Lower Intent to Comply with COVID-19 Public Health Recommendations Correlates to Higher Disease Burden in Following 30 Days

Robert P. Lennon, MD, JD, Aleksandra E. Zgierska, MD, PhD, Erin L. Miller, BS, Bethany Snyder, MPH, Aparna Keshaviah, ScM, Xindi C. Hu, ScD, Hanzhi Zhou, PhD, and Lauren Jodi Van Scoy, MD

Objectives: We sought to determine whether self-reported intent to comply with public health recommendations correlates with future coronavirus disease 2019 (COVID-19) disease burden.

Methods: A cross-sectional, online survey of US adults, recruited by snowball sampling, from April 9 to July 12, 2020. Primary measurements were participant survey responses about their intent to comply with public health recommendations. Each participant’s intent to comply was compared with his or her local COVID-19 case trajectory, measured as the 7-day rolling median percentage change in COVID-19 confirmed cases within participants’ 3-digit ZIP code area, using public county-level data, 30 days after participants completed the survey.

Results: After applying raking techniques, the 10,650-participant sample was representative of US adults with respect to age, sex, race, and ethnicity. Intent to comply varied significantly by state and sex. Lower reported intent to comply was associated with higher COVID-19 case increases during the following 30 days. For every 3% increase in intent to comply with public health recommendations, which could be achieved by improving average compliance by a single point for a single item, we estimate a 9% reduction in new COVID-19 cases during the subsequent 30 days.

Conclusions: Self-reported intent to comply with public health recommendations may be used to predict COVID-19 disease burden. Measuring compliance intention offers an inexpensive, readily available method of predicting disease burden that can also identify populations most in need of public health education aimed at behavior change.

Key Words: coronavirus, COVID-19, epidemiology, public health, SARS-CoV-2

Efforts to forecast coronavirus disease 2019 (COVID-19) case burden have been largely unsuccessful, and draconian measures predicated on flawed models have negatively affected economic and health issues beyond COVID-19. Widespread gaps in diagnostic testing have impeded population sampling as an efficient, reliable means to predict disease burden. Complex predictive techniques for population testing (eg, wastewater surveillance) show promise in providing predictive insights by mapping population disease burden, but require further methodological validation.

As COVID-19 data have increased, modeling efforts have improved; however, even the most sophisticated modeling remains limited by its requirements of implicit assumptions about human behavior, and the need to identify specific compliance parameters to test the model. For example, Reiner et al recently demonstrated the power of universal mask wearing using state-level US data and five susceptible-exposed-infectious-recovered models. They make a compelling case for the effectiveness of 95% mask use in public to ameliorate projected increases in case counts as public

Key Points
- Efforts to forecast coronavirus disease 2019 case burden have been largely unsuccessful.
- Comparing actual public behaviors with actual subsequent disease outcomes offers a pragmatic solution to predicting future disease burden.
- Electronic survey of public intent to comply with public health recommendations offers an inexpensive, rapid method of predicting coronavirus disease 2019 disease burden that may be applied to other ongoing or emergent health threats or used as a method to measure the success of public health education campaigns.
health–mandated restrictions are removed. Although their modeling is arguably the best published to date with a mean absolute percent error (MAPE) nearly 40% lower than the average COVID-19 model, it still has a MAPE of >20%. Also, their predetermined mask usage rates are 95% and 85%, which are markedly higher than the observed rate of 59% in US residents (https://covid19.healthdata.org). Furthermore, although we often think of the requirement of a “limited number of high-priority, evidence-based interventions” necessary for effective public health programs as a limiting factor, in this case it is an expansive one. Any one of the COVID-19 mitigation strategies may, with 95% adherence, show good results in a model. Effective public health policy on mitigation strategies should consider the effect of improvements across them all, and be able to demonstrate the impact of even modest improvements in compliance.

To help guide public health policy regarding infection control policies, we show the correlation between self-reported intent to comply with several public health recommendations and local COVID-19 disease burden within the following 30 days based on survey results from a national, demographically representative sample of adults in the United States and their actual, local COVID-19 case rates.

Methods

Design

We administered a cross-sectional online survey to US adults between April 9 and July 12, 2020. Participants were recruited using online snowball techniques leveraging social media and networks. The survey, refined from previous work and described in detail elsewhere, asked a series of questions related to COVID-19, including understanding of, and intent to comply with, public health recommendations, measured on a 5-point Likert scale (1 = “certainly not,” 5 = “most certainly”). The survey also collected geographic and demographic information, including three-digit ZIP code, age, sex, race, ethnicity, education, and health and social status, using the MacArthur Scale of Subjective Social Status. This study was deemed exempt by the Pennsylvania State University College of Medicine institutional review board. Written consent was obtained from all study participants.

Outcome Measures

Survey respondents indicated their intent to comply with the main public health recommendations put forth by the US Centers for Disease Control and Prevention (CDC) at the time of the survey: “Wash your hands often (for 20 seconds or more),” “Maintain social distancing/social isolation even if you have no symptoms,” “Avoid touching your eyes, nose, and mouth,” “Cough or sneeze into your elbow,” and “Stay at home if you feel unwell. If you have a fever, cough, and difficulty breathing seek medical attention and call in advance.” The survey was updated on June 5, 2020 (version 2) to reflect changes in the CDC public health recommendations: “Wash your hands often (for 20 seconds or more),” “Wear a cloth face cover (face mask) when out in public,” “Avoid touching your eyes, nose, and mouth with unwashed hands,” “Cover your mouth and nose with a tissue of the inside of your elbow when you cough or sneeze,” “Stay at home if you feel unwell,” “If you have a fever, cough, and difficulty breathing seek medical attention and call in advance,” “Stay at least 6 feet (2 meters, or about 2 arms’ lengths) from other people when outside of your home,” and “Stay out of crowded places and avoid mass gatherings.”

Analytical Methods

Analysis was conducted on individual “intent to comply” questions and also on an intent to comply index composite score, which was a weighted average of Likert scores for each individual question, weighted based on results from exploratory factor analyses generated for each version of the survey.

The trajectory of the number of COVID-19 cases during the 30, 14, and 7 days before and after a respondent’s survey completion date was based on each respondent’s local area, derived from his or her reported three-digit ZIP code. We chose the three-digit ZIP code as the spatial unit of analysis because this is the most granular spatial information we have from the survey. To characterize the COVID-19 case trajectory in the area in which each participant was living, we aligned county-level COVID-19 case data to the 3-digit ZIP code level. To do so, we first identified counties whose centroids fall within the boundary of the 3-digit ZIP code area, and then took a weighted average of COVID-19 cases in these counties. For 3-digit ZIP code areas that were too small to contain the centroid of a county, we spatially joined the boundary of the 3-digit ZIP code area with the boundary of the counties and identified the county that overlapped the most with the 3-digit ZIP code area. Participants from the same 3-digit ZIP code area had the same value for COVID-19 case trajectory. We did not stratify our analysis by 3-digit ZIP code and kept the model at the individual level. County-level COVID-19 data from the Johns Hopkins University Center for Systems Science and Engineering (https://github.com/CSSEGISandData/COVID-19) was used to calculate a 7-day rolling average of cases; new cases were identified as the difference between “today’s” 7-day rolling average and “yesterday’s” 7-day rolling average. Daily percentage change was calculated as the difference between new cases “today” and new cases “yesterday” over the absolute value of new cases “yesterday.” Lastly, cases in which the percentage change was undefined were filtered out (an artifact of dividing by 0 new cases “yesterday”), and a rolling median of 7-day percentage change was calculated.

Geospatial methods were used to interpolate the county-level results to the 3-digit ZIP code areas. First, the centroid (geometric center) for each county was calculated and assigned the 7-day percentage change to that point location. Second, a spatial join between county centroids and 3-digit ZIP code area boundaries was conducted to identify county centroids that fall within the regions of
each 3-digit ZIP code area. If multiple county centroids fell in the same 3-digit ZIP code area, then a weighted average of the 7-day percent change in disease burden, weighted using the estimated county population from the 2019 American Community Survey, was used. In some cases, no county centroid fell within a 3-digit ZIP code area, possibly because the ZIP code area was small and located in the corner of a county. In those cases, a spatial join of the 3-digit ZIP code area and county polygons was conducted and assigned the largest overlapping region to the 3-digit ZIP code area. Even after implementing these two steps, approximately 80 3-digit ZIP code areas were missing values for the 7-day percentage change in disease burden. For these 3-digit ZIP code areas, a nearest-neighbor search was conducted to identify the 5 closest county centroids (<100 km) and to calculate the weighted average percentage change in disease burden. After applying these methods, the case trajectory for the 30 days following survey response for 680 of the 728 3-digit ZIP code areas in which our participants reside was possible.

To generalize survey responses to the US population, raking techniques were used to calibrate the data via the calibrate() function from the survey package in R software for statistical analysis. Raking requires complete data, so missing values were imputed randomly, based on observed variable distribution. In general, the rate of missingness was extremely low; ethnicity was missing for 6.5% of respondents, but for all of the other variables used in the raking, <2% of respondents had a missing value. Specifically, raking was used to develop weights that aligned survey respondent distributions for demographics (age, sex, race, ethnicity, and education), geography (state), and presence of comorbidities known to increase the health risks of COVID-19 infection (eg, diabetes mellitus, heart disease) to national benchmarks. National-level weights and state-level weights for the 20 states with the largest numbers of survey respondents were generated. We chose to weight at the national and state levels based on feasibility and bias-variance tradeoff considerations. With respect to feasibility, we could obtain reliable population benchmarks for the characteristics used for raking only at the national level and individual state level. With respect to the bias-variance tradeoff, we were able to achieve a good balance between bias reduction and variance inflation (due to weighting effect) only at the individual state level or higher; at the 3-digit ZIP code or county level, sample sizes were too small to support reliable weights. These weights were applied when summarizing the intent to comply with public health recommendations scores and when generating models to assess the relation between the intent to comply and future COVID-19 disease caseload. Survey-weighted linear regression models were fitted with the svyglm() function from the R survey package that weights the importance of each case to make them representative of the entire US population in the national model or of the state population in the state-level model and properly accounts for this weighting effect when calculating standard errors of the regression coefficients.

All of the analyses used a 5% type I error rate and were completed using the R statistical package. This study adheres to the Strengthening the Reporting of Observational Studies in Epidemiology guidelines and was approved by an institutional review board.

**Results**

Of the 10,650 respondents, most self-identified as female, White, and younger than age 64 (Table). Raking generated a representative sample whose weighted demographic, geographic, and health status features matched national benchmarks (Table). Within this demographically representative sample, the weighted average intent to comply score was 93.18 out of 100 (range 20–100), indicating an overall high intent to comply. Female respondents reported a higher intent to comply than male respondents (Figure 1). Although this analysis focuses on the correlation of weighted average intent to comply score across all recommended behaviors, for comparison to other studies, deaggregated values of respondents’ intent to “most certainly” or “probably yes” comply with self-isolation (94%), social distancing (89%), and wearing a face mask (87%) are reported.

There are statistically significant differences in intent to comply scores by state (Figure 2A). A trend toward higher intent to comply scores in older adults and in individuals with chronic medical conditions was observed, but not significant (0.05 < P < 0.1). Differences in the intent to comply by race/ethnicity, education, geographic region, or social status also were not statistically significant. Case trajectories over the 30 days following survey completion for 680 of the 728 3-digit ZIP code areas represented by the survey respondents were computed and consolidated into state-level trajectories (Figure 2B). On average, the smoothed daily case count at the 3-digit ZIP code level increased by 0.14% (range −10% to 13.5%).

Intent to comply correlated with subsequent COVID-19 case trajectory nationally and within certain states (Figure 3). Overall, the regression coefficient capturing the effect of intent to comply on future disease burden was −0.029 (95% confidence interval −0.049 to −0.009), after accounting for differences in demographics, region, social status, health conditions that increase risk from COVID-19, and trust in information sources. In other words, these findings suggest that for every 3% increase in the composite intent to comply score, which could be achieved by increasing the average compliance rating by a single point (eg, from “maybe” to “probably yes”) for just one of the compliance items (eg, avoid touching your eyes, nose, and mouth with unwashed hands), 9% of COVID-19 cases could be prevented during the subsequent 30 days.

**Discussion**

This is one of the largest studies of self-reported intent to comply with COVID-19 public health recommendations, and, to our knowledge, the first to indicate that the public’s intent correlates with future COVID-19 case trajectory. In general, the modeling results showed that the more people intend to comply with public health recommendations, the fewer COVID-19 cases were reported in their community. This is true for the United States.
as a whole and for Colorado, Michigan, New Jersey, New York, and Pennsylvania in particular, where intent to comply was statistically negatively correlated with future COVID-19 disease rates. One possible explanation for a stronger negative association in those states compared with others is that those states were affected early by the pandemic, and by the time of the survey, residents in those states may have already begun to realize the importance of mitigation measures to reduce COVID-19 spread, and case increases may have already started to slow.

Compliance intent in this demographically representative US sample is similar to our previously reported preliminary data.7 This also is similar to what was reported by Lennon et al.13 Their focus was on low “most certain” compliance for individual recommendations; however, their aggregate average US compliance across behaviors for “most certainly” and “probably yes” was 95.8%.13 Because we include “maybe,” “probably not,” and “definitely not” for a weighted average, it is not surprising that our comparable US aggregate average compliance of 93.18% is slightly lower. Average compliance of 93.18% is higher than individual behavior intent reported by Czeisler et al.14 Their study of US adults reported 74.1% “always or often” wore face coverings, 79.5% “always or often” practiced social distancing, and 77.3% practiced self-isolation14; however, Czeisler and colleagues’ urban respondents (New York City and Los Angeles) reported that 89.7% wore face coverings, 84.2% practiced social distancing, and 83.8% practiced self-isolation.14 These are more similar to our comparable deaggregated intentions to “most certainly” or “probably yes” compliance intentions: 87% for wearing a face mask, 89% for social distancing, and 94% for self-isolation. Differences are partly the result of an aggregation effect; in our

Table. Respondent demographic and COVID-19 health risk characteristics: unweighted (survey respondents, N = 10,650) and weighted (to generate a representative US sample)

| Characteristics                      | Unweighted % (SE) | Weighted % (SE) | Benchmark |
|--------------------------------------|-------------------|-----------------|-----------|
| Age, y, mean                         | 47.8 (0.1)        | 49.7 (0.7)      | NA        |
| 18–64                                | 85.7 (0.3)        | 80.2 (1.6)      | 80.2      |
| ≥65                                  | 14.3 (0.3)        | 19.8 (1.6)      | 19.8      |
| Sex                                  |                   |                 |           |
| Female                               | 75.4 (0.4)        | 51.3 (2.2)      | 51.3      |
| Male                                 | 24.6 (0.4)        | 48.7 (2.2)      | 48.7      |
| Ethnicity                            |                   |                 |           |
| Hispanic                             | 3.6 (0.2)         | 17.8 (2.3)      | 17.8      |
| Non-Hispanic                         | 96.4 (0.2)        | 82.2 (2.3)      | 82.2      |
| Race                                 |                   |                 |           |
| White                                | 88.0 (0.3)        | 75.1 (2.1)      | 75.1      |
| African American                     | 2.9 (0.2)         | 14.1 (1.8)      | 14.1      |
| Other                                | 9.1 (0.3)         | 10.8 (1.5)      | 10.8      |
| Education                            |                   |                 |           |
| High school diploma or less          | 4.4 (0.2)         | 40.0 (2.4)      | 40.0      |
| Some college or associate degree     | 16.4 (0.4)        | 31.0 (1.8)      | 31.0      |
| Bachelor’s degree                    | 34.9 (0.5)        | 18.4 (1.2)      | 18.4      |
| Graduate degree                      | 44.3 (0.5)        | 10.6 (0.6)      | 10.6      |
| Region                               |                   |                 |           |
| Midwest                              | 28.6 (0.4)        | 20.8 (1.4)      | 20.7      |
| Northeast                            | 37.7 (0.5)        | 17.4 (1.5)      | 17.4      |
| South                                | 21.3 (0.4)        | 38.1 (2.3)      | 38.1      |
| West                                 | 12.5 (0.3)        | 23.8 (1.8)      | 23.8      |
| Lower perceived social status: 1–6   | 32.3 (0.5)        | 50.9 (2.3)      | NA        |
| Higher perceived social status: 7–10 | 67.7 (0.5)        | 49.1 (2.3)      | NA        |
| COVID-19 vulnerability                |                   |                 |           |
| No health conditions                 | 77.0 (0.4)        | 71.5 (1.8)      | NA        |
| ≥1 health conditions                 | 23.0 (0.4)        | 28.5 (1.8)      | NA        |

During modeling, all characteristics shown were controlled as potential confounders.

Perceived social status was measured using the MacArthur Scale of Subjective Social Status developed by Adler et al.8 COVID-19 vulnerability was based on the presence or absence of health conditions (heart disease, diabetes mellitus, lung disease, or any other condition that impairs immune function) identified by the CDC as risk factors for severe COVID-19 presentation. National and state benchmarks for demographic characteristics came from the 2014–2018 American Community Survey 5-year estimates and 5-year average, respectively; national and state benchmarks for individual health conditions (including prevalence of diabetes mellitus, and prevalence of heart disease) came from the CDC’s Behavioral Risk Factor Surveillance System (averaged across 2014–2018). CDC, Centers for Disease Control and Prevention; COVID-19, coronavirus disease 2019; NA, not applicable; SE, standard error.
comparable sample, 97% endorsed practicing cough etiquette and 96% endorsed staying home if unwell, raising our aggregate compliance. Also, at the time Czeisler et al collected data (May 5–12, 2020), the CDC recommended against general mask wearing as a mitigation strategy. In our study, mask-wearing data were collected only after the CDC endorsed general mask wearing in public. CDC

Fig. 1. Intent to comply with public health recommendation score by demographic, health, and geographic characteristics.

Fig. 2. Intent to comply with public health recommendations (A) and COVID-19 case trajectory (B) by states in the United States. The maps summarize these findings by state; the generated models were at the individual survey respondent level. COVID-19, coronavirus 2019.
opposition may have driven Czeisler and coworkers’ intentions lower, and publicity around the change may have driven our intentions higher. Our method of recruitment via snowball sampling and post hoc raking to achieve demographic representation also may have contributed to this difference. Given how much more similar our values are to Czeisler and colleagues’ urban data, our methods likely oversampled urban respondents to achieve demographic representation. Although this suggests some limitation in applying our aggregate data to nonurban centers, our key finding is not the intended compliance itself, but rather how intended compliance correlates to actual health outcomes.

Public health response to a novel pathogen pandemic is largely driven by prediction modeling. A systematic review of early COVID-19 models showed high bias, suggesting that the models would likely fare worse than expected. As the high bias values predicted, COVID-19 models have been largely unsuccessful. Even the best models remain limited, with MAPEs >20% and assumptions that are unrealistic (eg, 95% mask compliance). In contrast to these prediction models, our dependent variable was a rate that could take zero values. Also, rather than using the model for prediction, our major focus was on explaining the association between intent to comply and future COVID-19 burden, controlling for an array of respondent characteristics available from the survey and rigorously weighting the data to make the observed association generalizable to the population. For this reason, we chose to look at the $R^2$ in-sample goodness-of-fit statistic rather than the MAPE, which measures out-of-sample forecast accuracy. Although the substantial number of zero values for our dependent variable prohibited us from calculating a MAPE measure to compare with existing models in the limited literature, an $R^2$ of 17% for our model indicates an adequately fitted linear regression model. In that sense, our model can be used to supplement other prediction models by providing a lens into factors that explain differences in COVID-19 case trajectories.

Our approach avoids testing limitations and reliance on a priori levels of compliance with single mitigation behaviors. This enables us to identify the impact of overall local compliance on local disease burden, which in turn enables local public health decisions to be tailored to the specific needs of a given community. Communities can look for particular low-compliance behaviors or select behaviors easiest to remedy in an effort to improve overall compliance. Improving average compliance by a single point for a single item increases the composite intent to comply score by 3%, and is estimated to yield a 9% decrease COVID-19 cases during the subsequent 30 days.

Study limitations inherent to online, cross-sectional surveys include inability to verify the veracity of responses, assess actual compliance, and represent people without Internet access. Generalizability also may be limited by the fact that survey responses reflect a single moment in time per respondent. Social desirability bias also may have influenced responses. We would, however, expect this to change over time, and sensitivity analysis did not show a meaningful difference between an April-only model and the full-dataset model.

The key strengths of our study include the diversity and size of the sample, which allowed us to statistically generate a demographically representative sample of the United States. Furthermore, we are able to describe a compliance relation with cases

![Fig. 3. Correlation between intent to comply and future COVID-19 burden (subsequent 30 days): on the national level, the higher the intent to comply, the lower the future number of cases ($P < 0.01$). Models controlled for potential confounders including age, sex, race, ethnicity, region, self-reported social status, health conditions increasing risk from COVID-19, and trust in information sources composite score. The bars indicate the 95% confidence interval around the correlation coefficient. COVID-19, coronavirus 2019.](attachment:correlation_plot.png)
that demonstrates the benefit from even small improvements, and further offers localities the ability to build a compliance program
tailored to the specific needs of their community—for example, communities where social isolation is a near impossibility may
focus on other areas such as handwashing or mask wearing.

Conclusions
The findings of this study indicate that survey responses of population intent to comply with COVID-19 public health recommendations are
associated with subsequent actual COVID-19 infection rates. This approach offers health organizations an inexpensive, scalable method
for predicting outbreaks and targeting populations for education campaigns to encourage compliance, in turn reducing infection.

Acknowledgments
Without the assistance of the following individuals and
groups, the scope and scale of this project would not have been possible. We thank Cletis Earle, Susan Chobanoff, Neal Thomas,
Leslie Parent, Sarah Bronson, Young Ho, Mack Ruffin IV, and the Penn State Qualitative Mixed Methods Core team; Vikram Pendli
and Venkat Thumula at Genzeon; and Russell Branzell, Keith Fraidenburg, Candace Stuart, Amy Proffitt, Michelle Vibber,
and the rest of the extraordinary staff at the College of Information Healthcare Management Executives. The team at Mathematica
was instrumental in the survey analysis, promotion, and design and included support from Alex Bohl and Jill Feldman.

References
1. Ioannidis JPA, Cripps S, Tanner MA. Forecasting for COVID-19 has failed. 
   Int J Forecast 2020; DOI: 10.1016/j.ijforecast.2020.08.004.
2. Clason L. Lack of national COVID testing strategy drives confusion. 
   Washington, DC: Roll call; 2021. Available at: https://www.rollcall.com/2021/04/20/lack-of-national-covid-testing-strategy-drives-confusion/.
3. Daughton CG. Wastewater surveillance for population-wide Covid-19: the present and future. Sci Total Environ 2020;736:139631.
4. Reiner RC, Barber RM, Collins JK, et al. Modeling COVID-19 scenarios for the United States. Nat Med 2020;27:94–105.
5. Frieden TR. Six components necessary for effective public health program implementation. Am J Public Health 2014;104:17–22.
6. Van Scoy LJ, Miller EL, Snyder B, et al. Knowledge, perceptions, and preferred information sources related to COVID-19 among central Pennsylvania adults early in the pandemic: results of a mixed methods cross sectional survey. Ann Fam Med 2021;19:293–301.
7. Lennon RP, Sakya SM, Miller EL, et al. Public Intent to Comply with COVID-19 Public Health Recommendations. Health Lit Res Pract 2020;4:e161–e165.
8. Adler NE, Epel ES, Castellazzo G, et al. Relationship of subjective and objective social status with psychological and physiological functioning: preliminary data in healthy white women. Health Psychol 2000;19:586–592.
9. Deming WE, Stephan FF. On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. Ann Math Statist 1940;11:427–444.
10. Deville J-C, Särndal C-E. Calibration estimators in survey sampling. J Am Stat Assoc 1992;87:376–382.
11. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing.
12. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. Lancet 2007;370:1453–1457.
13. Lennon RP, Sakya SM, Miller EL, et al. Public intent to comply with COVID-19 public health recommendations. Health Lit Res Pract 2020;4:e161–e165.
14. Czeisler MÉ, Tynan MA, Howard ME, et al. Public attitudes, behaviors, and beliefs related to COVID-19, stay-at-home orders, nonessential business closures, and public health guidance—United States, New York City, and Los Angeles, May 5–12, 2020. MMWR Morb Mort Rep 2020;69:751–758.
15. Thompson RN. Epidemiological models are important tools for guiding COVID-19 interventions. BMC Med 2020;18:152.
16. Wynnats L, Van Calster B, Bonten MMJ, et al. Prediction models for diagnosis and prognosis of covid-19 infection: systematic review and critical appraisal. BMJ 2020;369:m3328.