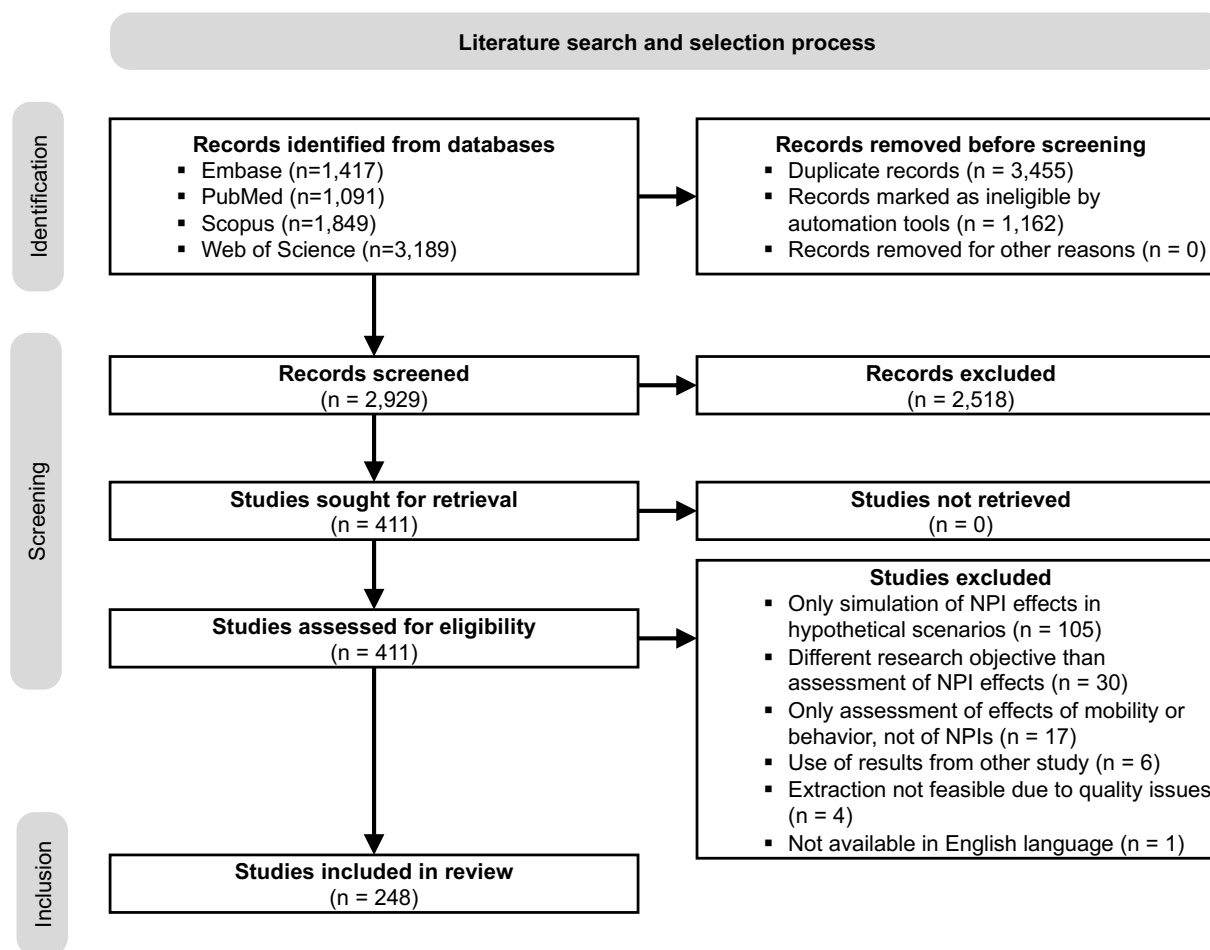
58 Here, we systematically review the methodologies for studying the effects of non-pharmaceutical  
59 interventions on health-related outcomes published between January 1, 2020 and January 12, 2021  
60 (n=248). Thereby, we aim to inform about different methodologies that were used by previous studies  
61 and promote common standards so that future studies can align with existing ones. In particular, we  
62 explore shortcomings of current studies and provide seven recommendations for subsequent studies  
63 on the effects of non-pharmaceutical interventions.

## 64 **2 Results**

65 Our review follows general guidelines for systematic literature reviews<sup>17</sup> and is reported according  
66 the PRISMA 2020 statement<sup>18</sup>. The methodology was preregistered in a review protocol at PROS-  
67 PERO<sup>19</sup>. Figure 1 shows the PRISMA flow diagram of our identification process. We conducted  
68 a systematic database search for peer-reviewed research articles from January 1, 2020 up to Jan-  
69 uary 12, 2021 (see [Materials and methods](#), Section 5), yielding 2,929 unique records of studies for  
70 screening. Through title and abstract screening, we identified 411 studies as potentially relevant  
71 and evaluated their full texts. Of these, we excluded 163 studies that did not meet the eligibility  
72 criteria. The most frequent reasons for exclusion were that (i) studies primarily simulated the effects  
73 of interventions in hypothetical scenarios rather than making inferences from observational data;  
74 (ii) studies had a different objective than assessing intervention effects, and (iii) studies only assessed  
75 the effects of population behavior (most often mobility) on health-related outcomes, but not the  
76 effects of interventions. The remaining n=248 studies met our eligibility criteria and were included  
77 for subsequent data extraction. Importantly, 35 studies in our review sample contained multiple  
78 (i. e. up to three) analyses, e. g. with different methodological approaches, leading to 285 different  
79 analyses included. If not indicated otherwise, our results are presented at the level of individual  
80 analyses (and *not* at the level of studies).

81 We characterized the analyses along five dimensions (SI Appendix D): study setting (D.1),  
82 outcome (D.2), intervention (D.3), methodological approach (D.4), and effect assessment (D.5).  
83 In the Results section, if not stated otherwise, we use the term *interventions* to refer to non-  
84 pharmaceutical interventions. Where appropriate, we also point to exemplary studies of specific  
85 characteristics. Due to the large size of our review sample, however, we refrain from referencing

86 all studies in the main manuscript and instead refer to our complete data extraction report in SI  
87 Appendix E.



**Fig 1. PRISMA flow diagram.** Overall, n=248 studies were included. Some studies contain multiple analyses, such that the number of analyses included in the review is 285.

## 88 2.1 Study setting

89 The analyses vary in their scope across populations, geographic areas, and study period. A systematic  
90 classification of the study setting is shown in Table 1.

### 91 **Population**

92 More than half of the analyses studied multiple populations (n=167; 59%), i. e. multiple countries  
93 or subnational regions (e. g. states or cities). The remainder focused on a single population (n=118;  
94 41%), i. e. a single country or subnational region. The analyses were performed at the national level  
95 (n=117; 41%), the subnational level (n=97; 34%), or both (n=71; 25%). If both levels were studied,

**Table 1.** Systematic classification and frequency of the study setting (D.1).

<b>D.1.1: Number of populations included</b>	Frequency
Single (country, state, city, etc.)	118 (41 %)
Multiple (countries, states, cities, etc.)	167 (59 %)
<b>D.1.2: Level of populations included</b>	
National (country-level)	117 (41 %)
Subnational (e. g. state-level)	71 (25 %)
Both national and subnational (country- and e. g. state-level)	97 (34 %)
<b>D.1.3: Geographic areas covered<sup>‡</sup></b>	
Asia	144 (51 %)
Europe	109 (38 %)
North America	91 (32 %)
Middle East and Africa	49 (17 %)
Central and South America	46 (16 %)
Oceania	42 (15 %)
<b>D.1.4: Number of countries covered</b>	
Multiple countries	66 (23 %)
Single country (including multiple populations from a single country) <sup>‡</sup>	219 (77 %)
↳ China	↳ 54 (25 %)
↳ United States	↳ 43 (20 %)
↳ India	↳ 11 ( 5 %)
↳ Italy	↳ 11 ( 5 %)
↳ Other	↳ 100 (46 %)
<b>D.1.5: Study period</b>	
Start and end date span first epidemic wave	161 (56 %)
One or more exceptions <sup>‡</sup>	124 (44 %)
↳ End date in growth phase of wave	↳ 44 (35 %)
↳ Same end date for several populations with diverse epidemic trajectories	↳ 38 (31 %)
↳ End date at peak of wave	↳ 16 (13 %)
↳ End date could not be evaluated	↳ 14 (11 %)
↳ Other	↳ 14 (11 %)

<sup>‡</sup> Multiple categories per analysis are possible. Frequencies refer to number of analyses to which category applies, proportions thus do not sum to 100 %.

96 the country and all its subnational regions were oftentimes considered, e. g. all states of the United  
97 States. Geographically, Asia (n=144; 51 %), Europe (n=109; 38 %), and North America (n=91;  
98 32 %) were much more frequently analyzed than the rest of the world. Regional disparities may be  
99 explained by the fact that analyses frequently focused on some specific countries. For instance, in  
100 analyses of a single country, the most frequent country was China (n=54; 25 %), from where the  
101 pandemic originated. Other frequently studied countries, such as the United States (n=43; 20 %),  
102 India (n=11; 5 %), and Italy (n=11; 5 %), may have received more attention due to particularly high  
103 incidence and mortality during the first epidemic wave.

## 104 **Study period**

105 Typically, the study period covered both a rise and decline in new cases of the first epidemic wave  
106 in the analyzed population, and started before and ended after the analyzed interventions were  
107 implemented (n=161; 56 %). However, many analyses also deviated from this pattern in one or  
108 several aspects. Most often, study periods had a comparatively early end date, i. e. the study period  
109 ended already at the peak (n=16; 13 %) or still in the growth phase (n=44; 35 %) of the wave. There  
110 was also a considerable number of analyses that used the same study period for populations which  
111 were in different epidemic phases (n=38; 31 %). In such cases, the end date of the study period was  
112 still within the epidemic growth phase for some populations but already in the control phase for  
113 other populations.

## 114 **2.2 Outcome**

115 The studies in our review sample used different types of health-related outcomes or surrogates. For  
116 every analysis, we identified the “raw outcome”, i. e. the outcome data which were self-collected or  
117 obtained from external sources and used as input for the analysis. In around half of the analyses, the  
118 raw outcome was analyzed directly to assess the effects of interventions. The other half of analyses,  
119 however, involved an intermediate step, in which another outcome was computed from the raw  
120 outcome. This “computed outcome” was then analyzed instead of the raw outcome, or sometimes in  
121 addition to it. A systematic classification of the outcomes are shown in Table 2.

### 122 **Raw outcome**

123 We identified three main types of raw outcome data used, namely (i) epidemiological population-level  
124 data, (ii) epidemiological individual-level data, and (iii) behavioral data.

#### 125 (i) Epidemiological population-level data

126 The majority of analyses used population-level data on epidemiological outcomes (n=223; 78 %).  
127 The most frequent types were surveillance data, mainly the number of confirmed cases, but also  
128 deaths, hospitalizations, recovered cases, and, less frequently, intensive care unit (ICU) admissions.  
129 Importantly, some of these outcomes, such as recovered cases, were predominantly used to fit  
130 transmission models, in which case the effect of interventions was rather measured in terms of a  
131 different latent outcome (see D.5 [Methodological approach](#), Section 2.4). Frequently, authors also

132 included several types of data (e. g. both cases and deaths), either to perform a separate analysis for  
133 each (e. g. as a robustness check) or to combine them in a joint model (e. g. a transmission model).  
134 Some analyses used surveillance data on other diseases than COVID-19, with influenza being the  
135 most popular choice. Such surrogate diseases have often been monitored over an extended period of  
136 time, which allows comparing their spread during the COVID-19 pandemic to earlier years. Notably,  
137 we found only three analyses that used external data on latent epidemiological population-level  
138 outcomes (e. g. the reproduction number). All other analyses using a latent outcome self-computed  
139 it from raw data in an intermediate step (see D.2.2 [Computed outcome](#), Section 2.2).

#### 140 (ii) Epidemiological individual-level data

141 Instead of population-level data, some analyses also used individual-level epidemiological data  
142 (n=23; 8%). These were in particular data about individual cases with case ID, demographics,  
143 and epidemiological characteristics (e. g. the date of symptom onset or travel history). In some  
144 instances, this included contact tracing data with links between index and secondary cases, allowing  
145 the reconstruction of transmission chains. Two analyses also used genome sequence data of clinical  
146 SARS-CoV-2 samples<sup>20,21</sup>.

#### 147 (iii) Behavioral data

148 In addition to epidemiological data, a relevant share of analyses employed data on population  
149 behavior (n=55; 19%), mainly mobility data. These data were usually obtained through tracking of  
150 mobile phone movements and provided as aggregates at the population level, based on summary  
151 statistics such as the daily number of trips made, distance travelled, time spent at certain locations,  
152 or population flow between regions. Another, less frequently used source of information on human  
153 behavior were surveys regarding social distancing practices, such as adherence to interventions, face  
154 mask usage, daily face-to-face contacts, or recent traveling.

**Table 2.** Systematic classification and frequency of the outcome (D.2).

<b>D.2.1: Raw outcome<sup>‡</sup></b>	Frequency
Epidemiological population-level outcome <sup>‡</sup>	223 (78%)
↳ Confirmed cases	↳ 186 (83%)
↳ Deaths	↳ 64 (29%)
↳ Recovered cases	↳ 20 (9%)
↳ Hospitalizations	↳ 18 (8%)
↳ Surrogate disease outcome	↳ 10 (4%)
↳ Other	↳ 24 (11%)
Epidemiological individual-level outcome <sup>‡</sup>	23 (8%)
↳ Individual cases	↳ 11 (48%)
↳ Individual cases and transmission chains	↳ 8 (35%)
↳ Genome sequence data	↳ 4 (17%)
Behavioral outcome <sup>‡</sup>	55 (19%)
↳ Mobility	↳ 50 (91%)
↳ Survey responses	↳ 6 (11%)
<b>D.2.2: Time resolution of raw outcome</b>	
Daily	269 (94%)
Other (weekly, biweekly, monthly, or not applicable)	16 (6%)
<b>D.2.3: Computed outcome<sup>‡</sup></b>	
None (only raw outcomes used)	150 (53%)
Measure of epidemic trend <sup>‡</sup>	34 (12%)
↳ Growth rate	↳ 24 (71%)
↳ Doubling time	↳ 11 (32%)
↳ Other	↳ 1 (3%)
Epidemiological parameter <sup>‡</sup>	89 (31%)
↳ Reproduction number	↳ 78 (88%)
↳ Transmission rate	↳ 6 (7%)
↳ Other	↳ 16 (18%)
Summary statistic	8 (3%)
Change points	7 (2%)
Other	9 (3%)
<b>D.2.4: Method to obtain the computed outcome</b>	
None (no computed outcome)	150 (53%)
One or several methods used <sup>‡</sup>	135 (47%)
↳ Simple computation (e. g. ratio, sum etc.)	↳ 35 (26%)
↳ Exponential growth model	↳ 11 (8%)
↳ Compartmental transmission model	↳ 35 (26%)
↳ Statistical estimation of reproduction number	↳ 43 (32%)
↳ Other	↳ 29 (21%)
<b>D.2.5: Data source<sup>‡</sup></b>	
Could not be evaluated	10 (4%)
Data from (sub)national authorities	141 (49%)
Data from publicly available cross-country selections	77 (27%)
Mobility data from corporate organizations	40 (14%)
Other	54 (19%)
<b>D.2.6: Data availability</b>	
Data access via source	173 (61%)
Data made available by the authors	76 (27%)
Data not accessible	36 (13%)

<sup>‡</sup> Multiple categories per analysis are possible. Frequencies refer to number of analyses to which category applies, proportions thus do not sum to 100%.



155 ***Time resolution of raw data***

156 Almost all raw data were obtained at daily resolution (n=269; 94%). Exceptions were data on  
157 surrogate diseases or from surveys, where reporting was usually only broken down in weekly, biweekly,  
158 or monthly intervals (n=16; 6%).

159 ***Computed outcome***

160 Around half of the analyses involved an intermediate step, in which a latent outcome was computed  
161 from the raw data before assessing the effects of interventions. We identified four main types of  
162 computed outcomes:

- 163 1) Measures of epidemic trend (n=34; 12%) were computed to describe the overall trend of the  
164 epidemic, e. g. through the growth rate or doubling time of confirmed cases or hospitalizations.
- 165 2) Epidemiological parameters (n=89; 31%) were computed to measure specific infection dynamics,  
166 oftentimes in terms of the reproduction number. A few analyses also used individual-level  
167 epidemiological data to compute epidemiologically relevant time spans such as the serial interval  
168 or the time from symptom onset to isolation.
- 169 3) Summary statistics (n=8; 3%) were computed to describe the progression of an epidemic in a  
170 certain population, e. g. the time until a certain number of documented cumulative cases was  
171 reached, or the time until the reproduction number first fell below one.
- 172 4) Change points in the outcome (n=7; 2%) were computed with the aim to find time points of  
173 presumably structural changes in epidemic dynamics and compare them with implementation  
174 dates of interventions in the subsequent analysis<sup>10,22,23</sup>. Typically, change points were computed  
175 for the time series of confirmed cases or mobility.

176 Of note, the raw outcome was not always used only for obtaining the computed outcome, e. g.  
177 changes both in the number of new confirmed cases (raw outcome) and in the reproduction number  
178 (computed outcome) were sometimes analyzed.

179 ***Method to obtain the computed outcome***

- 180 1) Measures of epidemic trend were often obtained through simple computation (e. g. growth rate  
181 as percentage change in confirmed cases). Other analyses used simple modeling approaches,

182 e. g. fitting an exponential growth model to the time series and extracting the exponential  
183 growth rate or doubling time from the estimated parameters.

184 2) Epidemiological parameters were mostly estimated from confirmed cases or deaths. Some  
185 approaches fitted a compartmental transmission model to the raw epidemiological outcome.  
186 For this, the parameter of interest was either allowed to vary over time, or the model was fitted  
187 independently on different time periods. Other approaches employed a statistical method to  
188 directly estimate reproduction numbers from the observed outcome. Here, the method by Cori  
189 et al.<sup>24</sup> as implemented in the popular software package “EpiEstim”<sup>25</sup> for estimation of the  
190 instantaneous effective reproduction number was used in a large number of analyses. However,  
191 we found that statistical methods were not always applied correctly, which could have led to  
192 bias in the inferred transmission dynamics (see SI Appendix A). Sometimes, authors also used  
193 methods to estimate reproduction numbers from contact matrices<sup>26</sup> (derived from surveys on  
194 personal contacts) or from transmission chains<sup>27,28</sup> (derived from contact tracing data).

195 3) Summary statistics were typically obtained through simple computation.

196 4) Change points in the outcome were obtained by fitting a compartmental transmission model  
197 with special parameters representing points in time when the transmission rate changes. Other  
198 analyses used special change point detection algorithms.

### 199 **Data source**

200 The majority of authors directly accessed surveillance data from national health authorities or other  
201 governmental bodies (n=141; 49 %). In the case of individual-level data, which may be subject  
202 to privacy regulations, authors were often themselves affiliated to the relevant health authority.  
203 To obtain population-level data, a considerable share of analyses also used publicly available data  
204 from cross-country selections (n=77; 27 %), e. g. the European Centre for Disease Prevention and  
205 Control (ECDC)<sup>29</sup>, the Johns Hopkins University (JHU)<sup>30</sup>, or Worldometer<sup>31</sup>, which offer aggregated  
206 surveillance data internationally from various sources for the pandemic. Mobile phone tracking  
207 data were usually provided by corporate organizations (n=40; 14 %) such as Google<sup>32</sup>, Apple<sup>33</sup>, or  
208 Baidu<sup>34</sup>. A few analyses were also based on data collected by the authors, e. g. survey data on  
209 behavioral outcomes, seroprevalence studies, or data collected at a local facility such as a hospital.

## 210 **Data availability**

211 Data for the raw outcome was usually publicly available, in particular for epidemiological population-  
212 level outcomes such as cases and deaths because such data could oftentimes be accessed via the  
213 source that is documented in the manuscript (n=173; 61%). In several cases, the data was made  
214 publicly available by the study authors (n=76; 27%), e. g. by depositing the analyzed data in a  
215 public repository. For a small, yet considerable number of analyses, data was not accessible (n=36;  
216 13%) as the data was neither made publicly available nor the source of the data could be identified.  
217 Of note, data on epidemiological individual-level data was typically not available due to privacy  
218 concerns. Furthermore, corporate mobility data was widely available in the past, but access has  
219 recently been restricted by many providers.

## 220 **2.3 Intervention**

221 The analyses vary in the types of exposures and non-pharmaceutical interventions. A systematic  
222 classification is shown in Table 3.

### 223 ***Terminology for non-pharmaceutical interventions***

224 Varying terminology was used by the literature to refer to non-pharmaceutical interventions. This is  
225 reflected in our search string, where we used a large set of terms in order to capture a broad range of  
226 relevant studies. One part of our search string considers the different terminology for interventions  
227 in general, while the other part considers the terminology for the specific type of non-pharmaceutical  
228 interventions. In our review, the most frequent terms for interventions were “measures” (n=135;  
229 54%), followed by “interventions” (n=65; 26%) and “policies” (n=16; 6%). The most frequent  
230 terms for the specific type of non-pharmaceutical interventions were “non-pharmaceutical” (n=49;  
231 16%), “control” (n=48; 16%), and “social distancing” (n=45; 15%). While terminology sometimes  
232 reflected the specific types of non-pharmaceutical interventions that were analyzed, differences in  
233 terminology may also be the result of different research backgrounds of the study authors.

**Table 3.** Systematic classification and frequency of the interventions (D.3).

<b>D.3.1: Terminology for interventions<sup>†‡</sup></b>	Frequency
Not applicable (only specific term for intervention type)	22 ( 9 %)
Measures	135 (54 %)
Interventions	65 (26 %)
Policies	16 ( 6 %)
Other	14 ( 6 %)
<b>D.3.2: Terminology for the specific type of non-pharmaceutical interventions<sup>†‡</sup></b>	Frequency
Not applicable (only general term for interventions)	3 ( 1 %)
Non-pharmaceutical	49 (16 %)
Control	48 (16 %)
Social distancing	45 (15 %)
Other	159 (52 %)
<b>D.3.3: Exposure types</b>	
One single intervention	43 (15 %)
Multiple separate interventions	31 (11 %)
One combination of interventions	84 (29 %)
Multiple combinations of interventions	20 ( 7 %)
All interventions together	70 (25 %)
Other	37 (13 %)
<b>D.3.4: Types of single interventions</b>	
Not applicable (no single interventions analyzed)	211 (74 %)
One or multiple single interventions analyzed (as defined in D.3.4 of the Documentation manual) <sup>‡</sup>	74 (26 %)
↳ Stay-at-home order	↳ 44 (59 %)
↳ Other	↳ 27 (36 %)
↳ School closure	↳ 25 (34 %)
↳ Workplace closure	↳ 20 (27 %)
↳ International travel restrictions	↳ 17 (23 %)
↳ Declaration of a state of emergency	↳ 13 (18 %)
↳ Bans of large gatherings	↳ 13 (18 %)
↳ Venue closure	↳ 12 (16 %)
↳ Bans of small gatherings	↳ 10 (14 %)
<b>D.3.5: Coding of interventions</b>	
↳ <b>D.3.6: Source of intervention data</b>	
Not applicable (no specific interventions analyzed)	74 (26 %)
Not necessary (no joint analysis of interventions across multiple populations)	137 (48 %)
↳ Could not be evaluated	↳ 98 (72 %)
↳ Government or news websites	↳ 30 (22 %)
↳ Other	↳ 9 ( 7 %)
Necessary (joint analysis of interventions across multiple populations)	74 (26 %)
↳ Could not be evaluated	↳ 9 (12 %)
↳ Coding done by authors	↳ 20 (27 %)
↳ Use of externally coded data	↳ 45 (61 %)
<b>D.3.7: Availability of data on exposure</b>	
Not applicable (no specific interventions analyzed)	73 (26 %)
Raw data documented in the manuscript	136 (48 %)
Access to externally coded data via source	32 (11 %)
Coded data made available by the authors	34 (12 %)
Coded data not available	10 ( 4 %)

<sup>†</sup> Results for this subdimension are reported at the study-level, and not the level of analysis (i.e. one study can contain multiple analyses). If a study uses more than one term predominantly, then both are counted and added to the total count.

<sup>‡</sup> Multiple categories per analysis are possible. Frequencies refer to number of analyses to which category applies, proportions thus do not sum to 100 %.

### 234 ***Exposure types and types of single interventions***

235 A considerable number of analyses examined one single (n=43; 15 %) <sup>5,35</sup> or multiple interventions  
236 separately (n=31; 11 %) <sup>2,7</sup>. Among these analyses, the effects of school closures (n=25; 34 %)   
237 and stay-at-home orders (n=44; 59 %) were assessed most frequently, which may be due to these  
238 interventions being particularly controversial in the public discourse <sup>36,37</sup>. The majority of analyses,  
239 however, did not analyze the separate effects of multiple interventions but rather analyzed the joint  
240 effect for a combination of multiple interventions (n=84; 29 %) <sup>38-40</sup>, which is often the case when  
241 multiple interventions were implemented on the same day and when thus the separate effects could not  
242 be disentangled. A considerable number of analyses were even less specific by only analyzing whether  
243 interventions were altogether effective but without attributing effects to specific interventions (n=70;  
244 25 %) <sup>41,42</sup>. Other ways to assess the effectiveness of interventions were: examining the start time of  
245 intervention <sup>23,43</sup>, e. g. to assess the effect of different delays with which governments responded to  
246 the pandemic <sup>43</sup>; dividing the public health response into different periods <sup>44,45</sup>; dividing interventions  
247 into different categories <sup>46,47</sup>; or summarizing the stringency of interventions to a numerical index at  
248 a specific time point <sup>48,49</sup>.

249 Of note, analyses that assessed the effects of a combination of interventions often referred to this  
250 combination as “lockdown”. In the underlying analysis, such lockdowns typically included multiple  
251 interventions implemented on the same day <sup>39,50</sup>. However, the specific interventions included in  
252 lockdowns varied considerably between populations. We therefore considered “lockdown” as an  
253 umbrella term for different combinations of interventions rather than as a specific type of intervention.  
254 Furthermore, some studies did not only assess the effects of non-pharmaceutical interventions on  
255 mobility, but also the effects of changes in mobility on population-level epidemiological outcomes.  
256 In these analyses, human mobility was typically defined as a continuous exposure. We extracted  
257 information on such complementary analyses of mobility as an addendum to the main review (see SI  
258 Appendix E).

### 259 ***Coding of interventions***

260 When multiple populations were jointly analyzed, coding of interventions may have been necessary  
261 in order to reconcile differences in the definitions of interventions between populations. For instance,  
262 the term “school closures” could refer to the closure of primary or secondary schools or universities.

263 Differences across populations are thus reconciled during coding by deciding upon the type of  
264 intervention and providing a common name and definition that is then applied to all populations.  
265 Such coding of interventions was necessary in around a quarter of analyses (n=74; 26 %).

#### 266 **Source of intervention data**

267 If coding of interventions was *not* necessary, authors often obtained intervention data (i. e. the date  
268 of interventions) from a government or news website (n=30; 22 %). Unfortunately however, the data  
269 source was often not provided by the authors and could thus not be evaluated (n=98; 72 %). If  
270 coding of interventions was necessary, then study authors either coded the data themselves (n=20;  
271 27, %), i. e. collected the data from government or news websites and systematically categorized  
272 them, or used externally coded data instead (n=45; 61 %). The most popular choices for externally  
273 coded data were the Oxford Government Response Tracker<sup>1</sup> and, for the United States, the New  
274 York Times<sup>51</sup>.

#### 275 **Availability of data on exposure**

276 Authors using exposure data (type of exposure, interventions, and implementation dates) where  
277 coding of interventions was not necessary usually documented the data in the manuscript. Some  
278 authors using externally coded data chose to make the data available themselves (n=34; 12 %), e. g.  
279 by depositing it in a public repository, although many only referenced external data (n=32; 11 %).

## 280 **2.4 Methodological approach**

281 A variety of methodological approaches were used to assess the effects of interventions. The  
282 methodological approaches extracted here describe the actual stage of estimating the intervention  
283 effect. A systematic classification of the methodological approaches is shown in Table 4.

#### 284 **Empirical approach**

285 We distinguished three general empirical approaches for assessing the effects of interventions, namely  
286 **(D)** descriptive, **(P)** parametric, and **(C)** counterfactual approaches.

- 287 • **(D)** Descriptive approaches (n=151; 53 %): These approaches provided descriptive summaries  
288 of the outcome over time or between populations, and related variation in these summaries to  
289 the presence or absence of different interventions. For example, some analyses compared changes  
290 in the growth rate of observed cases before and after interventions were implemented<sup>52,53</sup>. Of

**Table 4.** Systematic classification and frequency of the methodological approach (D.4).

<b>D.4.1: Empirical approach</b>				Total freq.
<b>D:</b> Descriptive				151 (53 %)
<b>P:</b> Parametric				94 (33 %)
<b>C:</b> Counterfactual				40 (14 %)
<b>D.4.2: Use of exposure variation</b>	(D)	(P)	(C)	
Only variation over time for a single population	78	23	24	125 (44 %)
Only variation over time for multiple populations	63	22	10	95 (33 %)
Only variation between populations	4	14	0	18 ( 6 %)
Both variation over time and between populations	6	35	6	47 (16 %)
<b>D.4.3: Method</b>				
Description of change over time	136	—	—	136 (48 %)
↳ Description of time course				↳ 49 (36 %)
↳ Comparison of time periods				↳ 87 (64 %)
Comparison of populations	8	—	—	8 ( 3 %)
Comparison of change points with intervention dates	7	—	—	7 ( 2 %)
Non-mechanistic model	—	61	17	78 (27 %)
↳ Generalized linear model				↳ 51 (65 %)
↳ Exponential growth model				↳ 11 (14 %)
↳ Other				↳ 16 (21 %)
Mechanistic model	—	30	13	43 (15 %)
↳ Compartmental single-population transmission model				↳ 29 (67 %)
↳ Compartmental meta-population transmission model				↳ 4 ( 9 %)
↳ Semi-mechanistic Bayesian transmission model				↳ 5 (12 %)
↳ Other				↳ 5 (12 %)
Synthetic controls	—	—	6	6 ( 2 %)
Other	0	3	4	7 ( 2 %)
<b>D.4.4: Code availability</b>				
None (not available)	121	66	33	220 (77 %)
Publicly available	30	28	7	65 (23 %)

*Empirical approach:* (D) descriptive, (P) parametric, and (C) counterfactual

‡ Multiple categories per analysis are possible. Frequencies refer to number of analyses to which category applies, proportions thus do not sum to 100 %.

291 note, descriptive approaches could involve modeling as part of an intermediate step, where a  
 292 latent outcome was computed from the raw outcome (see [Computed outcome](#)), while, afterward,  
 293 a descriptive approach was used to assess the effect of interventions on the latent outcome.  
 294 For example, some analyses used a single-population compartmental transmission model to  
 295 estimate the time-varying reproduction number and then compared the reproduction number  
 296 before and after interventions were implemented<sup>54–56</sup>.

- 297 • (P) Parametric approaches (n=94; 33 %): These approaches formulated an explicit link  
 298 between the intervention and the outcome, where the effects of interventions were quantified  
 299 via a parameter in a model. Most frequently these were regression-like links to estimate the

300 effects of interventions on the reproduction number<sup>2,8</sup>.

- 301 • (C) Counterfactual approaches (n=40; 14 %): These approaches assessed the effect of inter-  
302 ventions by comparing the observed outcome with a counterfactual outcome based on an  
303 explicit scenario in which the interventions were not implemented. For example, the observed  
304 number of cases was compared with the number of cases that would have been observed if the  
305 exponential growth in cases had continued as before the implementation of interventions<sup>57,58</sup>.

### 306 ***Use of exposure variation***

307 Effects of interventions were assessed by exploiting variation in the exposure to the intervention  
308 over time, between populations, or both. Assessments exploiting exposure variation over time  
309 contrasted the outcome in time periods when specific measures were in place with the outcome in  
310 time periods when they were not in place. In contrast, assessments exploiting exposure variation  
311 between populations were based on a comparison of the outcome between populations that were  
312 subject to specific measures with populations that were not. In the majority of analyses, a single  
313 population was studied and thus only variation over time could be exploited in order to assess the  
314 effects of interventions (n=125; 44 %). In around one third of analyses, multiple populations were  
315 studied but only variation over time was exploited (n=95; 33 %), i. e. the effects of interventions  
316 were assessed within each population separately. A small number of analyses exploited only variation  
317 between populations (n=18; 6 %). Less than one in five analyses exploited both variation over time  
318 and between populations (n=47; 16 %).

### 319 ***Method***

320 We grouped the different methods used into (i) description of change over time, (ii) comparison of  
321 populations, (iii) comparison of change points with intervention dates, (iv) non-mechanistic model,  
322 (v) mechanistic model, and (vi) synthetic controls. We review these in the following.

#### 323 *(i) Description of change over time*

324 The large majority of analyses following a descriptive approach examined the change of the outcome  
325 over time to assess the intervention effect (n=136; 48 %). In some of these analyses, the focus was on  
326 the course of the outcome over time, typically by attributing the observed change (e. g. a reduction  
327 in new cases over time) to the analyzed interventions (n=49; 36 %). For example, the outcome was



328 assessed at regular or irregular intervals, which were not necessarily aligned with the implementation  
329 dates of interventions<sup>41,42,59</sup>. The majority of analyses, however, followed the logic of an interrupted  
330 time series analysis, i.e. the outcome was explicitly compared between time periods before and after  
331 interventions (n=87; 64%)<sup>60-63</sup>.

### 332 (ii) Comparison of populations

333 A few descriptive analyses compared outcomes via summary statistics only between populations  
334 (i.e. without considering variation over time) to assess intervention effects (n=8; 3%). In such  
335 analyses, the outcomes were compared between populations that were stratified by different exposure  
336 to interventions (e.g. populations that implemented a certain intervention and populations that did  
337 not)<sup>64-66</sup>.

### 338 (iii) Comparison of change points with intervention dates

339 Some descriptive analyses checked whether the dates of estimated change points in outcomes and  
340 the implementation dates of interventions coincide (n=7; 2%)<sup>10,22,67</sup>. If both dates were more  
341 or less in agreement, this was taken as evidence confirming the effectiveness of the intervention.  
342 However, change point detection methods could also yield change points prior to the implementation  
343 of interventions, which was sometimes interpreted as a sign of additional factors influencing the  
344 outcome (e.g. proactive social distancing)<sup>22</sup>.

### 345 (iv) Non-mechanistic model

346 Non-mechanistic models are statistical models that typically make *no* explicit assumptions about  
347 the mechanisms that drive infection dynamics. Such models were used in both parametric and  
348 counterfactual approaches (n=78; 27%).

349 In parametric approaches, non-mechanistic models – almost always (generalized) linear regression  
350 models – were used to model a direct link between interventions and outcome. Typically, dummy  
351 variables were used to indicate when (variation over time)<sup>9,68,69</sup> or where (variation between  
352 populations)<sup>70-72</sup> interventions were implemented. Analyses exploiting both variation over time and  
353 between populations typically used panel regression methods<sup>5,73,74</sup>.

354 In counterfactual approaches, the non-mechanistic models used were mostly exponential growth  
355 models, and sometimes time series models (e.g. AR(I)MA or exponential smoothing)<sup>38,44,58</sup>. These

356 models were fitted using data prior to when an intervention was implemented and then extrapolated  
357 the outcome afterwards.

358 *(v) Mechanistic model*

359 Mechanistic models have a structure that makes, to some extent, explicit assumptions about the  
360 mechanisms that drive infection dynamics. They were used in both parametric and counterfactual  
361 approaches (n=43; 15 %).

362 In parametric approaches, the effect of an intervention was represented via a parameter that  
363 was functionally linked to the disease dynamics (i. e. via a latent variable) of the model. This was  
364 typically achieved by parameterizing the transmission rate or reproduction number as a function of  
365 binary variables, indicating whether interventions were implemented or not<sup>2,75-77</sup>. Others linked the  
366 effects of interventions to the contact rate, the transmission probability upon contact, or to entries  
367 in the contact matrix<sup>78-80</sup>. A few modeling approaches also represented the intervention via an  
368 explicit structure or dynamic in the model, e. g. a compartment for quarantined individuals with a  
369 quarantine rate<sup>47,81</sup> or an exponential decay of the susceptible population<sup>46,47,82</sup>.

370 The most popular mechanistic models used in parametric approaches were compartmental trans-  
371 mission models. These models were fitted to the time series of cases, hospitalizations, recovered  
372 cases, deaths, or several simultaneously. With the exception of one meta-population model<sup>83</sup>, all  
373 compartmental models used in analyses following a parametric approach were single-population  
374 models. If multiple populations were analyzed, each population was modeled separately. A few  
375 parametric analyses also used a semi-mechanistic Bayesian transmission model with a time-discrete  
376 renewal process, similar to the one in an early influential paper by Flaxman et al.<sup>8</sup>. These analyses  
377 fitted a Bayesian hierarchical model with stochastic elements for disease transmission and ascertain-  
378 ment on observed time series for cases, deaths, or both<sup>2,8,84</sup>. The model was usually fitted to data  
379 from several populations, modeling separately the time course in each population but estimating  
380 the parameters for the intervention effects jointly across populations. Rarely, analyses used highly  
381 complex models such as individual-based transmission models simulating the behavior of individual  
382 agents, or phylodynamic models inferring both virus phylogenies and transmission dynamics from  
383 genome sequence data.

384 In counterfactual approaches, mechanistic models were, similar to non-mechanistic models,

385 calibrated to data before the implementation of an intervention and then projected the outcome for  
386 the time after the intervention, while keeping the model parameters fixed<sup>85–87</sup>. Thus, no relationship  
387 between intervention and outcome is explicitly modeled. Regularly, these analyses used meta-  
388 population or individual-based models that incorporated migration dynamics through mobility data  
389 and a network between individuals or populations<sup>87–89</sup>.

#### 390 (vi) Synthetic controls

391 Some counterfactual approaches used synthetic control methods (n=6; 2%). Here, a counterfactual  
392 scenario was constructed by computing the counterfactual outcome as a weighted combination of ob-  
393 servations from a pool of “control” populations in which the intervention was not implemented<sup>43,90,91</sup>.  
394 Weights were fitted so as to give more importance to control populations similar to the intervention  
395 population. In these analyses, the course of the outcome before intervention was often used as the  
396 primary measure of similarity<sup>6,90,91</sup>. Sometimes, further factors such as geographic proximity or  
397 population characteristics were also considered<sup>90,92</sup>.

#### 398 **Code availability**

399 For around one in four analyses, a link to a publicly accessible repository containing the computer  
400 code implemented for a specific analysis was provided (n=65; 23%). Overall, the code availability  
401 was comparably higher for parametric approaches, where one in three analyses provided a link.

## 402 **2.5 Effect assessment**

403 The analyses in our review sample varied in their form of effect assessment, i.e. how the effect  
404 was quantified, whether uncertainty was reported, and whether sensitivity analyses or subgroup  
405 assessments were conducted. A systematic classification of the effect assessment is shown in Table 5.

#### 406 **Reporting of intervention effect, measure of effect, and reporting of uncertainty**

407 The are difference in how effects of interventions were reported. Around one in five analyses made  
408 a qualitative assessment of the intervention effect (n=53; 19%), e.g. by qualitatively describing  
409 the change in the outcome over time following the implementation of interventions. More frequent  
410 was the reporting of comparisons of outcome values to assess the intervention effect (n=73; 26%),  
411 i.e. by comparing the outcome values before an intervention with the outcome values after an  
412 intervention. Around half of the analyses reported a quantitative change in outcome values (n=159;

**Table 5.** Systematic classification and frequency of different effect assessments (D.5).

<b>D.5.1: Reporting of intervention effect</b>				Total freq.
QS: Qualitative statement				53 (19 %)
CO: Comparison of outcome values				73 (26 %)
QC: Quantification of change in outcome values				159 (56 %)
<b>D.5.2: Measure of effect<sup>‡</sup></b>	(QS)	(CO)	(QC)	
Change in reproduction number	22	44	29	95 (33 %)
Change in confirmed cases	16	15	38	69 (24 %)
Change in mobility	9	6	28	43 (15 %)
Other	18	29	100	147 (52 %)
<b>D.5.3: Reporting of uncertainty</b>				
Not applicable				52 (18 %)
Yes				154 (54 %)
No				79 (28 %)
<b>D.5.4: Sensitivity analysis (including computed outcomes)</b>				
None (no sensitivity analyses w.r.t effect)				217 (76 %)
One or more sensitivity analyses				68 (24 %)
↳ Model specification varied				↳ 36 (53 %)
↳ Epidemiological parameters varied				↳ 29 (43 %)
↳ Different or modified outcome used				↳ 17 (25 %)
↳ Same analysis with (sub)population excluded				↳ 16 (24 %)
↳ Different coding of interventions used				↳ 10 (15 %)
↳ Start or end date of study period varied				↳ 4 (6 %)
<b>D.5.5: Subgroup assessment</b>				
None (no subgroups)				250 (88 %)
One or more subgroups <sup>‡</sup>				35 (12 %)
↳ Based on socioeconomic indicators				↳ 23 (66 %)
↳ Based on epidemiological indicators				↳ 16 (46 %)
↳ Based on public health response				↳ 9 (26 %)
↳ Based on geographic areas				↳ 6 (17 %)

*Reporting of intervention effect:* (QS) qualitative statement, (CO) comparison of outcome values, and (QC) quantification of change in outcome values

<sup>‡</sup> Multiple categories per analysis are possible. Frequencies refer to number of analyses to which category applies, proportions thus do not sum to 100 %.

413 56 %), e.g. by computing the difference in the outcome values before and after an intervention,  
 414 or estimating the difference via a parameter in a statistical model. The intervention effect was  
 415 oftentimes measured in terms of a change in the reproduction number (n=95; 33 %), in confirmed  
 416 cases (n=69; 24 %), or in mobility (n=43; 15 %), but many other measures of effect were also common.  
 417 Notably, a relevant number of analyses in our review sample focused its methodology on estimating  
 418 the reproduction number as the computed outcome and then conducted only a qualitative assessment  
 419 of the intervention effects afterwards. Uncertainty was reported in around one half of the analyses  
 420 (n=154; 54 %), e.g. via standard error, confidence intervals, and credible intervals.

### 421 ***Sensitivity analyses***

422 We checked all works for sensitivity analyses that were specifically conducted to examine the  
423 robustness of the reported intervention effects. Many studies conducted sensitivity analyses only  
424 related to the predicted outcome or model fit, but not to the intervention effect. Overall, the vast  
425 majority of analyses did not conduct sensitivity analyses with regard to the intervention effect  
426 (n=217; 76 %).

427 Of those that did, sensitivity analyses focused on model extensions or adjustments in which the  
428 model specification was varied (n=36; 53 %), e. g. by changing the structure of a transmission model  
429 or by adjusting the estimated effects of interventions for additional variables in a regression model.  
430 Others analyzed sensitivity with respect to variations in epidemiological parameters (n=29; 43 %),  
431 e. g. by assuming a different basic reproduction number, generation or serial interval, infectious  
432 period, or reporting delay distribution. Only few analyses tested sensitivity with regard to data: i. e.  
433 using different or modified outcomes<sup>69,93</sup> (n=17; 25 %); using a different coding of interventions<sup>3,7</sup>  
434 (n=10; 15 %); or repeating the same analysis but excluding (sub)populations<sup>2</sup> (n=16; 24 %). Of the  
435 different methodological approaches, analyses with semi-mechanistic Bayesian transmission models  
436 generally conducted more comprehensive sensitivity analyses. For example, one work<sup>2</sup> performed  
437 checks regarding data, model, and epidemiological parameters and presented the results prominently  
438 in the main manuscript.

### 439 ***Subgroup assessment***

440 The effects of interventions were rarely assessed within subgroups of the population (n=35; 12 %).  
441 Two thirds of such assessments were within subgroups created based on socioeconomic indicators  
442 (n=23; 66 %), e. g. assessing intervention effects within groups of individuals of different age and  
443 gender<sup>41</sup> or assessing the effect of lockdown within low- and high-income income regions<sup>6</sup>. Less  
444 frequent were subgroups based on epidemiological indicators<sup>5</sup>, the public health response<sup>94</sup>, or  
445 geographic areas<sup>95</sup>.

## 446 **3 Discussion**

447 Our systematic review covers over 240 studies published between January 2020 and January 2021.  
448 Insights from this review can inform different types of future studies: (i) studies using data from

449 the same period that extend our knowledge on aspects that have so far been rarely investigated;  
450 (ii) studies using data from subsequent periods that generate new insights or corroborate existing  
451 ones; and (iii) studies using data from a future pandemic caused by another virus. Although the  
452 preconditions to conduct these studies differ, they share the goals and challenges of the studies in  
453 our review sample. Accordingly, the results from our systematic review allow us to make seven  
454 recommendations for the design of future studies, which are discussed in the following.

455 ***Recommendation 1: Exploring the value of rarely analyzed outcomes***

456 During the COVID-19 pandemic, both surveillance data on confirmed cases, hospitalizations, or  
457 deaths<sup>29,30</sup>, and mobility data from mobile phones<sup>32,33</sup> have become publicly available at scale. This  
458 has enabled a large number of studies assessing the effect of non-pharmaceutical interventions on  
459 population-level epidemiological outcomes and on human mobility (Table 2, D.2.1). However, there  
460 remains considerable potential to explore the value of other outcomes and data sources that have so  
461 far been rarely analyzed.

462 First, more detailed insights into the effects of non-pharmaceutical interventions could be gained  
463 from using individual-level data. This has been demonstrated by studies in our review sample, e. g.  
464 by relating the effects of non-pharmaceutical interventions to the serial interval using symptom onset  
465 data<sup>41</sup>, to transmission chains using contact tracing data<sup>87</sup>, or to virus migration rates using genome  
466 sequence data<sup>20</sup>. We hope to see more such analyses as more individual-level data becomes available.

467 Second, there can be great merit in analyses advancing our understanding of the mechanisms by  
468 which non-pharmaceutical interventions work. Our review sample contained many analyses assessing  
469 the effects of interventions on human behavior, or of human behavior on epidemiological outcomes.  
470 The majority of these analyses used mobility data from mobile phones (Table 2, D.2.1); however, more  
471 effort is needed to obtain a complete picture of how population behavior mediates the effects of non-  
472 pharmaceutical interventions. For example, interventions may influence behaviour and transmission  
473 through further factors not captured by previous studies, and, moreover, the relationship between  
474 mobility and disease transmission may change over time<sup>96,97</sup>. Additional insights can be gained from  
475 analyses using behavioral data from other sources, e. g. surveys evaluating compliance with mask  
476 mandates<sup>62</sup> or the number of daily contacts<sup>98</sup>. Moreover, we see value in analyzing interventions,  
477 behavior and epidemiological outcomes jointly, i. e. in the form of a mediation analysis<sup>99,100</sup>, allowing

478 to differentiate the direct and indirect effect of non-pharmaceutical interventions.

479 ***Recommendation 2: Exploiting variation in the exposure to interventions between populations***

480 Variation in the exposure to interventions (i. e. when, where and which interventions were imple-  
481 mented) is required in order to empirically assess their effects. Variation over time was the most  
482 common choice in our review sample (Table 4, D.4.2). Here, the effects were previously assessed  
483 within populations based on differences in outcomes before and after non-pharmaceutical interven-  
484 tions were implemented. However, such changes may falsely be attributed to non-pharmaceutical  
485 interventions if they are subject to confounding by concurring time trends. We thus recommend to  
486 also exploit exposure variation between populations, i. e. with respect to the timing and the types of  
487 single interventions that were implemented. This was done by only one in five analyses in our review  
488 sample (Table 4, D.4.2), although the types and timing of interventions varied considerably between  
489 the populations. This leaves a valuable source of variation largely untapped.

490 ***Recommendation 3: Dissecting the effects of single interventions***

491 Evidenced-based decision making requires empirical estimates for the effects of single non-  
492 pharmaceutical interventions (e. g. school closures or stay-at-home orders). However, the majority of  
493 analyses assessed the effects of population-specific combinations of interventions such as lockdowns  
494 (Table 3, D.3.3). The underlying analyses typically studied only a single population (or multiple  
495 populations separately) where multiple interventions were implemented on the same day, and, as a  
496 result, the separate effects of interventions cannot be disentangled. For future work, we recommend  
497 more effort to conduct analyses across multiple populations, so that the separate effects of single  
498 interventions can be dissected.

499 ***Recommendation 4: Careful coding of interventions***

500 Systematic coding of intervention data is necessary to compare the effects of non-pharmaceutical  
501 interventions across multiple populations (Table 3, D.3.5). However, the process of such systematic  
502 coding involves subjective decisions regarding if and when non-pharmaceutical interventions have  
503 been implemented<sup>2</sup>. These decisions are sometimes unavoidable, and they are also inherent to the  
504 development of public databases<sup>101</sup>. As a result, different coding of interventions could impact  
505 the results, and, therefore, coding decisions should be made transparent and documented carefully.  
506 In addition, we recommend to examine empirically the impact when using a different coding of

507 interventions, e. g. by comparing the results across different input data.

508 ***Recommendation 5: Promoting comparability across different analyses***

509 During a pandemic, public health policy has a strong focus on the number of confirmed cases, hospi-  
510 talizations, and deaths, making them obvious outcomes to evaluate the effectiveness of interventions.  
511 However, non-pharmaceutical interventions act only indirectly or with a certain delay on these  
512 observable outcomes. Typically, non-pharmaceutical interventions should influence the behaviour  
513 of the population, which should reduce transmission (e. g. by limiting the contact rate), which in  
514 turn should affect the number of new infections and, subsequently, observed outcomes like confirmed  
515 cases, hospitalizations, or deaths. The question of how to assess intervention effects along this path  
516 has been answered differently by the studies in our review sample. We identified four main types of  
517 analyses; see (1)–(4) in Box 1. In the following, we discuss the different types with regard to their  
518 ability of enabling a comparison of results between studies.



Box 1. Different types of analyses to assess the effects of non-pharmaceutical interventions.

**(1) Observed outcome directly linked to interventions**

A raw, observed outcome is analyzed directly by evaluating differences (i) over time with an interrupted time-series analysis comparing the outcome before vs. after an intervention, (ii) between populations with a cross-sectional analysis comparing populations exposed vs. not exposed to an intervention, or (iii) both over time and between populations with a panel data analysis. Mechanistic modeling is typically not involved in this type of analysis, with one exception, namely counterfactual approaches using a transmission model to project the observed outcome after intervention.

**(2) Computed, unobserved outcome linked to interventions**

In contrast to type (1), the intervention effect is measured in terms of an unobserved outcome. This is computed from the raw outcome and then analyzed in a similar manner as in (1). Mechanistic modeling can be involved in computing the unobserved outcome, for example by using a model to estimate the reproduction number or transmission rate from the number of new cases.

**(3) Observed outcome linked to interventions via unobserved outcome in mechanistic model**

Observed outcomes are used to fit a mechanistic model (e.g. compartmental transmission model) that includes a latent variable representing an unobserved outcome (e.g. the reproduction number), which in turn is parameterized as a function of interventions. For instance, a regression-like link is used within the mechanistic model to estimate the effect of interventions on the transmission rate as a latent variable.

**(4) Change points in outcome related to exposure**

Change points are estimated in the time series of an observed or unobserved outcome. The estimated change points are then related to the implementation dates of interventions. If the estimated change points agree well with the actual implementation dates of interventions, this is interpreted as evidence for the effectiveness of interventions.

520 Analyses of type (1), where the observed outcome is directly linked to interventions, can avoid  
521 mechanistic modeling by directly analyzing an observed outcome such as cases or deaths. Here, a  
522 central challenge is to take into account the uncertain delay between the implementation of non-  
523 pharmaceutical interventions and their effects on the observable outcome. The fact that infections  
524 and subsequent outcomes such as confirmed cases follow exponential dynamics during an epidemic  
525 wave makes it difficult to compare intervention effects measured by observable outcomes across  
526 different epidemic phases. In contrast to that, analyses following type (2) or (3) employ mechanistic  
527 modeling, allowing to assess the effects of interventions on latent, unobservable outcomes such as  
528 the transmission rate or the reproduction number. Since these latent outcomes can be inferred from  
529 different observed outcomes like cases or deaths, it becomes possible to compare analyses that use  
530 different raw data. The difference between type (2) and (3) is that for (2) the estimation of the latent  
531 outcome is separated from the effect assessment. Such separation reduces model complexity, however,  
532 often at the expense of incomplete uncertainty assessments. The reason is that most analyses based  
533 on type (2) only incorporated uncertainty involved in modeling the effect of interventions on the  
534 latent outcome, thus leaving out uncertainty involved in the computation of the latent outcome.  
535 This also means that differences in uncertainty assessments must be carefully taken into account  
536 when synthesizing results from multiple studies.

537 Finally, analyses following type (4), where change points in the outcome are related to the  
538 exposure, take a very different approach that shares few assumptions with the other approaches. A  
539 comparison of change points can verify the presence of an effect, yet without quantifying its size. As  
540 a result, such findings are best complemented with an analysis of type (1), (2), or (3).

#### 541 ***Recommendation 6: Understanding variation in effectiveness across subgroups***

542 The effects of non-pharmaceutical interventions may vary between and within populations. However,  
543 only one in ten analyses in our review sample examined variation across subgroups or populations  
544 (Table 4, D.4.2). Estimating and explaining such variation could help understand the conditions  
545 under which interventions are more or less effective for a specific subgroup, potentially allowing  
546 policy makers to tailor interventions to a specific subgroup or setting. Our review points out  
547 two approaches that can contribute towards a better understanding of the varying effectiveness  
548 of non-pharmaceutical interventions: (i) comparing the effects of interventions between subgroups

549 of the same population (e. g. between the young and elderly population); and (ii) comparing the  
550 effects of interventions between different populations and relating differences to population-specific  
551 characteristics (e. g. population density).

#### 552 ***Recommendation 7: Assessing sensitivity within and between studies***

553 While variation in methodologies can complicate the comparability among studies, it may help to  
554 identify the influence of certain methodological choices on the results. Here, the public availability of  
555 data for outcomes and interventions holds potential for sensitivity analyses within studies as well as  
556 comparisons between studies. Specifically, the same analysis could be repeated with different sets of  
557 publicly available data as part of the same study. This way, sensitivity of the findings with respect to  
558 the choice of outcome and intervention data could be assessed within studies, reducing the risk of bias  
559 from specific outcome data (e. g. incomplete case ascertainment due to limited testing capacity etc.)  
560 or the specific coding of interventions. For example, the number of new cases, deaths, or both could  
561 be used as the raw outcome in mechanistic models with a comparable latent outcome<sup>2</sup>. However,  
562 other aspects, in particular the specific setting and methodologies used, are presumably more difficult  
563 to vary as part of a sensitivity analysis, and may therefore need to be compared between different  
564 studies. Important for such comparisons is giving access to the preprocessed data, yet most study  
565 authors accessed these data directly from national authorities or external data providers without  
566 making the preprocessed data additionally available (Table 2, D.2.5-D.2.6 and Table 3, D.3.6-D.3.7).  
567 To promote reproducibility, study authors should always make the preprocessed data available for  
568 two reasons. First, future access to the data via the original source may be restricted (this has  
569 already been observed for corporate mobility data). Second, originally analyzed and recently accessed  
570 data may be different as publicly available epidemiological data and coded intervention data can be  
571 subject to updates, revisions, and corrections. In addition, publication of computed code is crucial  
572 to support comparisons of methodological approaches, specific modeling choices, and input data.

## 573 **4 Conclusions**

574 Our review of more than 240 studies on the effects of non-pharmaceutical interventions revealed  
575 substantial variation in methodologies. Until specific best practices emerge, further heterogeneity in  
576 studies is inevitable and can also be beneficial, e. g. for assessing robustness of the results with respect

577 to method and input data. Nevertheless, some standardization is required in order to synthesize  
578 evidence on the effects of non-pharmaceutical interventions from multiple studies. So far, a lack of  
579 common standards and substantial variation in the methodologies used have created a challenge  
580 for meta-analyses to summarize and compare the reported effects from existing studies<sup>11–15</sup>. Here,  
581 our methodology review can serve as a basis for subsequent meta-analyses to factor in the variety  
582 of existing methodologies when pooling and comparing the large number of effects that have been  
583 reported for non-pharmaceutical interventions on health-related outcomes. More importantly though,  
584 our review promotes common standards and reduces barriers to comparability across studies by  
585 making recommendations for the design of future studies.

586 A general limitation of the studies included in our review is that it is difficult to estimate causal  
587 effects based on population-level observational data and to rule out unobserved confounding. This  
588 is also highlighted by the persistent debate on the role of voluntary behavioral change in curbing  
589 transmission<sup>102–105</sup>. Evidence from observational studies on the effects of non-pharmaceutical  
590 interventions should thus be evaluated in conjunction with evidence from other fields. For example,  
591 evidence on the infectiousness of school children and parental strategies to fill the care gap can  
592 produce independent predictions about the effects of school closures. Similarly, laboratory evidence  
593 regarding the effectiveness of masks together with evidence on compliance with masks can produce  
594 independent predictions of the effectiveness of mask mandates.

595 During the COVID-19 pandemic, a tremendous amount of publicly available epidemiological  
596 data has been generated. The ease of access to this data allowed many researchers to contribute  
597 work, using a variety of methodologies to assess the effects of non-pharmaceutical interventions  
598 on health-related outcomes. With researchers from diverse fields contributing, there is a unique  
599 opportunity to benefit from the various inputs in developing a methodological foundation for timely  
600 and robust assessments during future pandemics. This will however require a thorough examination  
601 of the present methodologies in order to share lessons learnt and develop best practices. Our  
602 systematic review can be viewed as a first such attempt.

## 603 **5 Materials and methods**

604 We tailored our review to the challenge of mapping a potentially diverse set methodologies from  
605 a large number of studies. To ensure rigour and consistency, we preregistered the procedures for  
606 all stages of the review process, following common guidelines for systematic literature reviews<sup>17</sup>.  
607 Certain guidelines were not applicable to a methodology review as ours. In particular, our eligibility  
608 criteria and risk of bias assessment reflect the objective of this review, which was not to evaluate the  
609 effects of non-pharmaceutical interventions, but to map the variation in methodologies used. The  
610 preregistered methodology was documented in a review protocol at PROSPERO<sup>19</sup>.

611 We report our review according to the PRISMA 2020 statement<sup>18</sup>. A completed PRISMA 2020  
612 checklist is provided in SI Appendix G. We conceptualized the review by drawing on experience from  
613 our own primary research in the field<sup>4,97</sup>. The search strategy was developed jointly and executed by  
614 an experienced information consultant. Then, two authors (NB and AL) performed study selection,  
615 data extraction, and synthesis, while having regular meetings with the complete author team.

### 616 **5.1 Eligibility criteria for studies**

617 In the following, we describe our eligibility criteria, which informed our search strategy and were  
618 systematically applied during study selection. Importantly, if a study contained multiple analyses of  
619 which only some fulfilled our eligibility criteria, we included the study but extracted only the eligible  
620 analyses. This may sometimes not correspond to the main analysis of a paper or may include more  
621 than one analysis per study.

#### 622 ***Study design***

623 In this review, we considered observational studies assessing the effects of non-pharmaceutical  
624 interventions on outcomes related to the COVID-19 disease. We focused on retrospective analyses  
625 that used real-world observational data to assess the effects of non-pharmaceutical interventions.  
626 Specifically, we excluded modeling studies that predominantly worked with synthetic data or  
627 projected future transmission dynamics based on hypothetical scenarios without assessing the effects  
628 of interventions empirically.

629 **Population**

630 We considered studies assessing the effects of non-pharmaceutical interventions on the population in  
631 one or several geographic regions. Our review was not limited to a specific geographic region, i. e.  
632 all national and subnational regions worldwide were considered. We furthermore included studies  
633 analyzing specific subpopulations in a certain region (e. g. certain age groups). We also considered  
634 analyses using individual-level data, as long as the intervention effect was assessed on a population  
635 level.

636 **Outcome**

637 The main outcomes considered were health-related outcomes at population level that are associated  
638 with COVID-19 (e. g. confirmed cases, hospitalizations, and deaths), and epidemiological outcomes  
639 characterizing infection dynamics such as reproduction numbers or transmission rates. We also  
640 considered similar outcomes associated with other infectious diseases (e. g. influenza), if used as  
641 a surrogate for COVID-19. Moreover, behavioral outcomes potentially mediating the effect of  
642 non-pharmaceutical interventions were included (e. g. human mobility). In contrast, we excluded  
643 analyses assessing the effects of non-pharmaceutical interventions solely on other outcomes not  
644 directly related to infectious diseases (e. g. psychological well-being or economic activities).

645 **Intervention**

646 As non-pharmaceutical interventions, we considered the implementation of health policy measures in  
647 the context of the COVID-19 pandemic. Specifically, we included any intervention related to social  
648 distancing (e. g. school closures, venue closures, workplace closures), containment (e. g. contact  
649 tracing, quarantining), population flow (e. g. border closures), or personal protection (e. g. facial  
650 mask mandates). Analyses were considered regardless of whether they assessed the effect of a  
651 single intervention, the effects of multiple interventions separately, or the effect of a combination  
652 of interventions. For simplicity, we refer to these as non-pharmaceutical interventions throughout  
653 the review, while recognizing that also other terms have been used in the literature. Importantly,  
654 we accounted for various alternative terms in our literature search (see [Search strategy](#) below). We  
655 excluded interventions not directly related to disease control (e. g. economic measures like social  
656 benefits).

## 657 **5.2 Search strategy**

658 Our search strategy was informed by our personal experience with studies in the field from our own  
659 primary research on the impact of non-pharmaceutical interventions. The search was supported by  
660 an experienced information consultant.

661 We searched for peer-reviewed original research articles in English language that were accepted,  
662 published, or in press between January 1, 2020 and January 12, 2021. In our review protocol,  
663 we specified that we would also include preprints in our search. However, due to their enormous  
664 volume, we eventually decided not to consider gray literature or preprints in our review. Our results  
665 therefore only cover methodologies used by articles peer-reviewed at the time of search, among  
666 which we already found considerable variation. To account for potentially new methodologies in  
667 articles published after the time of search, we also considered further recent studies on the effects  
668 of non-pharmaceutical interventions in our discussion and put them into the context of our review  
669 findings.

670 We searched the databases Embase, PubMed, Scopus, and Web of Science. These databases  
671 include, among others, MEDLINE, Biological Abstracts, CAB Abstracts, and Global Health. We  
672 composed our search query of four components to be contained in the publication title or abstract:  
673 (1) a synonym for “non-pharmaceutical intervention”, (2) a synonym for “estimation” or “assessment”,  
674 (3) a synonym for “effect”, and (4) a synonym for “COVID-19”. Starting from a precompiled list  
675 of 18 references based on our primary research on the effects of non-pharmaceutical interventions,  
676 we created and repeatedly extended a collection of synonyms for each of the above components,  
677 thereby achieving a broad search while keeping the number of selected studies manageable. The  
678 strings for our search queries are provided in SI Appendix B. Importantly, we decided not to include  
679 search terms for single interventions such as face masks or travel restrictions, as this would have  
680 resulted in an unmanageable number of studies that were not concerned with the population-level  
681 impact of the non-pharmaceutical intervention. Nevertheless, our search query found studies on  
682 single interventions through other terms describing non-pharmaceutical interventions.

### 683 **5.3 Data collection and analysis**

#### 684 ***Study selection***

685 As a first step, we screened the titles of the studies retrieved from the database search for keywords  
686 clearly suggesting that the study would not meet our predefined eligibility criteria (e. g. “mental  
687 health” or “air quality”). The compiled set of keywords (see SI Appendix B) was used to automatically  
688 identify cases for exclusion via the publication title. For all remaining studies, two authors (AL and  
689 NB) checked the eligibility criteria and individually decided on inclusion or exclusion. Each of the  
690 two authors checked the eligibility for one half of the studies via the following process: First, studies  
691 were checked by title and, if in doubt, by abstract. Then, if still in doubt, studies were checked by  
692 full text and discussed by both authors. Any disagreements were resolved with involvement of a  
693 third author (WV). Generally, we followed an inclusive approach by keeping all studies that could  
694 not be excluded with high confidence. At each stage, all decisions were recorded in a spreadsheet.

#### 695 ***Data extraction***

696 We extracted data from all included studies in a spreadsheet. Our extraction strategy reflected the  
697 exploratory nature of our analysis and thus allowed for new data items to be added throughout  
698 the process. Therefore, we maintained a detailed manual with all data items and the potential  
699 values for each item (see SI Appendix D). Before extraction, a preliminary version of the data  
700 extraction form was created based on reporting items from checklists for observational studies  
701 (STROBE<sup>106</sup>, RECORD<sup>107</sup>), a template for public health policy interventions (TIDieR-PHP<sup>108</sup>),  
702 and our personal experience with studies in the field from our own primary research on the impact  
703 of non-pharmaceutical interventions. Aside from bibliographic information, the data to be extracted  
704 consisted of information on the study setting, outcome, intervention, methodological approach, and  
705 effect assessment.

706 The extraction process was structured in four rounds. During the first round, two authors (AL  
707 and NB) extracted data from an initial set of 20 publications, blinded to each other’s coding. The  
708 coding was then compared, and any differences were discussed to resolve ambiguities. Corresponding  
709 changes were recorded by updating the extraction form and manual, and applied subsequently. In  
710 the second round, the two reviewers each extracted data from one half of the remaining publications  
711 and checked the other half coded by their colleague. Color-coding was used to highlight uncertain



712 or ambiguous entries for the other reviewer or to mark such entries for further discussion. Regular  
713 meetings were held between the two authors (AL and NB) to discuss these uncertainties and  
714 ambiguities. All disagreements were resolved through discussion, if needed by involving a third  
715 author (WV). Thereby, the data extraction manual and form were continuously refined and kept  
716 up-to-date. In particular, the list of values that could be potentially assigned to each data item was  
717 continuously extended and harmonized as new studies were extracted. In the third round, the data  
718 extraction form and manual were simplified by merging data items or categories that, retrospectively,  
719 were found redundant, or by relabeling items and categories to define them more precisely. This  
720 was done with particular attention to enable comparability among the extracted analyses as well as  
721 readability of the results. In the fourth round, the final scheme was applied to all studies.

### 722 ***Quality assessment***

723 The goal of this systematic review was not to perform a meta-analysis or narrative synthesis of the  
724 effects of non-pharmaceutical interventions, but to compare the included studies along methodological  
725 dimensions and to analyze the variation in study setting, outcome, intervention, methodological  
726 approach, and effect assessment. Therefore, no risk of bias assessment with regard to the study  
727 results was conducted. Our minimum requirement for quality was that most information on the  
728 aforementioned dimensions could be extracted from the manuscript and/or supplementary material.  
729 This minimum requirement was not met by four studies, which were thus excluded. For other studies  
730 where only some methodological information was missing, we noted in the data extraction sheet  
731 that this information “could not be evaluated”.

### 732 ***Data synthesis***

733 The results of the data extraction were synthesized in tabular form by recording the frequency of  
734 categories per item. We reported the frequency for each item of the main dimensions (study setting,  
735 outcome, intervention, methodological approach, and effect assessment) individually. For some items,  
736 we conducted further specialized analyses, for example by computing the frequencies of categories  
737 for different methodologies separately, or by qualitatively evaluating the supplementary information  
738 added to certain entries during extraction. Insights from these additional analyses were reported  
739 textually. Furthermore, we synthesized common analysis types based on patterns identified in the  
740 methodological approaches. Lastly, based on our findings, we derived specific recommendations for

741 future studies with regard to scope, robustness, and comparability, and put them into the context of  
742 more recent studies that were not part of our review sample.

## 743 **References**

- 744 **1.** Hale T, Angrist N, Goldszmidt R, Kira B, Petherick A, Phillips T, et al. A Global Panel  
745 Database of Pandemic Policies (Oxford COVID-19 Government Response Tracker). *Nature*  
746 *Human Behaviour*. 2021;5(4):529–538.
- 747 **2.** Brauner JM, Mindermann S, Sharma M, Johnston D, Salvatier J, Gavenčiak T, et al. Inferring the  
748 Effectiveness of Government Interventions against COVID-19. *Science*. 2021;371(6531):eabd9338.
- 749 **3.** Haug N, Geyrhofer L, Londei A, Dervic E, Desvars-Larrive A, Loreto V, et al. Ranking the  
750 Effectiveness of Worldwide COVID-19 Government Interventions. *Nature Human Behaviour*.  
751 2020;4(12):1303–1312.
- 752 **4.** Banholzer N, van Weenen E, Lison A, Cenedese A, Seeliger A, Kratzwald B, et al. Estimating the  
753 Effects of Non-Pharmaceutical Interventions on the Number of New Infections with COVID-19  
754 during the First Epidemic Wave. *PLOS ONE*. 2021;16(6):e0252827.
- 755 **5.** Auger KA, Shah SS, Richardson T, Hartley D, Hall M, Warniment A, et al. Association  
756 between Statewide School Closure and COVID-19 Incidence and Mortality in the US. *JAMA*.  
757 2020;324(9):859–870.
- 758 **6.** Bennett M. All Things Equal? Heterogeneity in Policy Effectiveness against COVID-19 Spread  
759 in Chile. *World Development*. 2021;137:105208.
- 760 **7.** Courtemanche C, Garuccio J, Le A, Pinkston J, Yelowitz A. Strong Social Distancing Measures  
761 in the United States Reduced the COVID-19 Growth Rate. *Health Affairs*. 2020;39(7):1237–1246.
- 762 **8.** Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, et al. Estimating the Effects  
763 of Non-Pharmaceutical Interventions on COVID-19 in Europe. *Nature*. 2020;584(7820):257–261.
- 764 **9.** Hsiang S, Allen D, Annan-Phan S, Bell K, Bolliger I, Chong T, et al. The Effect of Large-Scale  
765 Anti-Contagion Policies on the COVID-19 Pandemic. *Nature*. 2020;584(7820):262–267.
- 766 **10.** Lemaitre JC, Perez-Saez J, Azman AS, Rinaldo A, Fellay J. Assessing the Impact of Non-  
767 Pharmaceutical Interventions on SARS-CoV-2 Transmission in Switzerland. *Swiss Medical Weekly*.  
768 2020;150:w20295.

- 769 **11.** Mendez-Brito A, Bcheraoui CE, Pozo-Martin F. Systematic Review of Empirical Studies  
770 Comparing the Effectiveness of Non-Pharmaceutical Interventions against COVID-19. *Journal of*  
771 *Infection*. 2021;83(3):281–293.
- 772 **12.** Poeschl J, Larsen RB. How Do Non- Pharmaceutical Interventions Affect the Spread of  
773 COVID-19? A Literature Review. *Danmarks Nationalbank*; 2021. 4.
- 774 **13.** Rizvi RF, Craig KJT, Hekmat R, Reyes F, South B, Rosario B, et al. Effectiveness of Non-  
775 Pharmaceutical Interventions Related to Social Distancing on Respiratory Viral Infectious Disease  
776 Outcomes: A Rapid Evidence-Based Review and Meta-Analysis. *SAGE Open Medicine*. 2021;9.
- 777 **14.** Iezadi S, Gholipour K, Azami-Aghdash S, Ghiasi A, Rezapour A, Pourasghari H, et al. Effec-  
778 tiveness of Non-Pharmaceutical Public Health Interventions against COVID-19: A Systematic  
779 Review and Meta-Analysis. *PLOS ONE*. 2021;16(11):e0260371.
- 780 **15.** Talic S, Shah S, Wild H, Gasevic D, Maharaj A, Ademi Z, et al. Effectiveness of Public Health  
781 Measures in Reducing the Incidence of COVID-19, SARS-CoV-2 Transmission, and Covid-19  
782 Mortality: Systematic Review and Meta-Analysis. *BMJ*. 2021;375:e068302.
- 783 **16.** Perra N. Non-Pharmaceutical Interventions during the COVID-19 Pandemic: A Review. *Physics*  
784 *Reports*. 2021;913:1–52.
- 785 **17.** Higgins J, Thomas J, Chandler J, Cumpston M, Li T, Page M, et al. *Cochrane Handbook for*  
786 *Systematic Reviews of Interventions*. Cochrane; 2021.
- 787 **18.** Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD, et al. The PRISMA  
788 2020 Statement: An Updated Guideline for Reporting Systematic Reviews. *BMJ*. 2021;372:n71.
- 789 **19.** Banholzer N, Lison A, Özcelik D, Feuerriegel S, Vach W. A Comparison of Studies Estimating the  
790 Effectiveness of Non-Pharmaceutical Interventions: A Systematic Review Protocol. *PROSPERO*;  
791 2021.
- 792 **20.** Candido DS, Claro IM, de Jesus JG, Souza WM, Moreira FRR, Dellicour S, et al. Evolution  
793 and Epidemic Spread of SARS-CoV-2 in Brazil. *Science*. 2020;369(6508):1255–1260.
- 794 **21.** Moreno GK, Braun KM, Riemersma KK, Martin MA, Halfmann PJ, Crooks CM, et al. Re-  
795 vealing Fine-Scale Spatiotemporal Differences in SARS-CoV-2 Introduction and Spread. *Nature*  
796 *Communications*. 2020;11(1):5558.

- 797 **22.** Wieland T. A Phenomenological Approach to Assessing the Effectiveness of COVID-19 Related  
798 Nonpharmaceutical Interventions in Germany. *Safety Science*. 2020;131:104924.
- 799 **23.** Karnakov P, Arampatzis G, Kii I, Wermelinger F, Wlchli D, Papadimitriou C, et al. Data-  
800 Driven Inference of the Reproduction Number for COVID-19 before and after Interventions for  
801 51 European Countries. *Swiss Medical Weekly*. 2020;150:w20313.
- 802 **24.** Cori A, Ferguson NM, Fraser C, Cauchemez S. A New Framework and Software to Estimate  
803 Time-Varying Reproduction Numbers During Epidemics. *American Journal of Epidemiology*.  
804 2013;178(9):1505–1512.
- 805 **25.** Cori A. EpiEstim: Estimate Time Varying Reproduction Numbers from Epidemic Curves; 2021.
- 806 **26.** Diekmann O, Heesterbeek JAP, Metz JAJ. On the Definition and the Computation of the Basic  
807 Reproduction Ratio  $R_0$  in Models for Infectious Diseases in Heterogeneous Populations. *Journal*  
808 *of Mathematical Biology*. 1990;28(4):365–382.
- 809 **27.** Wallinga J, Teunis P. Different Epidemic Curves for Severe Acute Respiratory Syndrome Reveal  
810 Similar Impacts of Control Measures. *American Journal of Epidemiology*. 2004;160(6):509–516.
- 811 **28.** Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the Effect of Individual  
812 Variation on Disease Emergence. *Nature*. 2005;438(7066):355–359.
- 813 **29.** European Centre for Disease Prevention and Control. COVID-19 Datasets; 2022.  
814 <https://www.ecdc.europa.eu/en/covid-19/data>.
- 815 **30.** Johns Hopkins University & Medicine. Coronavirus Resource Center; 2022.  
816 <https://coronavirus.jhu.edu/>.
- 817 **31.** Worldometer. Coronavirus Statistics; 2022. <https://www.worldometers.info/coronavirus/>.
- 818 **32.** Google. COVID-19 Community Mobility Reports; 2022.  
819 <https://www.google.com/covid19/mobility/>.
- 820 **33.** Apple. COVID-19 Mobility Trends Reports; 2022. <https://covid19.apple.com/mobility>.
- 821 **34.** China Data Lab. Baidu Mobility Data; 2021. <https://doi.org/10.7910/DVN/FAEZIO>.
- 822 **35.** Kucharski AJ, Russell TW, Diamond C, Liu Y, Edmunds J, Funk S, et al. Early Dynamics of  
823 Transmission and Control of COVID-19: A Mathematical Modelling Study. *The Lancet Infectious*  
824 *Diseases*. 2020;20(5):553–558.

- 825 **36.** Couzin-Frankel J, Vogel G. School Openings across Globe Suggest Ways to Keep Coronavirus  
826 at Bay, despite Outbreaks. *Science*; 2021.
- 827 **37.** Berry CR, Fowler A, Glazer T, Handel-Meyer S, MacMillen A. Evaluating the Effects of  
828 Shelter-in-Place Policies during the COVID-19 Pandemic. *Proceedings of the National Academy  
829 of Sciences*. 2021;118(15):e2019706118.
- 830 **38.** Bönisch S, Wegscheider K, Krause L, Sehner S, Wiegel S, Zapf A, et al. Effects of Coronavirus  
831 Disease (COVID-19) Related Contact Restrictions in Germany, March to May 2020, on the  
832 Mobility and Relation to Infection Patterns. *Frontiers in Public Health*. 2020;8:568287.
- 833 **39.** Kraemer MUG, Yang CH, Gutierrez B, Wu CH, Klein B, Pigott DM, et al. The Effect of Human  
834 Mobility and Control Measures on the COVID-19 Epidemic in China. *Science*. 2020;368(6490):493–  
835 497.
- 836 **40.** Salvatore M, Basu D, Ray D, Kleinsasser M, Purkayastha S, Bhattacharyya R, et al. Com-  
837 prehensive Public Health Evaluation of Lockdown as a Non-Pharmaceutical Intervention on  
838 COVID-19 Spread in India: National Trends Masking State-Level Variations. *BMJ Open*.  
839 2020;10(12):e041778.
- 840 **41.** Ali ST, Wang L, Lau EHY, Xu XK, Du Z, Wu Y, et al. Serial Interval of SARS-CoV-2 Was  
841 Shortened over Time by Nonpharmaceutical Interventions. *Science*. 2020;369(6507):1106–1109.
- 842 **42.** Price DJ, Shearer FM, Meehan MT, McBryde E, Moss R, Golding N, et al. Early Analysis of  
843 the Australian COVID-19 Epidemic. *eLife*. 2020;9:e58785.
- 844 **43.** Huber M, Langen H. Timing Matters: The Impact of Response Measures on COVID-19-related  
845 Hospitalization and Death Rates in Germany and Switzerland. *Swiss Journal of Economics and  
846 Statistics*. 2020;156(1):1–19.
- 847 **44.** Pullano G, Valdano E, Scarpa N, Rubrichi S, Colizza V. Evaluating the Effect of Demographic  
848 Factors, Socioeconomic Factors, and Risk Aversion on Mobility during the COVID-19 Epidemic in  
849 France under Lockdown: A Population-Based Study. *The Lancet Digital Health*. 2020;2(12):e638–  
850 e649.
- 851 **45.** Jefferies S, French N, Gilkison C, Graham G, Hope V, Marshall J, et al. COVID-19 in New  
852 Zealand and the Impact of the National Response: A Descriptive Epidemiological Study. *The*

- 853 Lancet Public Health. 2020;5(11):e612–e623.
- 854 **46.** Maier BF, Brockmann D. Effective Containment Explains Subexponential Growth in Recent  
855 Confirmed COVID-19 Cases in China. *Science*. 2020;368(6492):742–746.
- 856 **47.** Collins OC, Duffy KJ. Estimating the Impact of Lock-down, Quarantine and Sensitization in a  
857 COVID-19 Outbreak: Lessons from the COVID-19 Outbreak in China. *PeerJ*. 2020;8:e9933.
- 858 **48.** Braithwaite J, Tran Y, Ellis LA, Westbrook J. The 40 Health Systems, COVID-19 (40HS, C-19)  
859 Study. *International Journal for Quality in Health Care*. 2020;33(1):mzaa113.
- 860 **49.** Koh WC, Naing L, Wong J. Estimating the Impact of Physical Distancing Measures in  
861 Containing COVID-19: An Empirical Analysis. *International Journal of Infectious Diseases*.  
862 2020;100:42–49.
- 863 **50.** Gupta M, Mohanta SS, Rao A, Parameswaran GG, Agarwal M, Arora M, et al. Transmission  
864 Dynamics of the COVID-19 Epidemic in India and Modeling Optimal Lockdown Exit Strategies.  
865 *International Journal of Infectious Diseases*. 2021;103:579–589.
- 866 **51.** New York Times. See Reopening Plans and Mask Manadates for All 50 States; 2021.  
867 <https://www.nytimes.com/interactive/2020/us/states-reopen-map-coronavirus.html>.
- 868 **52.** McGrail DJ, Dai J, McAndrews KM, Kalluri R. Enacting National Social Distancing  
869 Policies Corresponds with Dramatic Reduction in COVID19 Infection Rates. *PLOS ONE*.  
870 2020;15(7):e0236619.
- 871 **53.** Scarabel F, Pellis L, Bragazzi NL, Wu J. Canada Needs to Rapidly Escalate Public Health  
872 Interventions for Its COVID-19 Mitigation Strategies. *Infectious Disease Modelling*. 2020;5:316–  
873 322.
- 874 **54.** Guirao A. The COVID-19 Outbreak in Spain. A Simple Dynamics Model, Some Lessons, and a  
875 Theoretical Framework for Control Response. *Infectious Disease Modelling*. 2020;5:652–669.
- 876 **55.** Krishna MV. Mathematical Modelling on Diffusion and Control of COVID–19. *Infectious*  
877 *Disease Modelling*. 2020;5:588–597.
- 878 **56.** Zhang B, Zhou H, Zhou F. Study on SARS-CoV-2 Transmission and the Effects of Control  
879 Measures in China. *PLOS ONE*. 2020;15(11):e0242649.

- 880 **57.** Sebastiani G, Massa M, Riboli E. COVID-19 Epidemic in Italy: Evolution, Projections and  
881 Impact of Government Measures. *European Journal of Epidemiology*. 2020;35(4):341–345.
- 882 **58.** Valencia M, Becerra JE, Reyes JC, Castro KG. Assessment of Early Mitigation Measures  
883 against COVID-19 in Puerto Rico: March 15-May 15, 2020. *PLOS ONE*. 2020;15(10):e0240013.
- 884 **59.** Riccardo F, Ajelli M, Andrianou XD, Bella A, Manso MD, Fabiani M, et al. Epidemiological  
885 Characteristics of COVID-19 Cases and Estimates of the Reproductive Numbers 1 Month into  
886 the Epidemic, Italy, 28 January to 31 March 2020. *Eurosurveillance*. 2020;25(49):2000790.
- 887 **60.** Gao S, Rao J, Kang Y, Liang Y, Kruse J, Dopfer D, et al. Association of Mobile Phone Location  
888 Data Indications of Travel and Stay-at-Home Mandates with COVID-19 Infection Rates in the  
889 US. *JAMA Network Open*. 2020;3(9):e2020485.
- 890 **61.** Lurie MN, Silva J, Yorlets RR, Tao J, Chan PA. Coronavirus Disease 2019 Epidemic Doubling  
891 Time in the United States before and during Stay-at-Home Restrictions. *The Journal of Infectious  
892 Diseases*. 2020;222(10):1601–1606.
- 893 **62.** Jarvis CI, Zandvoort aKV, Gimma A, Prem K, Klepac P, Rubin GJ, et al. Quantifying the  
894 Impact of Physical Distance Measures on the Transmission of COVID-19 in the UK. *BMC  
895 Medicine*. 2020;18(1).
- 896 **63.** Ng Y, Li Z, Chua YX, Chaw WL, Zhao Z, Er B, et al. Evaluation of the Effectiveness of  
897 Surveillance and Containment Measures for the First 100 Patients with COVID-19 in Singapore  
898 — January 2–February 29, 2020. *Morbidity and Mortality Weekly Report*. 2020;69(11):307–311.
- 899 **64.** Cobb JS, Seale MA. Examining the Effect of Social Distancing on the Compound Growth Rate  
900 of COVID-19 at the County Level (United States) Using Statistical Analyses and a Random  
901 Forest Machine Learning Model. *Public Health*. 2020;185:27–29.
- 902 **65.** Jardine R, Wright J, Samad Z, Bhutta ZA. Analysis of COVID-19 Burden, Epidemiology and  
903 Mitigation Strategies in Muslim Majority Countries. *Eastern Mediterranean Health Journal*.  
904 2020;26(10):1173–1183.
- 905 **66.** Murillo-Zamora E, Guzmán-Esquivel J, Sánchez-Piña RA, Cedeño-Laurent G, Delgado-Enciso  
906 I, Mendoza-Cano O. Physical Distancing Reduced the Incidence of Influenza and Supports a



- 907 Favorable Impact on SARS-CoV-2 Spread in Mexico. *The Journal of Infection in Developing*  
908 *Countries*. 2020;14(9):953–956.
- 909 **67.** Verma BK, Verma M, Verma VK, Abdullah RB, Nath DC, Khan HTA, et al. Global Lockdown:  
910 An Effective Safeguard in Responding to the Threat of COVID-19. *Journal of Evaluation in*  
911 *Clinical Practice*. 2020;26(6):1592–1598.
- 912 **68.** Islam N, Sharp SJ, Chowell G, Shabnam S, Kawachi I, Lacey B, et al. Physical Distancing  
913 Interventions and Incidence of Coronavirus Disease 2019: Natural Experiment in 149 Countries.  
914 *BMJ*. 2020;370:m2743.
- 915 **69.** Wagner AB, Hill EL, Ryan SE, Sun Z, Deng G, Bhadane S, et al. Social Distancing Merely  
916 Stabilized COVID-19 in the United States. *Stat*. 2020;9(1):e302.
- 917 **70.** Silva L, Filho DF, Fernandes A. The Effect of Lockdown on the COVID-19 Epidemic  
918 in Brazil: Evidence from an Interrupted Time Series Design. *Cadernos de Saúde Pública*.  
919 2020;36(10):e00213920.
- 920 **71.** Medline A, Hayes L, Valdez K, Hayashi A, Vahedi F, Capell W, et al. Evaluating the Impact of  
921 Stay-at-Home Orders on the Time to Reach the Peak Burden of COVID-19 Cases and Deaths:  
922 Does Timing Matter? *BMC Public Health*. 2020;20:1750.
- 923 **72.** Arshed N, Meo MS, Farooq F. Empirical Assessment of Government Policies and Flattening of  
924 the COVID 19 Curve. *Journal of Public Affairs*. 2020;20:e2333.
- 925 **73.** Holtz D, Zhao M, Benzell SG, Cao CY, Rahimian MA, Yang J, et al. Interdependence and  
926 the Cost of Uncoordinated Responses to COVID-19. *Proceedings of the National Academy of*  
927 *Sciences*. 2020;117(33):19837–19843.
- 928 **74.** Li Y, Campbell H, Kulkarni D, Harpur A, Nundy M, Wang X, et al. The Temporal Association of  
929 Introducing and Lifting Non-Pharmaceutical Interventions with the Time-Varying Reproduction  
930 Number (R) of SARS-CoV-2: A Modelling Study across 131 Countries. *The Lancet Infectious*  
931 *Diseases*. 2021;21(2):193–202.
- 932 **75.** Salje H, Kiem CT, Lefrancq N, Courtejoie N, Bosetti P, Paireau J, et al. Estimating the Burden  
933 of SARS-CoV-2 in France. *Science*. 2020;369(6500):208–211.

- 934 **76.** Gatto M, Bertuzzo E, Mari L, Miccoli S, Carraro L, Casagrandi R, et al. Spread and Dynamics  
935 of the COVID-19 Epidemic in Italy: Effects of Emergency Containment Measures. Proceedings  
936 of the National Academy of Sciences. 2020;117(19):10484–10491.
- 937 **77.** Dehning J, Zierenberg J, Spitzner FP, Wibral M, Neto JP, Wilczek M, et al. Inferring  
938 Change Points in the Spread of COVID-19 Reveals the Effectiveness of Interventions. Sci-  
939 ence. 2020;369(6500):eabb9789.
- 940 **78.** Anderson SC, Edwards AM, Yerlanov M, Mulberry N, Stockdale JE, Iyaniwura SA, et al.  
941 Quantifying the Impact of COVID-19 Control Measures Using a Bayesian Model of Physical  
942 Distancing. PLOS Computational Biology. 2020;16(12):e1008274.
- 943 **79.** Wang T, Wu Y, Lau JYN, Yu Y, Liu L, Li J, et al. A Four-Compartment Model for the  
944 COVID-19 Infection — Implications on Infection Kinetics, Control Measures, and Lockdown Exit  
945 Strategies. Precision Clinical Medicine. 2020;3(2):104–112.
- 946 **80.** McCarthy Z, Xiao Y, Scarabel F, Tang B, Bragazzi NL, Nah K, et al. Quantifying the  
947 Shift in Social Contact Patterns in Response to Non-Pharmaceutical Interventions. Journal of  
948 Mathematics in Industry. 2020;10:28.
- 949 **81.** Dandekar R, Rackauckas C, Barbastathis G. A Machine Learning-Aided Global Diagnostic  
950 and Comparative Tool to Assess Effect of Quarantine Control in COVID-19 Spread. Patterns.  
951 2020;1(9):100145.
- 952 **82.** Crokidakis N. COVID-19 Spreading in Rio de Janeiro, Brazil: Do the Policies of Social Isolation  
953 Really Work? Chaos, Solitons & Fractals. 2020;136:109930.
- 954 **83.** Ge J, He D, Lin Z, Zhu H, Zhuang Z. Four-Tier Response System and Spatial Propagation of  
955 COVID-19 in China by a Network Model. Mathematical Biosciences. 2020;330:108484.
- 956 **84.** Manevski D, Gorenjec NR, Kejžar N, Blagus R. Modeling COVID-19 Pandemic Using Bayesian  
957 Analysis with Application to Slovene Data. Mathematical Biosciences. 2020;329:108466.
- 958 **85.** Li BZ, Cao NW, Zhou HY, Chu XJ, Ye DQ. Strong Policies Control the Spread of COVID-19  
959 in China. Journal of Medical Virology. 2020;92(10):1980–1987.
- 960 **86.** Zhao S, Chen H. Modeling the Epidemic Dynamics and Control of COVID-19 Outbreak in  
961 China. Quantitative Biology. 2020;8(1):11–19.

- 962 **87.** Shi Q, Hu Y, Peng B, Tang XJ, Wang W, Su K, et al. Effective Control of SARS-CoV-2  
963 Transmission in Wanzhou, China. *Nature Medicine*. 2020;27(1):86–93.
- 964 **88.** Adekunle A, Meehan M, Rojas-Alvarez D, Trauer J, McBryde E. Delaying the COVID-19  
965 Epidemic in Australia: Evaluating the Effectiveness of International Travel Bans. *Australian and  
966 New Zealand Journal of Public Health*. 2020;44(4):257–259.
- 967 **89.** Li Y, Wang LW, Peng ZH, Shen HB. Basic Reproduction Number and Predicted Trends of  
968 Coronavirus Disease 2019 Epidemic in the Mainland of China. *Infectious Diseases of Poverty*.  
969 2020;9:94.
- 970 **90.** Kendall M, Milsom L, Abeler-Dörner L, Wymant C, Ferretti L, Briers M, et al. Epidemiological  
971 Changes on the Isle of Wight after the Launch of the NHS Test and Trace Programme: A  
972 Preliminary Analysis. *The Lancet Digital Health*. 2020;2(12):e658–e666.
- 973 **91.** Kang N, Kim B. The Effects of Border Shutdowns on the Spread of COVID-19. *Journal of  
974 Preventive Medicine and Public Health*. 2020;53(5):293–301.
- 975 **92.** Tian T, Luo W, Tan J, Jiang Y, Chen M, Pan W, et al. The Timing and Effectiveness of  
976 Implementing Mild Interventions of COVID-19 in Large Industrial Regions via a Synthetic  
977 Control Method. *Statistics and Its Interface*. 2021;14(1):3–12.
- 978 **93.** Marschner IC. Back-Projection of COVID-19 Diagnosis Counts to Assess Infection Incidence  
979 and Control Measures: Analysis of Australian Data. *Epidemiology and Infection*. 2020;148:e97.
- 980 **94.** Valcarcel B, Avilez JL, Torres-Roman JS, Poterico JA, Bazalar-Palacios J, Vecchia CL. The  
981 Effect of Early-Stage Public Health Policies in the Transmission of COVID-19 for South American  
982 Countries. *Revista Panamericana de Salud Pública*. 2020;44:e148.
- 983 **95.** Wong CKH, Wong JYH, Tang EHM, Au CH, Lau KTK, Wai AKC. Impact of National  
984 Containment Measures on Decelerating the Increase in Daily New Cases of COVID-19 in 54  
985 Countries and 4 Epicenters of the Pandemic: Comparative Observational Study. *Journal of  
986 Medical Internet Research*. 2020;22(7):e19904.
- 987 **96.** Nouvellet P, Bhatia S, Cori A, Ainslie KEC, Baguelin M, Bhatt S, et al. Reduction in Mobility  
988 and COVID-19 Transmission. *Nature Communications*. 2021;12:1090.

- 989 **97.** Lison A, Persson J, Banholzer N, Feuerriegel S. Estimating the Effect of Mobility on SARS-  
990 CoV-2 Transmission during the First and Second Wave of the COVID-19 Epidemic, Switzerland,  
991 March to December 2020. *Eurosurveillance*. 2022;27(10):2100374.
- 992 **98.** Coletti P, Wambua J, Gimma A, Willem L, Vercruysse S, Vanhoutte B, et al. CoMix: Comparing  
993 Mixing Patterns in the Belgian Population during and after Lockdown. *Scientific Reports*.  
994 2020;10:21885.
- 995 **99.** Chernozhukov V, Kasahara H, Schrimpf P. Causal impact of masks, policies, behavior on early  
996 covid-19 pandemic in the U.S. *Journal of Econometrics*. 2021;220(1):23–62.
- 997 **100.** Persson J, Parie JF, Feuerriegel S. Monitoring the COVID-19 Epidemic with Na-  
998 tionwide Telecommunication Data. *Proceedings of the National Academy of Sciences*.  
999 2021;118(26):e2100664118.
- 1000 **101.** Shen Y, Powell G, Ganser I, Zheng Q, Grundy C, Okhmatovskaia A, et al. Monitoring  
1001 Non-Pharmaceutical Public Health Interventions during the COVID-19 Pandemic. *Scientific*  
1002 *Data*. 2021;8:225.
- 1003 **102.** Allcott H, Boxell L, Conway JC, Ferguson BA, Gentzkow M, Goldman B. What Explains  
1004 Temporal and Geographic Variation in the Early US Coronavirus Pandemic? *National Bureau of*  
1005 *Economic Research*; 2020. 27965.
- 1006 **103.** Yan Y, Malik AA, Bayham J, Fenichel EP, Couzens C, Omer SB. Measuring Voluntary and  
1007 Policy-Induced Social Distancing Behavior during the COVID-19 Pandemic. *Proceedings of the*  
1008 *National Academy of Sciences*. 2021;118(16):e2008814118.
- 1009 **104.** Grossman G, Kim S, Rexer JM, Thirumurthy H. Political partisanship influences behav-  
1010 ioral responses to governors' recommendations for COVID-19 prevention in the United States.  
1011 *Proceedings of the National Academy of Sciences*. 2020;117(39):24144–24153.
- 1012 **105.** Herby J. A First Literature Review: Lockdowns Only Had a Small Effect on COVID-19. *Social*  
1013 *Science Research Network*; 2021. 3764553.
- 1014 **106.** Vandembroucke JP, von Elm E, Altman DG, Gøtzsche PC, Mulrow CD, Pocock SJ, et al.  
1015 Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): explanation  
1016 and elaboration. *Epidemiology (Cambridge, Mass)*. 2007;18(6):805–835.

- 1017 **107.** Benchimol EI, Smeeth L, Guttman A, Harron K, Moher D, Petersen I, et al. The REporting  
1018 of studies Conducted using Observational Routinely-collected health Data (RECORD) statement.  
1019 PLoS Medicine. 2015;12(10):e1001885.
- 1020 **108.** Campbell M, Katikireddi SV, Hoffmann T, Armstrong R, Waters E, Craig P. TIDieR-PHP: a  
1021 reporting guideline for population health and policy interventions. BMJ. 2018;361:k1079.

## 1022 **Declarations**

### 1023 **Ethics approval**

1024 Ethics approval was not required for this study.

### 1025 **Competing interests**

1026 SF reports membership in a COVID-19 working group of the World Health Organization but without  
1027 competing interest. SF reports grants from the Swiss National Science Foundation outside of the  
1028 submitted work. All authors declare no competing interests.

### 1029 **Funding**

1030 NB and SF acknowledge funding from the Swiss National Science Foundation (SNSF) as part of the  
1031 Eccellenza grant 186932 on “Data-driven health management”. The funding bodies had no control  
1032 over design, conduct, data, analysis, review, reporting, or interpretation of the research conducted.

### 1033 **Contributions**

1034 NB and AL contributed to conceptualization, literature search, study selection, data extraction,  
1035 data synthesis, and manuscript writing, review & editing. DO contributed to literature search and  
1036 manuscript review & and editing. TS contributed to manuscript review & editing. SF contributed  
1037 to manuscript writing, review & editing. WV contributed to conceptualization, study selection, data  
1038 extraction, data synthesis, manuscript writing, review & editing.

### 1039 **Data availability**

1040 The full data extracted from the studies in this review is available in machine-readable format at  
1041 <https://github.com/adrian-lison/methodologies-npi-effects>.