

1 **Title:** Mask mandates can limit COVID spread: Quantitative assessment of month-over-month
2 effectiveness of governmental policies in reducing the number of new COVID-19 cases in 37 US States
3 and the District of Columbia

4
5 **Authors:** Michael J. Maloney¹, Nathaniel J. Rhodes^{2-4#}, Paul R. Yarnold⁵

6
7 **Affiliations:** ¹Proof School, San Francisco, CA, USA; ²Department of Pharmacy, Northwestern Memorial
8 Hospital, Chicago, IL, USA; ³Department of Pharmacy Practice, Midwestern University Chicago College
9 of Pharmacy, Downers Grove, IL, USA; ⁴Pharmacometrics Center of Excellence, Midwestern University
10 Chicago College of Pharmacy, Downers Grove, IL, USA; ⁵Optimal Data Analysis, LLC, Chicago, IL,
11 USA

12
13 **#Contact information corresponding author:** Nathaniel J. Rhodes, PharmD, MSc, BCPS AQ-ID
14 Midwestern University Chicago College of Pharmacy, Department of Pharmacy Practice. 555 31st Street
15 Downers Grove, IL 60515, USA; Tel: 630-515-7376; Email: nrhode@midwestern.edu

16
17 **Running title:** Mask Mandates and New COVID-19 Cases

18
19 **Funding Information:** The author(s) received no specific funding for this work.

20
21 **Competing Interests:** The authors have declared that no competing interests exist.

22
23 **Keywords:** SARS-CoV-2; COVID-19; masks; public health; epidemiology, policy

24
25 **Abstract word count:** 250

Manuscript word count: 4191

26 **ABSTRACT**

27 **Introduction:** SARS-CoV-2 is the beta-coronavirus responsible for COVID-19. Facemask use has been
28 qualitatively associated with reduced COVID-19 cases, but no study has quantitatively assessed the
29 impact of government mask mandates (*MM*) on new COVID-19 cases across multiple US States.

30 **Data and Methods:** We utilized a non-parametric machine-learning algorithm to test the *a priori*
31 hypothesis that *MM* were associated with reductions in new COVID-19 cases. Publicly available data
32 were used to analyze new COVID-19 cases from 37 States and the District of Columbia (i.e., “38
33 States”). We conducted confirmatory All-States and State-Wise analyses, validity analyses [e.g., leave-
34 one-out (LOO) and bootstrap resampling], and covariate analyses.

35 **Results:** No statistically significant difference in the daily number of new COVID-19 infections was
36 discernable in the All-States analysis. In State-Wise LOO validity analysis, 11 States exhibited reductions
37 in new COVID-19 and the reductions in four of these States (AK, MA, MN, VA) were significant in
38 bootstrap resampling. Only the Social Capital Index predicted *MM* success (training $p < 0.028$ and LOO
39 $p < 0.013$).

40 **Conclusion:** Results obtained when studying the impact of *MM* on COVID-19 cases varies as a function
41 of the heterogeneity of the sample being considered, providing clear evidence of Simpson's Paradox and
42 thus of confounded findings. As such, studies of *MM* effectiveness should be conducted on disaggregated
43 data. Since transmissions occur at the individual rather than at the collective level, additional work is
44 needed to identify optimal social, psychological, environmental, and educational factors which will
45 reduce the spread of SARS-CoV-2 and facilitate *MM* effectiveness across diverse settings.

46 **INTRODUCTION**

47 The significance of respiratory droplets and airborne viral particles in the spread of SARS-CoV-2 and
48 propagation of the COVID-19 pandemic gained early attention.¹ Asymptomatic carriage of SARS-CoV-2
49 is important to community spread because these individuals can transmit the virus in exhaled breath.^{2,3}
50 Corollary research reported that face masks are protective⁴⁻⁷ and may limit the severity of COVID-19
51 among individuals who become infected.⁸ Indeed, the Centers for Disease Control and Prevention
52 (CDC) has advocated wearing a facemask when social isolation is impossible or impractical,⁹ and
53 recently reported that dining-out is among the riskiest known activities during the coronavirus
54 pandemic—since face masks *are not used* when people are eating and drinking.¹⁰ Other activities
55 associated with an increased risk of SARS-CoV-2 transmission include singing¹¹ and aerosol generating
56 procedures in the healthcare setting.¹² Small droplet formation, which can lead to aerosolization of the
57 virus, is a major driver of infection risk,¹³ thus wearing a facemask is expected to reduce inhalation of
58 aerosolized virus.⁴ Indeed, low rates of positive serology were observed (i.e., < 4% n=1/27) when
59 healthcare workers followed aerosol minimizing procedures *and* used personal protective equipment to
60 perform deep respiratory sampling in patients with COVID-19.¹⁴ Thus, there is consensus in the scientific,
61 clinical, and business communities that appropriate wearing of facemasks is a “best-practice” personal
62 behavior, which can reduce the chances of being infected by, or of infecting others with, SARS-CoV-2
63 and other airborne disease-causing microbes.

64 Mathematical and time-series models of COVID-19 data collected in the US investigated salutary effects
65 of masking on reducing the spread of the pandemic.^{15,16} These models suggest that a consistent
66 application of best-practice public-health orders and citizen behaviors is needed to slow the spread of the
67 pandemic.¹⁷ A hybrid modeling approach used by the Institute for Health Metrics and Evaluation
68 predicted that if 95% of people in the US wore masks outside their homes, the number of projected deaths
69 from COVID-19 would drop by half within four months.¹⁸ Thus, experts hypothesize that Mask Mandate

70 (*MM*) orders should result in reduced COVID-19 transmission and fatalities—assuming that the
71 individuals in the community adhere to the *MM*.

72 Encouraged by research findings that wearing masks reduces the number of new COVID-19 cases, and
73 motivated to intervene by increasing numbers of cases, Governors of 37 US States and the District of
74 Columbia (hereafter called the “38 States”) issued *MM* orders. However, the lack of clear, consistent
75 Global, Federal, and State guidance on the use of face coverings (e.g., mandates *vs.* recommendations),
76 and on who should be using masks (e.g., essential workers *vs.* private citizens), has initiated a nation-wide
77 naturalistic experiment.¹⁵ In the absence of clear guidance and direct leadership, variations in *MM*
78 acceptance and enforcement have arisen within and between US States.^{17, 19-22}

79 Promulgating and exacerbating confusion regarding the effectiveness of wearing face masks, adjudicating
80 the effectiveness of *MM* orders has largely been left to *qualitative interpretations* and *visual/numerical*
81 *inspections* of the epidemic curves depicting the number of new cases over time. Qualitative assessment
82 cannot objectively test the *a priori* hypothesis that wearing a mask in public spaces reduces the incidence
83 of COVID-19 cases. Accordingly, using publicly-available data, we employ non-parametric maximum-
84 accuracy machine-learning to *quantitatively* test the confirmatory (*a priori*) hypothesis that imposing a
85 *MM reduced* daily number of new COVID-19 cases in the month *after vs. before* the *MM*.

86 **METHODS**

87 **Hypotheses**

88 First, merging data from all 38 States in which the Governor issued a *MM*, we test the *a priori* hypothesis
89 that “*All-States*” imposition of a *MM reduced* the daily number of new COVID-19 cases in the month
90 *after vs.* in the month *before* the *MM*. States were included if State leadership (e.g., Governors) had
91 mandated the use of facemasks by all public-facing employees.

92 Second, considering each State individually, we test the *a priori* hypothesis that “*State-Wise*” imposition
93 of *MM reduced* the daily number of new COVID-19 cases in the month *after vs.* in the month *before* the
94 *MM*.

95 **Data**

96 The daily number of new COVID-19 cases was obtained separately for each State in the month *before* the
97 *MM* was made (maximum of 30 days), and in the month *after* the *MM* was made (maximum of 31 days).
98 Case reports occurring before the *MM* were dummy-coded as class=0, and case reports occurring after the
99 *MM* were coded as class=1. Data from The New York Times, based on reports from State and local health
100 agencies, were initially downloaded from GitHub on August 8, 2020 (data re-confirmed on September 23,
101 2020)²³ and cross referenced against the COVID-19 State Policy Database (data through June 1, 2020) for
102 dates that *MM* were issued in each State.²⁴

103 Our decision to compare daily number of new cases in the month before *vs.* the month after the *MM* was
104 based on methodological and statistical considerations. First, *MM* orders were made because the day-
105 over-day increases in the number of new cases threatened the ability of available health care resources to
106 meet patient and healthcare worker needs. Therefore, the month following the *MM* naturally offers great
107 opportunity for increased mask wearing to meaningfully reduce the number of new cases. Second, some
108 amount of time is required after the *MM* is made for citizens (who wish to comply) to obtain masks, learn

109 how and when to wear them, and how to care for them, and for new infections occurring in the two-week
110 period prior to the *MM* to develop symptoms after the *MM* was made. Third, to maximize the available
111 statistical power of our analyses for estimating short-term effects (see *Statistical Analysis*), we extracted
112 the number of cases for up to 30 days before and up to 31 days after the *MM* was implemented (yielding a
113 maximum of n=61 data points for each state).

114 Additional attributes, obtained for individual States, were used in an attempt to discriminate States for
115 which a *MM* did vs. did not reduce the number of new COVID-19 cases. Numerical attributes included
116 population estimates,²⁶ gross domestic product (State and per capita),²⁷ age (with arbitrary brackets of 0-
117 18, 18-25, 26-34, 35-54, 55-64, and 65+),²⁸ pre-pandemic (January 2020) number of unemployed persons
118 and rate,²⁹ homelessness,³⁰ shelter beds,³¹ incarceration number and rate,³² number of nursing home
119 residents,³³ population density (people per square mile in 2015),³⁴ urban overcrowding (number of houses
120 having >1 person per room),³⁵ severe urban overcrowding (number of houses having >1.5 people per
121 room),³⁵ ethnicity (white, black, and Hispanic categories),³⁶ Social Capital Index (a measure of social
122 cohesion, where higher positive numbers indicate greater societal cohesion),³⁷ and obesity.³⁸ Categorical
123 attributes included if the State's Governor wore a mask in public after the *MM* was made (0=Did not wear
124 mask, 1=wore mask) which was assessed using media reports and public database image searches, the
125 Governor's political party (0=Republican, 1=Democrat),³⁹ and the Governor's gender (0=Female,
126 1=Male).⁴⁰

127 **Statistical Analysis**

128 We used confirmatory *Optimal Discriminant (or Data) Analysis (ODA)* to ascertain the *predictive*
129 *accuracy* attained by using maximum-accuracy models to evaluate the *hypothesized reduction in the*
130 *number of new cases* occurring subsequent to the *MM*.⁴¹⁻⁴³ ODA is the moniker of the statistically-
131 motivated non-parametric machine-learning algorithm which identifies the cut-point for an ordered
132 *attribute* (i.e., "independent variable"), or the assignment rule for a categorical attribute, which yields the

133 *strongest achievable accuracy* in discriminating between two or more class categories (i.e., “dependent
134 variable”) for a given sample and hypothesis.⁴¹⁻⁴³ In the present study all hypotheses were confirmatory.⁴²
135 For both the All-States (38-State) and the State-Wise analysis, the *a priori* hypothesis was that the *MM*
136 (up to 31 days) reduced the number of new cases.^{44,45}

137 An ODA model is obtained by iterating through every possible different assignment rule that is consistent
138 with the *a priori* hypothesis, and identifying the model achieving the highest effect strength for sensitivity
139 (ESS) statistic.^{41,42} An index of classification accuracy adjusted to remove the effect of chance, ESS is a
140 function of the mean sensitivity of the model achieved across class categories, standardized such that 0
141 represents the discriminatory accuracy which is expected by chance, 100 represents errorless (perfect)
142 discrimination, and values of $-100 \leq \text{ESS} < 0$ represent accuracy which is worse than expected by chance.
143 For a two-category *class variable* such as studied herein (i.e., month before *MM*=class 0 vs. month after
144 *MM*=class 1),

$$145 \quad \text{Mean Classification Accuracy (\%)} = 100 \times [(\text{sensitivity for class 0} + \text{sensitivity for class 1})/2], \quad (1)$$

146 and

$$147 \quad \text{ESS} = 100 \times [(\text{Mean Classification Accuracy} - 50)/50]. \quad (2)$$

148 A nonparametric permutation test which requires no distributional assumptions is used to assess the
149 statistical significance (*p* value) of the achieved ESS.^{41,44,45} Point estimates of *p* values for tests of
150 statistical hypotheses reported herein are indicated as being statistically significant ($p \leq 0.05$) at either the
151 *experimentwise* (confirmatory $p \leq p_{\text{Bonferroni}}$) or the *per-comparison* (confirmatory $p \leq 0.05$) criterion.⁴²

152 Internal model validation was conducted using bootstrap resampling^{46,47} ($n=25,000$) of the State-Wise
153 models within the *ODA* package for R.⁴⁸ Specifically, resampling with 50% replacement was conducted
154 using the LOO (i.e., validity) ODA model confusion matrix for each State. Exact discrete 95% confidence
155 intervals (CI) for the given ODA model (“Model”) and for randomly scrambled observations from the

156 model (“Chance”) were obtained.⁴⁶ Graphical depictions of new cases organized across days, and
157 histograms of bootstrapped 95% CIs, were created using *ggplot2* for R.⁴⁹

158 Potential generalizability of the ODA model applied to classify an independent random sample is assessed
159 using a one-sample “leave-one-out” (LOO) jackknife analysis.⁴² The sample size offers greater than 90%
160 power to identify an ODA model of moderate strength (i.e., $25 \leq \text{ESS} < 50$).⁵⁰

161 Exploratory ODA analyses were conducted to model the association of inter-State differences 28
162 variables (e.g., age, wealth, population density) conceivably capable of discriminating States which did
163 *vs.* did not exhibit the predicted decline in number of new cases in the month following a *MM* (see *Data*
164 *Supplement*).⁴²

165 Sensitivity analyses were conducted comparing the initial data (obtained August 8, 2020) to the data
166 obtained September 23, 2020. The percentage difference between the two datasets were calculated for
167 each value [i.e., $100 \times (\text{data}_{\text{new}} - \text{data}_{\text{old}}) / \text{data}_{\text{old}}$] to quantify differences between datasets. In sensitivity
168 analyses, any differences between datasets were compared according to the ODA model cut-points
169 obtained using the initial data to determine whether the model was invariant to new (updated) data.

170 RESULTS

171 All-States Analysis

172 The *mean* number of new COVID-19 cases before *vs.* after the *MM* for the sample of 38 States was 654
173 (N=1138, SD=1357) *vs.* 639 (N=1177, SD=975), respectively. In three instances, the daily report of new
174 cases appeared to be outliers (GA 4/12/20, LA 5/29/20, and VA 5/6/20) and were coded as missing (i.e.,
175 coded as -999). Confirmatory ODA was used to test the *a priori* hypothesis that imposition of a *MM*
176 reduced the *number* of new COVID-19 cases (the attribute) one-month *after vs.* one-month *before* the
177 *MM* (the class variable). The confirmatory ODA model was: if >4 new cases-per-day then predict pre-
178 *MM*; otherwise if ≤4 new cases-per-day, then predict post-*MM*. In training (using all observations from
179 the sample) analysis, this model correctly classified 98.1% of the pre-*MM* cases, and 6.1% of the post-
180 *MM* cases: ESS=4.1 (a weak effect), $p < 0.13$.

181 Thus, when considering all 38 States having a *MM* as a single sample, there was no statistically
182 significant difference in the daily number of new COVID-19 infections in the month before *vs.* the month
183 after the *MM*.

184 State-Wise Analysis

185 As seen in Table 1, after the *MM* the *mean* number of new daily cases was numerically *lower* in 13 States
186 (AK, GA, HI, LA, MA, MI, MN, NY, VA, VT, WA, WV, WY) post-*MM*, and numerically *higher* in the
187 remaining 25 States. Table 2 gives the findings of ODA used to test the *a priori* hypothesis that State-
188 Wise imposition of a *MM* (made by 38 different State Governors) *reduced the number of new COVID-19*
189 *cases* in the month *after vs.* in the month *before* the *MM*—with each State considered individually. In
190 each State-Wise analysis, the class variable was a binary coded variable signifying the period before *vs.*
191 after the *MM* occurred, and the attribute was the daily number of new cases. In Table 2, the States are
192 ordered from strongest to weakest ESS of the confirmatory ODA model in training analysis.

193 Among all 38 States, training data for Minnesota (MN, the first State in Table 2) offer the greatest support
194 for the *a priori* hypothesis. The ODA model is: if >480 new cases-per-day, then predict pre-*MM*;
195 otherwise if ≤ 480 cases-per-day, then predict post-*MM*. This ODA model correctly classified 28 of 30
196 (93.3%) of the pre-*MM* days, and 27 of 31 (87.1%) of the post-*MM* days, thereby yielding a *strong*⁴²
197 ESS=80.4, $p < 0.0001$. However, this model showed modest instability in LOO jackknife analysis
198 (*relatively strong*⁴² ESS=73.9, $p < 0.0001$). In contrast, data for Alaska (AK, second row in Table 2) had
199 the second-strongest training effect (*strong* ESS=80.3, $p < 0.0001$), which was stable in LOO analysis.
200 Among all 38 States, only AK returned a *strong* confirmation of the *a priori* hypothesis in training and
201 LOO analysis.

202 Of the total of 38 States included in this study, 11 (23.7%) achieved confirmatory $p < 0.05$ at the
203 experimentwise criterion, and two (5.3%) achieved confirmatory $p < 0.05$ at the per-comparison criterion⁴²
204 (Table 2). Three States (MN, AK, HI) yielded a training model indicating a *strong* confirmation⁴² (i.e.,
205 $ESS \geq 75$) of the *a priori* hypothesis. Five States (MA, VT, VA, MI, NY) yielded a training model
206 indicating a *relatively strong* confirmation⁴² (i.e., $50 \leq ESS < 75$). And, three States (LA, WA, WY) yielded
207 a training model indicating a *moderate* confirmation⁴² (i.e., $25 \leq ESS < 50$) of the *a priori* hypothesis.

208 Except for GA and WY, States having a numerically greater *mean* number of new daily cases pre- vs.
209 post-*MM* (Table 1) also had statistically-significant (at either the experimentwise or the per-comparison
210 criterion⁴²) confirmatory (Table 2) and LOO (Table 3) ODA models supporting the *a priori* hypothesis.

211 For all 11 States in Table 2 having training $p < 0.05$, Table 3 presents 95% exact discrete CIs obtained for
212 model and for chance.⁴⁶ For AK, VA, MA, and MN, the *lower bound* of the 95% exact discrete CI *for*
213 *model* is greater than the *upper bound* of the 95% exact discrete CI *for chance*—therefore the result is
214 statistically significant. For the remaining seven States, the upper bound for chance exceeded the lower
215 bound for the model, so the result was not statistically significant in bootstrap validity analysis.

216 For example, the daily number of new COVID-19 cases one-month pre- and one-month post-*MM* is
217 illustrated in Figure 1.A for MN. Figure 1.B is a graphical summary of the model and chance bootstrap
218 results: the blue histogram represents chance, the red histogram represents the model, the Y axis is
219 frequency of the result in the Bootstrap analysis, and the X axis is training ESS. The dashed vertical lines
220 inside the Figure represent the lower 95% CI for the model (red dashes), and the upper 95% CI for chance
221 (blue dashes). When the blue dashes are on the same side as the blue histogram, and the red dashes are on
222 the same side as the red histogram, the effect is statistically significant. Thus, evaluating MN in training
223 analysis, the *a priori* hypothesis is accepted with experimentwise $p < 0.05$.

224 In contrast, Figure 2.A presents the daily number of new COVID-19 cases one-month pre- and one-month
225 post-*MM* for WA, and Figure 2.B is a graphical summary of the model and chance bootstrap results.

226 Since the blue dashes are on the side of the red histogram, and the red dashes are on the side of the blue
227 histogram, the hypothesized effect is not statistically significant for WA.

228 **Modeling Success vs. Failure of a *MM* Order in Reducing the Number of New COVID-19 Cases**

229 Finally, ODA was used to discriminate the four States (AK, MA, MN, VA) having a significantly lower
230 number of new cases after the *MM* vs. the 34 States which did not support the *a priori* hypothesis. The
231 only attribute which met the per-comparison criterion for statistical significance⁴² was SCI, a measure of
232 societal cohesion and the closeness of social ties.⁵¹ The ODA model was: if $SCI \leq 0.311$, then predict the
233 *MM* did not reduce the number of new cases of COVID-19; otherwise if $SCI > 0.311$ then predict the *MM*
234 reduced the number of new cases. In training ($p < 0.028$) and LOO ($p < 0.013$) analysis this model correctly
235 classified all four States in which the *MM* reduced the number of new cases, and 25 of 35 (71.43%) States
236 in which the *MM* failed to reduce the number of new cases. This level of classification accuracy
237 corresponds to $ESS = 71.43$, indicating a relatively strong effect.⁴²

238 **Sensitivity analyses**

239 The dataset obtained on August 8, 2020 to was compared to a dataset obtained on September 23, 2020
240 and after exclusion of outliers (n=3), case numbers in the new dataset matched the original dataset (i.e.,
241 0% difference) for all states except Oregon (n=12/61) and Virginia (n=5/12), for which the new data fell
242 within $\pm 6\%$ of the original data. These numerical changes in data values (none of which were near the
243 ODA model cut-points) did not change the ODA model, the accuracy (ESS), or statistical significance (p)
244 of the model in training or LOO analysis for any model.

245 **DISCUSSION**

246 An *All-States* analysis was conducted using merged data for 38 US States. Results indicated *no*
247 *statistically significant decrease* in the number of new daily COVID-19 cases in the month before *vs.*
248 after the *MM* was made. In contradistinction, a *State-Wise* analysis identified *clear evidence of State-*
249 *specific MM effectiveness in reducing the number of new cases within one month of enactment* for
250 multiple States (Tables 2 and 3). Such clear inconsistency is attributable to improper analysis of merged
251 samples creating an analytic anomaly called *Simpson’s Paradox*—which may identify spurious
252 relationships, may fail to identify actual relationships, and/or may over- or under-weight the strength of
253 effects—in every area of quantitative empirical science.^{42,47} *Paradoxical confounding* exists when the
254 finding obtained by statistical analysis conducted for a combined sample is different than the finding
255 obtained by the same analysis conducted separately for the constituent samples.⁵² In the setting of a public
256 health emergency, such as the current pandemic, accurate knowledge of the effectiveness of public health
257 interventions is essential so that public health authorities can communicate this to the public.

258 Combining disparate groups is the primary operating procedure in government, scientific, and media
259 discussions about the number of new and cumulative daily, monthly, and total number of COVID-19
260 cases and fatalities. Examples of such combined groups include global (the maximum-possible combined
261 group), regional (e.g., the northern hemisphere), continental (e.g., Europe, Africa), and country (e.g.,
262 China, US, India, Brazil, France, UK, Israel) daily and cumulative case counts.⁵³ However, disparities in
263 regional increases in numbers of new cases prompted some analysts to issue a warning against focusing
264 on National trends.⁵⁴ Findings presented herein—that four of 38 (10.5%) States reliably demonstrated the
265 hypothesized reduction in number of new cases in the month after the *MM* was made—similarly suggest a
266 warning against focusing on State trends. Cognizance of the need to combine groups in a manner that
267 prevents invalid statistical conclusions begs the question of how small must a “catchment area” be to
268 count new cases so as to prohibit paradoxical confounding. Unfortunately, it has been shown that
269 Simpson’s Paradox can even occur for single-subject, “N-of-1” designs.⁵⁵ Thus, while accurate case

270 counts and definitions are important for tracking the epidemiology of the pandemic, such measures do not
271 necessarily reflect the “reality on the ground” in any specific community, nor can they reliably discern the
272 individual effectiveness of public health interventions.

273 The current methodology of conducting after-the-fact pandemic accounting for convenient and familiar
274 (statistically unmotivated) groups is obviously inadequate when it is considered from a problem-solving
275 perspective that focuses on getting (and staying) ahead of the pandemic curve. Others have argued, early
276 in the pandemic, that rapid and comprehensive contact tracing will empower understanding of the
277 situations and the processes underlying new COVID-19 infections.⁵⁶ As is the case in fighting a wildfire,
278 progress is made by going to the active location each day, extinguishing blazes, and reducing available
279 fuel, rather than by assessing the number of acres which burned the day before. However, contact tracing
280 is most effective when community spread is low, so multifaceted interventions are needed to control the
281 spread of the pandemic and presently wearing a mask is the most efficacious known means of slowing the
282 pandemic.^{5, 6, 8, 15, 16}

283 Temporal counts of the number of new cases are an imprecise measure—including the month-over-month
284 methodology used presently—for multiple reasons. Even with perfect *MM* compliance, there is an up to
285 14-day lag time between infection and presentation of symptoms, and potentially an even longer lag
286 between infection and presentation at a healthcare facility. After a *MM* order is made, there are logistical
287 problems in getting masks to people, and there is a learning curve for users—where, when and how to
288 wear and care for masks. *MM* compliance is imperfect, and enforcement is difficult. Indeed, individual
289 *MM* compliance is expected to be heterogeneous based on experimental evidence from behavioral
290 psychology. Individuals are susceptible to cognitive biases and often use heuristics to address complex
291 problems when uncertainty is high. Examples of such biases and heuristics related to *MM* use include
292 availability bias (e.g., knowing someone who is sick vs. not), anchoring bias (e.g., fixation on initial
293 recommendations to avoid mask use to preserve PPE), and substitution heuristics (e.g., physical

294 discomfort with wearing a mask is substituted for mask efficacy).⁵⁷ The impact of these and other biases
295 and heuristics on mask effectiveness is currently understudied.

296 Because transmissions occur at the level of the individual, rather than at a State or National level,
297 governments should emphasize best-practice personal behaviors, adequate contact tracing, and
298 minimization of high-risk exposures. Indeed, we posit that the efficacy of mask wearing is not at question
299 as this depends only on the use of the mask. Instead, the effectiveness of *MM* can be eroded by multiple
300 factors including asymptomatic carriage, willful ignorance of *MM* and lack of knowledge of how and
301 when to wear masks. Moreover, our State-Wise analysis supports the *a priori* hypothesis that *MM* do
302 indeed have a significant salutary effect on the spread of the pandemic, but more work is needed to define
303 the social, psychological, environmental, and educational facilitators that maximize *MM* effectiveness.

304 Experimentally speaking, the best test of public health intervention effectiveness (e.g., *MM*)—using a
305 methodology called an “A (before intervention), B (after intervention)” design—would involve
306 simultaneous uniform application, enforcement, and education around the intervention, coupled with
307 exacting real-time contact-tracing. A science-based, consistent approach to disease management could
308 reduce public misunderstanding in this realm, and diminish mistrust of the other public health
309 interventions (e.g., vaccines currently under development).⁵⁸ Our analysis is limited because we were not
310 able to utilize a rigorous AB analytic framework to test our *a priori* hypotheses. Rather, we took
311 advantage of the naturalistic experiment created by the heterogeneous response to the pandemic across
312 States to highlight the paradoxical results obtained when applying an All-States followed by a State-Wise
313 analytic approach. Our analysis was limited to testing our *a priori* hypothesis and it is worth considering
314 the time-dependent impact of *MM*. Additional work is needed to evaluate the impact of *MM* on the
315 growth of the pandemic. Our analysis is conservative in that we focused only on reductions in new cases
316 and may underestimate the true impact of *MM*. Future studies should evaluate the impact of *MM* on new
317 deaths. Large-scale serological studies may provide more insights into the spread of SARS-CoV-2 *MM* in
318 States with vs. those without early *MM*.

319 Finally, we found that SCI, a measure of social closeness and ties,⁵¹ was a relatively strong predictor of
320 *MM* effectiveness. This finding is notable in that SCI could be a facilitating factor that makes mandated
321 public health interventions more efficient. For example, populations with higher levels of social cohesion
322 are inherently more likely to work together when faced with a common problem based on a shared sense
323 of identity and social bonds. However, this finding has limitations. States with higher SCI should not be
324 interpreted as being inherently better at reducing new SARS-CoV-2 cases, but rather we believe this
325 finding also supports a new corollary hypothesis that States with lower SCI may require different and
326 multifaceted interventions to address the pandemic.

327 **CONCLUSION**

328 We identified paradoxical confounding whereby the effect of *MM* orders when analyzed using the pooled
329 data from all 38 States available differed when compared to a State-Wise analysis. Overemphasis placed
330 on National or State-level trends may obscure the true effectiveness of important public health
331 interventions such as mask wearing, because transmission events are stochastic and occur at the level of
332 individuals. Thus, greater emphasis on the impact of personal behaviors on *MM* effectiveness and
333 thorough contact tracing are needed to combat the pandemic. While societal factors reflected by measures
334 like the SCI may facilitate the efficiency of interventions like *MM* in select regions, more work is needed
335 to define the optimal set of interventions that reduce SARS-CoV-2 transmission for socially diverse
336 populations. Finally, optimal experimental designs, measurements, and analytic approaches are needed to
337 provide the public with a clearer understanding of the evolution of the pandemic, and feedback on how
338 current public health interventions are performing, in order to begin to restore broad-based trust in the
339 scientific method.

340 **ACKNOWLEDGEMENT**

341 We would like to acknowledge Dr. Alan Maloney for his helpful edits and suggestions in reviewing the
342 drafts of this manuscript.

343

344 **FUNDING:**

345 No funding was received. This study was completed as part of our normal work.

346

347 **DISCLOSURES:**

348 All authors: No relevant disclosures.

349 **REFERENCES**

- 350 [1] **Zhang R, Li Y, Zhang AL, Wang Y and Molina MJ** (2020) Identifying airborne transmission as the
351 dominant route for the spread of COVID-19. *Proceedings of the National Academy of Sciences* 117:
352 14857-14863. doi: 10.1073/pnas.2009637117
- 353 [2] **Bae SH, Shin H, Koo H-Y, Lee SW, Yang JM, et al.** (2020) Asymptomatic Transmission of SARS-
354 CoV-2 on Evacuation Flight. *Emerging Infectious Diseases* 26. doi: 10.3201/eid2611.203353
- 355 [3] **Centers for Disease Control and Prevention (CDC)** (2020) How COVID-19 Spreads. Available at:
356 <https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/how-covid-spreads.html> Accessed on:
357 September 21, 2020
- 358 [4] **Leung NHL, Chu DKW, Shiu EYC, Chan K-H, McDevitt JJ, et al.** (2020) Respiratory virus
359 shedding in exhaled breath and efficacy of face masks. *Nature Medicine* 26: 676-680. doi:
360 10.1038/s41591-020-0843-2
- 361 [5] **Wang X, Ferro EG, Zhou G, Hashimoto D and Bhatt DL** (2020) Association Between Universal
362 Masking in a Health Care System and SARS-CoV-2 Positivity Among Health Care Workers. *Jama* 324:
363 703. doi: 10.1001/jama.2020.12897
- 364 [6] **Clase CM, Fu EL, Joseph M, Beale RCL, Dolovich MB, et al.** (2020) Cloth Masks May Prevent
365 Transmission of COVID-19: An Evidence-Based, Risk-Based Approach. *Annals of Internal Medicine*
366 173: 489-491. doi: 10.7326/m20-2567
- 367 [7] **Mitze T, Kosfeld R, Rode J and Wälde K** (2020) Face Masks Considerably Reduce Covid-19 Cases
368 in Germany [pre-print]. medRxiv. doi: 10.1101/2020.06.21.20128181
- 369 [8] **Gandhi M, Beyrer C and Goosby E** (2020) Masks Do More Than Protect Others During COVID-
370 19: Reducing the Inoculum of SARS-CoV-2 to Protect the Wearer. *Journal of General Internal Medicine*.
371 doi: 10.1007/s11606-020-06067-8
- 372 [9] **Centers for Disease Control and Prevention (CDC)** (2020) Considerations for Wearing Masks:
373 Help Slow the Spread of COVID-19. Available at: [https://www.cdc.gov/coronavirus/2019-ncov/prevent-](https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/cloth-face-cover-guidance.html)
374 [getting-sick/cloth-face-cover-guidance.html](https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/cloth-face-cover-guidance.html) Accessed on: August 7, 2020
- 375 [10] **Fisher KA, Tenforde MW, Feldstein LR, Lindsell CJ, Shapiro NI, et al.** (2020) Community and
376 Close Contact Exposures Associated with COVID-19 Among Symptomatic Adults ≥ 18 Years in 11
377 Outpatient Health Care Facilities — United States, July 2020. *MMWR Morbidity and Mortality Weekly*
378 *Report* 69: 1258-1264. doi: 10.15585/mmwr.mm6936a5
- 379 [11] **Hammer L, Dubbel P, Capron I, Ross A, Jordan A, et al.** (2020) High SARS-CoV-2 Attack Rate
380 Following Exposure at a Choir Practice — Skagit County, Washington, March 2020. *MMWR Morbidity*
381 *and Mortality Weekly Report* 69: 606-610. doi: 10.15585/mmwr.mm6919e6
- 382 [12] **Semple MG, Tran K, Cimon K, Severn M, Pessoa-Silva CL, et al.** (2012) Aerosol Generating
383 Procedures and Risk of Transmission of Acute Respiratory Infections to Healthcare Workers: A
384 Systematic Review. *PLoS ONE* 7: e35797. doi: 10.1371/journal.pone.0035797
- 385 [13] **Anfinrud P, Stadnytskyi V, Bax CE and Bax A** (2020) Visualizing Speech-Generated Oral Fluid
386 Droplets with Laser Light Scattering. *New England Journal of Medicine* 382: 2061-2063. doi:
387 10.1056/NEJMc2007800
- 388 [14] **Gao CA, Bailey JI, Walter JM, Coleman JM, Malsin ES, et al.** (2020) Bronchoscopy on
389 Intubated COVID-19 Patients is Associated with Low Infectious Risk to Operators at a High-Volume
390 Center Using an Aerosol-minimizing Protocol [pre-print]. medRxiv. doi: 10.1101/2020.08.30.20177543
- 391 [15] **Lyu W and Wehby GL** (2020) Community Use Of Face Masks And COVID-19: Evidence From A
392 Natural Experiment Of State Mandates In The US. *Health Affairs* 39: 1419-1425. doi:
393 10.1377/hlthaff.2020.00818
- 394 [16] **Eikenberry SE, Mancuso M, Iboi E, Phan T, Eikenberry K, et al.** (2020) To mask or not to mask:
395 Modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic.
396 *Infectious Disease Modelling* 5: 293-308. doi: 10.1016/j.idm.2020.04.001

- 397 [17] **Althouse BM, Wallace B, Case B, Scarpino SV, Berdhal A, et al.** (2020) The unintended
398 consequences of inconsistent pandemic control policies [pre-print]. medRxiv. doi:
399 10.1101/2020.08.21.20179473
- 400 [18] **Institute for Health Metrics and Evaluation** (2020) New IHME COVID-19 Forecasts See Nearly
401 300,000 Deaths by December 1, However, Consistent Mask-Wearing Could Save about 70,000 Lives.
402 Available at: [https://www.prnewswire.com/news-releases/new-ihme-covid-19-forecasts-see-nearly-300-](https://www.prnewswire.com/news-releases/new-ihme-covid-19-forecasts-see-nearly-300-000-deaths-by-december-1--however-consistent-mask-wearing-could-save-about-70-000-lives-301107858.html)
403 [000-deaths-by-december-1--however-consistent-mask-wearing-could-save-about-70-000-lives-](https://www.prnewswire.com/news-releases/new-ihme-covid-19-forecasts-see-nearly-300-000-deaths-by-december-1--however-consistent-mask-wearing-could-save-about-70-000-lives-301107858.html)
404 [301107858.html](https://www.prnewswire.com/news-releases/new-ihme-covid-19-forecasts-see-nearly-300-000-deaths-by-december-1--however-consistent-mask-wearing-could-save-about-70-000-lives-301107858.html) Published on: August 6, 2020
- 405 [19] **Tapp T** (2020) Donald Trump Asks Reporter To Take His Mask Off At News Conference After He
406 Asks About Atlantic Article. Available at: [https://deadline.com/2020/09/donald-trump-asks-reporter-to-](https://deadline.com/2020/09/donald-trump-asks-reporter-to-take-his-mask-off-at-news-conference-atlantic-story-1234572311/)
407 [take-his-mask-off-at-news-conference-atlantic-story-1234572311/](https://deadline.com/2020/09/donald-trump-asks-reporter-to-take-his-mask-off-at-news-conference-atlantic-story-1234572311/) Published on: September 07, 2020
- 408 [20] **Niedzwiadek N and Atterbury A** (2020) Colleges crack down on student behavior as virus
409 threatens more closures. Available at: [https://www.politico.com/news/2020/08/30/college-students-](https://www.politico.com/news/2020/08/30/college-students-coronavirus-closures-404567)
410 [coronavirus-closures-404567](https://www.politico.com/news/2020/08/30/college-students-coronavirus-closures-404567) Published on: August 30, 2020
- 411 [21] **Moreno JE** (2020) Sturgis Motorcycle Rally was 'superspreading event' that cost public health \$12.2
412 billion: analysis. Available at: [https://thehill.com/homenews/state-watch/515453-sturgis-motorcycle-](https://thehill.com/homenews/state-watch/515453-sturgis-motorcycle-rally-was-superspreading-event-that-cost-public)
413 [rally-was-superspreading-event-that-cost-public](https://thehill.com/homenews/state-watch/515453-sturgis-motorcycle-rally-was-superspreading-event-that-cost-public) Published on: September 8, 2020
- 414 [22] **Groves S and Kolpack D** (2020) Dakotas lead US in virus growth as both reject mask rules.
415 Available at: <https://www.yahoo.com/news/virus-rises-dakotas-freedom-argument-144101521.html>
416 Accessed on: 9/12/2020
- 417 [23] **The New York Times** (2020) Coronavirus (Covid-19) Data in the United States. Available at:
418 <https://github.com/nytimes/covid-19-data> Accessed on: September 22, 2020
- 419 [24] **Raifman J, Nocka K, Jones D, Bor J, Lipson S, et al.** (2020) COVID-19 US state policy database.
420 Available at: www.tinyurl.com/statepolicies Accessed on:
- 421 [25] **The Atlantic Monthly Group** (2020) COVID Tracking Project at the Atlantic. Available at:
422 <https://covidtracking.com/data/download> Accessed on: September 23, 2020
- 423 [26] **United States Census Bureau Population Division** (2019) Annual Estimates of the Resident
424 Population for the United States, Regions, States, and Puerto Rico: April 1, 2010 to July 1, 2019. 2010-
425 2019 Population Estimates. Available at: [https://www.census.gov/data/tables/time-](https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-total.html)
426 [series/demo/popest/2010s-state-total.html](https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-total.html) Accessed on: December 30, 2019
- 427 [27] **Bureau of Economic Activity** (2019) Gross Domestic Product by State, 4th Quarter and Annual
428 2019 (PDF). Available at: www.bea.gov Accessed on: April 7, 2020
- 429 [28] **Kaiser Family Foundation** (2018) Population Distribution by Age. Available at:
430 [https://www.kff.org/other/state-indicator/distribution-by-](https://www.kff.org/other/state-indicator/distribution-by-age/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D)
431 [age/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22](https://www.kff.org/other/state-indicator/distribution-by-age/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D)
432 [%7D](https://www.kff.org/other/state-indicator/distribution-by-age/?currentTimeframe=0&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D) Accessed on: December 04, 2019
- 433 [29] **Bureau of Labor Statistics** (2020) STATE EMPLOYMENT AND UNEMPLOYMENT --
434 JANUARY 2020. Available at: https://www.bls.gov/news.release/archives/laus_03162020.htm Accessed
435 on: August 29, 2020
- 436 [30] **Department of Housing and Urban Development** (2020) 2007 - 2019 PIT Counts by State
437 (XLSX). Available at: <https://www.hudexchange.info/resource/3031/pit-and-hic-data-since-2007/>
438 Accessed on: January 2020
- 439 [31] **Department of Housing and Urban Development** (2020) 2007 - 2019 HIC Data by State (XLSX).
440 Available at: <https://www.hudexchange.info/resource/3031/pit-and-hic-data-since-2007/> Accessed on:
441 January 2020
- 442 [32] **Kaeble D, Cowhig M and Bureau of Justice Statistics** (2016) Correctional Populations In The
443 United States, 2016. Available at: <https://www.bjs.gov/index.cfm?ty=pbdetail> Accessed on: August 29,
444 2020
- 445 [33] **Kaiser Family Foundation** (2020) Total Number of Residents in Certified Nursing Facilities.
446 Available at: [https://www.kff.org/other/state-indicator/number-of-nursing-facility-](https://www.kff.org/other/state-indicator/number-of-nursing-facility-residents/?currentTimeframe=0)
447 [residents/?currentTimeframe=0](https://www.kff.org/other/state-indicator/number-of-nursing-facility-residents/?currentTimeframe=0) Accessed on: August 29, 2020

- 448 [34] **United States Census Bureau** (2020) National Population Totals and Components of Change: 2010-
449 2019. Available at: [https://www.census.gov/content/census/en/data/tables/time-series/demo/popest/2010s-](https://www.census.gov/content/census/en/data/tables/time-series/demo/popest/2010s-national-total.html#par_textimage_2011805803)
450 [national-total.html#par_textimage_2011805803](https://www.census.gov/content/census/en/data/tables/time-series/demo/popest/2010s-national-total.html#par_textimage_2011805803) Accessed on: September 23, 2020
- 451 [35] **Bennefield R and Bonnette R** (2003) Structural and Occupancy Characteristics of Housing: 2000.
452 Census 2000 Brief. Available at: <https://www.census.gov/prod/2003pubs/c2kbr-32.pdf> Accessed on:
453 January 8, 2019
- 454 [36] **Kaiser Family Foundation** (2020) Population Distribution by Race/Ethnicity. Available at:
455 <https://www.kff.org/other/state-indicator/distribution-by-raceethnicity/?currentTimeframe=0> Accessed
456 on: August 29, 2020
- 457 [37] **US Congress Joint Economic Committee (JEC) Committee** (2018) The Geography of Social
458 Capital in America. Available at: [https://www.jec.senate.gov/public/index.cfm/republicans/2018/4/the-](https://www.jec.senate.gov/public/index.cfm/republicans/2018/4/the-geography-of-social-capital-in-america)
459 [geography-of-social-capital-in-america](https://www.jec.senate.gov/public/index.cfm/republicans/2018/4/the-geography-of-social-capital-in-america) Accessed on: August 29, 2020
- 460 [38] **Centers for Disease Control and Prevention (CDC)** (2018) CDC 2018 obesity map, Adult Obesity
461 Prevalence Maps. Available at: <https://www.cdc.gov/obesity/data/prevalence-maps.html> Accessed on:
462 August 29, 2020
- 463 [39] **270 to Win** (2020) 2020 Gubernatorial Elections Map. Available at:
464 <https://www.270towin.com/2020-governor-election/> Accessed on: September 05, 2020
- 465 [40] **Center for American Women in Politics (CAWP)** (2020) Women in Statewide Elective Executive
466 Office 2020. Available at: <https://cawp.rutgers.edu/women-statewide-elective-executive-office-2020>
467 Accessed on: September 05, 2020
- 468 [41] **Yarnold PR and Soltysik RC** (1991) Theoretical distributions of optima for univariate
469 discrimination of random data. *Decision Sciences* 22: 739-752. doi:
470 [42] **Yarnold PR and Soltysik RC** (2005) *Optimal data analysis: Guidebook with software for Windows*.
471 Washington, D.C.: APA Books. doi:
472 [43] **Yarnold PR** (2017) What is optimal data analysis? *Optimal Data Analysis* 6: 26-42. doi:
473 [44] **Soltysik RC and Yarnold PR** (1994) Univariable Optimal Discriminant Analysis: One-Tailed
474 Hypotheses. *Educational and Psychological Measurement* 54: 646-653. doi:
475 10.1177/0013164494054003007
- 476 [45] **Carmony L, Yarnold PR and Naeymi-Rad F** (1998) One-tailed Type I error rates for balanced
477 two-category UniODA with a random ordered attribute. *Annals of Operations Research* 74: 223-238. doi:
478 [46] **Rhodes NJ and Yarnold PR** (2020) Generating Novometric Confidence Intervals in R: Bootstrap
479 Analyses to Compare Model and Chance ESS. *Optimal Data Analysis* 9: 172-177. doi:
480 [47] **Yarnold PR and Soltysik RC** (2016) *Maximizing Predictive Accuracy*. ODA Books. doi:
481 10.13140/RG.2.1.1368.3286
- 482 [48] **Rhodes NJ** (2020) *ODA: a package and R-interface for the MegaODA software suite*. R package
483 version 1.0.1.3. doi:
484 [49] **Wickham H** (2016) *ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag. doi:
485 [50] **Rhodes NJ** (2020) *Statistical Power Analysis in ODA, CTA and Novometrics*. *Optimal Data*
486 *Analysis* 9: 21-25. doi:
487 [51] **Hendrix M** (2018) *The Surprising Geography of Social Capital in America*. Available at:
488 [https://medium.com/@michael_hendrix/the-surprising-geography-of-social-capital-in-america-](https://medium.com/@michael_hendrix/the-surprising-geography-of-social-capital-in-america-c6cef6bae50b)
489 [c6cef6bae50b](https://medium.com/@michael_hendrix/the-surprising-geography-of-social-capital-in-america-c6cef6bae50b) Published on: June 29, 2018
- 490 [52] **Yarnold PR** (2016) Characterizing and Circumventing Simpson's Paradox for Ordered Bivariate
491 Data. *Educational and Psychological Measurement* 56: 430-442. doi: 10.1177/0013164496056003005
- 492 [53] **Anonymous** (2020) Covid-19: New fear grips Europe as cases top 30m worldwide. Available at:
493 <https://www.bbc.com/news/world-54199825> Published on: 9/18/2020
- 494 [54] **Scott D** (2020) These 4 Midwestern states are seeing worrying Covid-19 spikes. Available at:
495 <https://www.vox.com/2020/9/2/21418812/covid-19-coronavirus-us-cases-midwest-surge> Published on:
496 September 2, 2020
- 497 [55] **Yarnold PR** (2013) Ascertaining an individual patient's symptom dominance hierarchy: Analysis of
498 raw longitudinal data induces Simpson's Paradox. *Optimal Data Analysis* 2: 159-171. doi:

499 [56] **Kretzschmar ME, Rozhnova G, Bootsma MCJ, van Boven M, van de Wiggert JHHM, et al.**
500 (2020) Impact of delays on effectiveness of contact tracing strategies for COVID-19: a modelling study.
501 The Lancet Public Health 5: e452-e459. doi: 10.1016/s2468-2667(20)30157-2
502 [57] **Tversky A and Kahneman D** (1974) Judgment under Uncertainty: Heuristics and Biases. Science
503 185: 1124-1131. doi: 10.1126/science.185.4157.1124
504 [58] **Holcombe M, Yan H and Waldrop T** (2020) 'Mix of science and politics' leading to people's
505 uncertainty about Covid-19 vaccine, NIH director says. Available at:
506 <https://www.cnn.com/2020/09/16/health/us-coronavirus-wednesday/index.html> Published on: September
507 17, 2020

508

Table 1: State-Wise Descriptive Statistics for Number of New Cases Pre- vs. Post-Mask Mandate (*MM*)

State	<u>Before <i>MM</i></u>		<u>After <i>MM</i></u>	
	Mean	SD	Mean	SD
AK	9.8	4.8	2.4	2.0
AL	229.4	64.1	390.3	158.0
AR	93.7	75.4	205.0	114.8
AZ	245.6	88.2	547.7	355.1
CA	1417.9	390.6	2147.9	464.8
CO	338.6	81.2	421.1	148.5
DC	68.0	52.3	155.3	52.7
DE	147.2	100.1	153.7	69.2
FL	754.3	263.8	863.7	305.2
GA	704.1	315.6	653.4	218.1
HI	17.1	9.4	3.4	4.3
IL	1564.2	504.7	2182.9	668.9
IN	530.7	134.4	559.0	132.2
KY	159.5	87.9	178.8	85.7
LA	760.2	606.1	399.2	237.0
MA	1925.7	718.4	1041.5	660.5
MD	385.2	292.1	911.0	240.9
ME	26.4	12.6	39.7	16.7
MI	1147.5	344.6	572.2	255.6
MN	637.7	116.4	383.6	101.5
MS	222.9	68.5	277.7	81.8
NE	179.6	152.3	296.1	109.6
NH	59.3	17.4	80.8	29.6
NJ	1480.3	1528.0	2936.7	808.8
NM	117.1	47.0	146.3	51.3
NV	120.8	38.8	124.9	42.7
NY	7479.5	2836.8	4170.2	2206.4
OH	503.5	288.7	574.1	106.9
OR	61.0	14.9	59.9	29.0
PA	1051.8	605.8	1150.4	355.1
RI	138.1	120.7	278.0	87.1
TX	923.8	222.6	1266.6	380.9
UT	65.9	56.0	137.8	34.7
VA	933.2	243.9	656.0	230.7
VT	25.1	16.2	5.5	5.2
WA	299.2	132.8	252.7	108.5
WV	31.8	17.7	28.4	19.5
WY	14.6	12.9	11.1	6.6

Legend: SD=Standard Deviation

Table 2: Maximum-Accuracy Confirmatory Models Assessing State-Wise Mask Mandate (*MM*) Effectiveness

State	Pre- <i>MM</i>	Sensitivity (%) ^a	Post- <i>MM</i>	Specificity (%) ^b	Training		LOO ^d	
					<i>p</i> <	ESS (%) ^c	<i>p</i> <	ESS (%) ^c
MN	>480 Cases	28/30 (93.3)	≤480 Cases	27/31 (87.1)	0.0001	80.4	0.0001	73.9
AK	>4 Cases	27/30 (90.0)	≤4 Cases	28/31 (90.3)	0.0001	80.3	0.0001	80.3
HI	>6 Cases	26/30 (86.7)	≤6 Cases	28/31 (90.3)	0.0001	77.0	0.0001	64.0
MA	>1174 Cases	29/30 (96.7)	≤1174 Cases	23/31 (74.2)	0.0001	70.9	0.0001	64.3
VT	>14 Cases	23/30 (76.7)	≤14 Cases	29/31 (93.6)	0.0001	70.2	0.0001	66.8
VA	>691 Cases	27/29 (93.1)	≤691 Cases	23/31 (74.2)	0.0001	67.3	0.0001	67.3
MI	>732 Cases	26/30 (86.7)	≤732 Cases	24/31 (77.4)	0.0001	64.1	0.0001	50.8
NY	>4762 Cases	26/30 (86.7)	≤4762 Cases	22/31 (71.0)	0.0001	57.6	0.0001	51.0
LA	>340 Cases	26/30 (86.7)	≤340 Cases	17/30 (56.7)	0.0027	43.3	0.0037	36.7
WA	>294 Cases	17/30 (56.7)	≤294 Cases	25/31 (80.6)	0.011	37.3	0.0028	37.3
WY	>15 Cases	13/30 (43.3)	≤15 Cases	27/31 (87.1)	0.036	30.4	0.04	24.0
PA	>1389 Cases	12/30 (40.0)	≤1389 Cases	26/31 (83.4)	0.16	23.9	0.07	20.5
OR	>41 Cases	28/30 (93.3)	≤41 Cases	8/31 (25.8)	0.25	19.1	0.23	9.5
GA	>886 Cases	8/29 (27.6)	≤886 Cases	28/31 (90.3)	0.34	17.9	0.21	11.0
WV	>36 Cases	11/30 (36.7)	≤36 Cases	25/31 (80.6)	0.31	17.3	0.18	14.0
DE	>203 Cases	6/30 (20.0)	≤203 Cases	28/31 (90.3)	0.65	10.3	0.34	7.0
OH	>768 Cases	4/30 (13.3)	≤768 Cases	30/31 (96.8)	0.68	10.1	0.30	6.8
KY	>64 Cases	28/30 (93.3)	≤64 Cases	5/31 (16.1)	0.7	9.5	0.30	9.5
IN	>396 Cases	27/30 (90.0)	≤396 Cases	6/31 (19.4)	0.72	9.4	0.38	6.1
NV	>189 Cases	3/30 (10.0)	≤189 Cases	30/31 (96.8)	0.81	6.8	0.99	-77.1
NE	>455 Cases	3/30 (10.0)	≤455 Cases	30/31 (96.8)	0.81	6.8	0.82	-3.0
AL	>107 Cases	30/30 (100)	≤107 Cases	2/31 (6.4)	0.84	6.4	0.52	3.1
NJ	>4296 Cases	1/30 (3.33)	≤4296 Cases	31/31 (100)	0.94	3.3	0.88	-3.1
TX	>477 Cases	30/30 (100.0)	≤477 Cases	1/31 (3.2)	0.96	3.2	0.51	3.2
FL	>534 Cases	25/30 (83.3)	≤534 Cases	6/31 (19.4)	0.97	2.7	0.77	-3.9
NM	>216 Cases	2/30 (6.7)	≤216 Cases	29/31 (93.6)	0.98	0.2	0.94	-6.3
MS	>138 Cases	28/30 (93.3)	≤138 Cases	2/31 (6.4)	0.99	-0.2	0.95	-6.8
UT	>176 Cases	2/30 (6.7)	≤176 Cases	28/31 (90.3)	0.99	-3.0	0.99	-12.8
DC	>224 Cases	1/30 (3.3)	≤224 Cases	29/31 (93.6)	0.99	-3.1	0.88	-3.1
MD	>1154 Cases	1/30 (3.3)	≤1154 Cases	29/31 (93.6)	0.99	-3.1	0.94	-6.3
CO	>210 Cases	28/30 (93.3)	≤210 Cases	1/31 (3.2)	0.99	-3.4	0.89	-3.4
NH	>34 Cases	28/30 (93.3)	≤34 Cases	1/31 (3.23)	0.99	-3.4	0.95	-6.8
ME	>61 Cases	1/30 (3.3)	≤61 Cases	27/31 (87.1)	0.99	-9.6	0.99	-93.4
IL	>1133 Cases	26/30 (86.7)	≤1133 Cases	1/31 (3.2)	0.99	-10.1	0.98	-10.1
AR	>34 Cases	24/30 (80.0)	≤34 Cases	1/31 (3.2)	0.99	-16.8	0.99	-20.1
AZ	>161 Cases	24/30 (80.0)	≤161 Cases	1/31 (3.2)	0.99	-16.8	0.99	-20.1
RI	>273 Cases	9/30 (30.0)	≤273 Cases	16/31 (51.6)	0.99	-18.4	0.99	-41.7
CA	>2315 Cases	2/30 (6.7)	≤2315 Cases	21/31 (67.7)	0.99	-25.6	0.99	-32.2

Footnotes: a) Sensitivity gives the results for the predicted (numerator) and the actual (denominator) number of new cases after *MM*. b) Specificity gives the results for predicted (numerator) and actual (denominator) number of new cases before *MM*. c) ODA models explicitly maximize classification (i.e., predictive) accuracy adjusted for chance, which is assessed

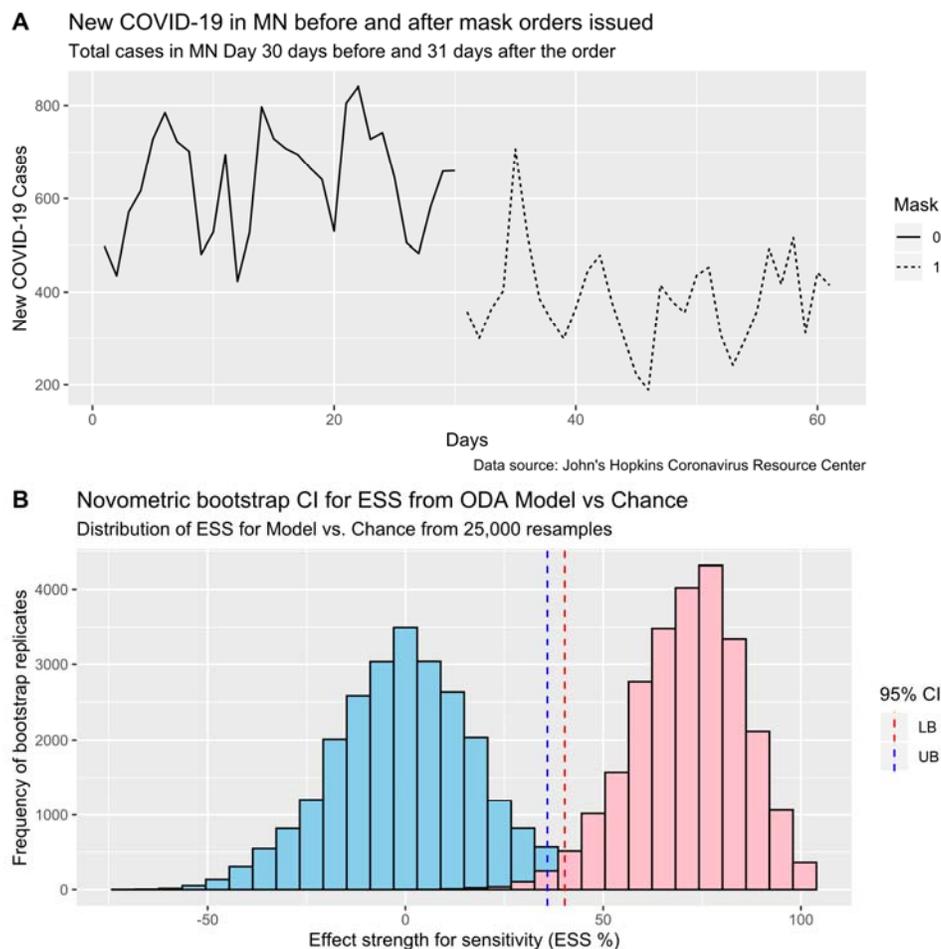
by the effect strength for sensitivity (ESS) statistic: $ESS=0$ is the accuracy expected by chance; $ESS<25$ is a relatively weak effect; $ESS<50$ is a moderate effect; $ESS<75$ is a relatively strong effect; $ESS<90$ is a strong effect; and $100\leq ESS<0$ is accuracy which is worse than expected by chance. In this column, MN has a training model with an ESS that provides the strongest support for the *a priori* hypothesis and is statistically significant whereas CA has a training model that provides the least support for this hypothesis and is not significant. d) Classification performance in validity analysis was computed using a one-sample leave-one-out (LOO) jackknife validity assessment. In this column, AK has a validity model with an ESS that provides the strongest support for the *a priori* hypothesis and is statistically significant whereas ME has a validity model that provides the least support for this hypothesis and is not significant.

Table 3: LOO Confirmatory p , and Bootstrap 95% CI for LOO ESS, for 11 States Demonstrating Mask Mandate (MM) Effectiveness vis-à-vis Training $p < 0.05$

State	$p \leq$	Chance 95% CI ^a	Model 95% CI ^a	95% CI overlap ^b
AK	0.0001	-36.1 – 35.9	64.6 – 100	No
VA	0.0001	-35.8 – 35.9	49.9 – 100	No
MA	0.0001	-35.0 – 34.8	41.7 – 92.9	No
MN	0.0001	-35.9 – 36.1	40.3 – 95.2	No
HI	0.0001	-36.1 – 35.9	34.9 – 91.7	Yes
VT	0.0001	-35.7 – 34.9	30.3 – 86.7	Yes
NY	0.0001	-36.1 – 35.8	25.5 – 86.1	Yes
LA	0.0037	-34.8 – 35.7	25.0 – 84.6	Yes
MI	0.0001	-35.8 – 36.1	11.1 – 76.6	Yes
WA	0.0028	-34.4 – 35.0	0.00 – 68.4	Yes
WY	0.04	-33.3 – 33.5	-10.1 – 58.3	Yes

Footnotes: a) Exact Discrete 95% Confidence Intervals (CIs) for Model and Chance ESS were estimated using bootstrap resampling of the ODA model confusion matrix from each respective State obtained in LOO analysis. b) For a given State, if the upper bound (UB) of the CI for chance falls *beneath* the lower bound (LB) of the CI for the model, this column indicates NO 95% CI overlap—so the confirmatory hypothesis is *accepted* (i.e., the MM reduced the number of new cases). And, for a given State, if the UB of the CI for chance falls *above* the LB of the CI for the model, this column indicates YES 95% CI overlap—so the confirmatory hypothesis is *rejected* (i.e., the MM did not reduce the number of new cases). In other words, if the exact discrete CIs for model and chance do not overlap, then the ODA model identified in LOO analysis is considered statistically robust and the findings are likely to be generalizable.

Figure 1. Representative state (MN) with a strong impact of Mask Mandate (*MM*) orders



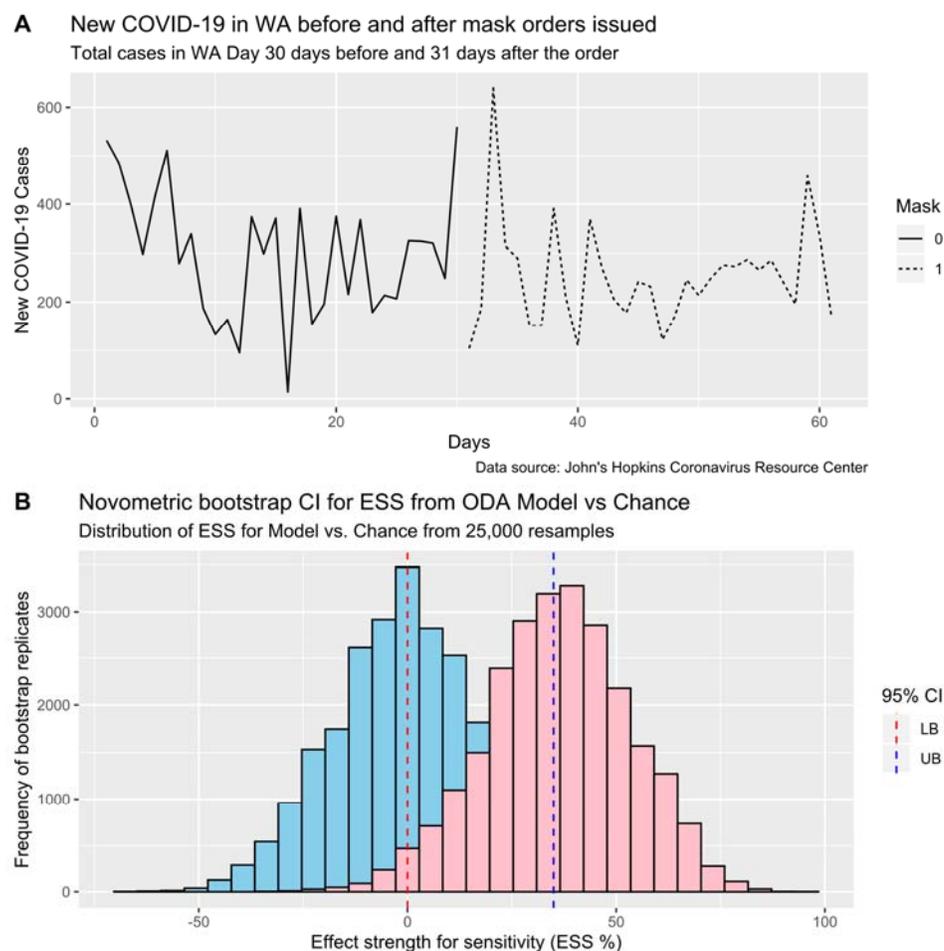
Legend:

Panel A. Number of new COVID-19 cases reported within MN daily in the month before (solid black line) vs. daily in the month after (broken black line) *MM* orders were issued.

Panel B. Distribution of bootstrap replicates for model and chance for the LOO analysis for MN by ODA. Here, the ODA model was resampled using bootstrap resampling with replacement ($n=25,000$ resamples with 50% replacement) for the LOO model confusion matrix (see Table 3) as well as for chance (i.e., the final model predictions randomly scrambled relative to the class variable). The lack of overlap between the Chance UB and Model LB is denoted by the alignment (rather than reversal) of the dashed lines with the corresponding Chance (blue) and Model (red) bootstrap distributions.

Abbreviations: LB = lower bound of the 95% CI (i.e., 2.5th percentile); UB = upper bound of the 95% CI (i.e., 97.5th percentile); LOO = leave-one-out one-sample jackknife.⁴²

Figure 2. Representative state (WA) with a moderate impact of Mask Mandate (*MM*) orders



Legend:

Panel A. Number of new COVID-19 cases reported within WA daily in the month before (solid black line) vs. daily in the month after (broken black line) *MM* orders were issued.

Panel B. Distribution of bootstrap replicates for model and chance for the LOO analysis for WA by ODA. Here, the ODA model was resampled using bootstrap resampling with replacement ($n=25,000$ resamples with 50% replacement) for the LOO model confusion matrix (see Table 3) as well as for chance (i.e., the final model predictions randomly scrambled relative to the class variable). The overlap between the Chance UB and Model LB is denoted by the reversal (rather than alignment) of the dashed lines with the corresponding Chance (blue) and Model (red) bootstrap distributions.

Abbreviations: LB = lower bound of the 95% CI (i.e., 2.5th percentile); UB = upper bound of the 95% CI (i.e., 97.5th percentile); LOO = leave-one-out one-sample jackknife.⁴²