Journal of Public Health International

Open Access - Peer-Reviewed - ISSN 2641-4538

Portico Archived | Crossref DOI | CC BY 4.0 Licensed

Methodology Article

Controlling the Covid-19 Pandemic without Killing the Economy: About Data Driven Decision Making with a Data Model Assessing Local Transmission Risk

Manfred Stapff[1*]

[1] LG Chem Life Sciences Innovation Center, Cambridge, MA, USA

*** Correspondence:** Manfred Stapff, LG Chem Life Sciences Innovation Center, 139 Main St, Ste 200, Cambridge, MA 02142

Published citation: Manfred Stapff (2020) **Controlling the Covid-19 Pandemic without Killing the Economy: About Data Driven Decision Making with a Data Model Assessing Local Transmission Risk**. *Journal of Public Health International* - 3(2):12-19. https://doi.org/10.14302/issn.2641-4538.jphi-20-3621

Article DOI: 10.14302/issn.2641-4538.jphi-20-3621

Abstract

Context

In the face of further waves of the COVID-19 pandemic it becomes essential to find a balance between protective actions to guard public health and restrictive measures which can collapse our economy.

Background

As a basis for public health decisions, officials still rely on metrics which were helpful in the beginning of the pandemic but are now not precise enough for a focused and targeted approach to keep the spread of the infection under control. This can lead to public mistrust, “pandemic tiredness”, and can cause unnecessary damage to the economy without having the desired protective effect on public health.

Methods

This article discusses various metrics, their advantages and caveats, and it provides suggestions for use in a more targeted and risk-based approach, as an alternative to the current “general lock-down” practice. It suggests the notion of including a concept of “risk contacts per area” to better describe the possibility of virus transmission than currently published metrics do. The article also suggests specific analyses of real-world data for identification of populations at risk for severe courses of COVID-19 to allow more targeted protective actions.

Discussion

Data currently used to describe the COVID-19 pandemic lack important parameters like population density and local likelihood of potentially infectious contacts. The currently often used “all or nothing”

approach of shut-down orders needs to be replaced by more sophisticated tactics considering individual local exposure risks and need to be balanced towards metrics on economic short term and long-term impact. In addition, smart analyses of real-world data may contribute to effective protection of individuals at risk.

Keywords:
COVID-19; epidemiology; infection rate

Article Information	
Academic Editor	Aroma Oberoi, Professor & Head, Department of Microbiology, Christian Medical College & Hospital, India.
Received Date	12 November 2020
Accepted Date	14 November 2020
Published Date	20 November 2020
Online	20 November 2020
Publication	Volume 3; Issue 2; Pages 12–19
Copyright	© 2020 Manfred Stapff
License	This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. https://creativecommons.org/licenses/by/4.0/

Context

More than 8 Million people in the US have slipped into poverty due to COVID-19 related lockdown measures[1].

As we now approach new waves of the pandemic, it becomes more essential than ever to balance protective actions to guard the health of communities versus restrictive measures which not only can ruin our economy but also may turn into a gamble with people's trust and compliance. Politicians and other players in the public health sector repeatedly state that the way the pandemic is dealt with should not be politicized but should be based on data and science. There is hope that the public can be convinced by facts rather than by beliefs. This approach sounds reasonable, but officials need to follow this approach themselves, have to use adequate data and must interpret their own metrics correctly.

Especially as more politicians test positive themselves (e.g. D. Trump, US President and several White House staff members; J. Spahn, German Ministry of Health, B. Johnson, Prime Minister of UK, V. Zelensky, President Ukraine) or need to be quarantined (T. Adhanom Ghebreyesus, WHO Director), general trust in their ability of decision making may become undermined.

Since the very beginning of the pandemic, popular publicly available trackers[2],[3] provide the number of cases and deaths with the possibility to drill them down by geography or by time. However, what was suitable in the beginning, to watch the increasing spread of the disease, may not be adequate anymore on a long-term basis to keep it under control and eventually eliminate it.

Limitation of Currently Used Metrics

As dramatic and saddening as the total number of cases [3] (~51 million globally) and the cumulative count of COVID-19 related deaths (~1.3 million globally) may be, they don't teach us anything more than the overall burden of the pandemic. Similarly, on the regional level of a state or a county, the total numbers increase with the time passed since the beginning of the pandemic and do not tell anything about a current infection risk. Furthermore, it is important to acknowledge that "cases" usually refer

to virus tests (antigen or molecular) which provide only a short look into a narrow time window of a few days during the course of an infection, and that the cumulative number correlates with the intensity with which testing is done in a region.

It has been estimated that perhaps only about 20% of people who were infected with SARS-CoV-2 develop symptoms or the full picture of COVID-19. This means that about 80% of infected people may be unknown (if not tested just out of curiosity) and may further spread the virus. Other sources assume that between 20% and 40% of infected persons stay asymptomatic.[4] Thus, 60 to 80% are not identified by virus test results if these tests were only accessible to symptomatic patients or to persons who were in contact with a COVID-19 patient (most testing sites apply these restrictions). In addition, infectiousness before symptom onset and with only minimal contact has been reported.[5] Therefore, virus tests and tracing activities which are just built on positive tests show only the "tip of the iceberg". They don't see the asymptomatic spreaders and are not representative of the general population as they are based on a preselected group.

Overall, the totals (case load) provide a good insight into how much a region is, and has been, generally impacted by SARS-CoV-2. However, in order to understand the expansion of the disease and the (protective) effect of interventions the totals need a correct denominator.

Need for a Correct Denominator

Population

It is obvious that highly populated geographies, countries, or states can have more COVID cases than small ones. Therefore, the most important denominator to be used in COVID-19 statistics is an indicator of the population, usually given as "per 100 000". As an example, the US with currently 10.1 million cases[6] in 328 million people has endured a much higher cumulative case load over time than Italy with 960 000 cases in a population of 60 million: 3079 per 100k in the US versus 1600 per 100k in Italy. The epidemic started in the US approximately one month before Italy. Since the cumulative numbers grow continuously, one could argue that the total is expected to be higher in regions that had to deal with the pandemic for a longer time, but such differences should fade out over time.

Time

The incubation period for COVID-19 is thought to extend to 14 days, with a median time of 4-5 days from exposure to symptom onset[7]. Therefore, the current disease activity and infection rate can be best estimated by looking at a time period of the past 1 to 2 weeks. For example, New York State had seen 11 new cases per 100 000[8] in the first week of November 2020. In Minnesota the disease was more than triple as active, with 46 new cases per 100 000 during the same period of 7 days.[9]

Many European countries consider numbers higher than 50 new cases in seven days per 100k as an indicator of a "risk country" which triggers travel restrictions, however without a convincing rationale why this specific threshold has been chosen.[10]

Area and Population Density

The definitions of "risk country" or "risk state" which can generate significant travel restrictions and quarantine requirements vary from country to country, in the US even from state to state. Some guidelines use the number of cases per week per 100k population[10], others per day[11], others the number of cases per day per million. None of these definitions takes population density into consideration although density is a critical parameter in epidemiology, especially when it comes to spreading a disease of infectious nature. As it is obvious and addressed by the requirement of "social distancing", infectious diseases spread easier if people are closely together. Consequently, the exposure risk can be expected to be higher in densely populated areas. Density and congregation of people in streets, restaurants, beaches or parks are highly variable and can never be fully taken into consideration by a calculation. However, the area of a region can be used as denominator for an overall indicator of concentration of infected persons. It provides an impression about the density of potentially infectious people and the likelihood to encounter them in that area. For example, Cambridge, MA has an area of 7.1 square miles (sqmi). In the last week of October there were a total of 51 new cases[12] over a 7-day period, i.e. 7 potentially

infectious people per sqmi. In New York City 6650 new cases[13] in a week within an area of 303 sqmi represent a much higher density of 22 newly infected people per sqmi. Therefore, one could expect a generally higher infection risk in New York City. As the actual risk depends a whole heap on individual behavior, e.g. in parks, restaurants, bars, churches and other places of congregation, it may be more important to regulate such behavior in locations with a higher general infection density.

(Note: The 7 day period as a basis to calculate the currently infectious population may be too short, and 14 days may be a better metric as studies have shown that the virus can be shed much longer.[14])

Testing Intensity

Testing frequency has been commonly used as an excuse to explain high numbers of positive cases.[15] While it is a general truth that more search generates more findings, it is only the rate of positive tests (positive SARS-CoV-2 PCR) per number of tests conducted in the past 1 to 2 weeks which allows an objective interpretation. The ratio of positives to total tests is supposed to be stable against changes in the total number of tests. However, repeated tests on the same person (e.g. routine tests for health care workers) numerically reduce the percentage of positive tests since the denominator increases[16]. On the other hand, the narrower the indication for the test is set (e.g. tests are restricted to first responders or to patients with symptoms) the higher the positive rate will be, due to a preselection of persons with a higher likelihood to test positive. On the flip side, if tests are widely available, many people will get tested without a specific reason, just for "peace of mind", which reduces pretest probability and decreases the likelihood of a positive test.

Test Specificity

The currently authorized SARS-CoV-2 virus tests are optimized for their ability to correctly detect positive cases and to avoid false negative results. High sensitivity is often inversely correlated to the ability to determine a negative case (specificity). With high sensitivity the likelihood of false positive results can increase, especially in diseases with low prevalence. For mass-test interpretation, the positive predictive value (PPV, percent of positive test results that are true positives) varies with disease prevalence. As disease prevalence decreases, the percent of test results that are false positives increases. The following hypothetical example is based on an average test sensitivity of 96% and a specificity of 95% among authorized tests and an assumed prevalence of 7% in the US population. Table 1.

The consequence of these (realistically assumed) numbers for test sensitivity, specificity, and disease prevalence is that only approximately 60% of people with a positive test may actually have a SARS-CoV-2 infection. This over-estimation, together with other uncertainties (e.g. selection bias of test population by test indication) needs to be considered when interpreting infection rates as a basis for decision making. Low specificity together with low prevalence can lead to overestimating the COVID-19 incidence and the extent of asymptomatic infection, in the worst case with the consequence of a misdirection of policies regarding lockdowns and school closures.[17]

Positive Test as Surrogate Parameter: What is the Clinical Relevance?

The estimated proportion of asymptomatic SARS-CoV-2 infections ranges from 18% to 81%.[18] One may argue that asymptomatic infections have no clinical relevance, and a positive test alone is no more

Table 1. Calculation of a positive predictive value based on test sensitivity (96%), specificity (95%), and disease prevalence (7%)

	SARS-CoV-2 infected	Not infected	total
Positive PCR	0.96*7=6.72	93-88.35=4.65	6.72+4.65=11.37
Negative PCR	7-6.72=0.28	0.95*93=88.35	0.28+88.35=88.63
total	7	93	100

PPV= 6.72 / (6.72 + 4.65) = 6.72 / 11.37 = 0.591

than a surrogate parameter. However, asymptomatic infections are a key contributor in the spread of COVID-19. Therefore, asymptomatic cases should be reported in COVID-19 statistics.

Better than just positive tests, the number of hospitalizations or the frequency of COVID-19 related deaths proves the clinical relevance of an outbreak. Mortality is obviously an indicator which is heavily impacted by comorbidities, demography, and by the quality of medical care. Furthermore, as a metric for public health measures the respective time lag needs to be considered. The number of deaths is a parameter for activities four to eight weeks earlier, i.e. the time it takes from infection to symptoms to hospitalization to intensive care to death.

In extreme cases hospitalizations may be falsely low if the health care system is already over-burdened (hospitals at limit) and cannot take new patients anymore. Depending on the applied definition of "COVID related", the number of deaths may be falsely high and may not be easily comparable across legislations.

Conclusions for Public Health Decisions

Decisions about closing or opening public life should be based on a balance between protecting public health and minimalizing economic damage (well acknowledging that these two principles are not completely separated but can have a long-term effect on each other).

The Public Health side Should be Based on Three Principles

Reduce exposure to potentially infectious persons

Protect vulnerable population

Count on reliable data

Lock down orders intend to reduce the number of potentially infectious contacts. The denser a vulnerable population is, and the more potentially infectious persons can be found in an area, the higher is the risk, and the more important it is to reduce potential contacts. A two-step approach could use the density of newly infected persons in a geographic area as an indicator of the importance of any regulating actions. Since the likelihood of an infection depends mainly on individual behavior and distancing in small areas, in a second step individual businesses should be regulated purely based on the individual risk and business or their ability to reduce it. A data driven approach would be a calculation of the likelihood that a potentially infectious person meets a vulnerable person within a radius of 6ft (as this is the currently used definition for "close contact"[19]), which is an area of approximately 100 square feet (sqft).

Examples

A recent study[20] found that 20% of workers in a grocery store had positive viral assays at the same time (by the way, 76% of them asymptomatic). With an average area of a mid-size grocery store of 20000 sqft and 10 employees working there at the same time, this would mathematically result in ((0.2*10)/20000)*100 = 0.01 infectious workers within a 6 feet radius (100 sqft). If there are 20 customers in the store, = i.e. 0.1 per 100 sqft, the likelihood of a potentially infectious contact happening within a 6 ft radius (if social distancing is not followed) is 0.01 * 0.1 = 0.1%, which seems very low. However, using the same precautionary definitions as currently recommended for "risk threshold" in COVID-19 alert apps (3 minutes within 6ft radius in contact with an infectious person), the risk increases to 2% per hour. Such calculations can support rules on the maximum allowed number of customers in a store.

As far as we know, President Trump's Rose Garden event on September 26th to introduce the Supreme Court Candidate created approximately 15 new SARS-CoV-2 infections (not counting further secondary spread). Assuming there was initially only one infectious person among the 200 guests in an area of 7500 sqft[21], this would account for 0.013 infectious and 2.66 vulnerable persons per 100 sqft (radius of 6ft) allowing 0.35 risk contacts at any given time, or 40 potentially infectious contacts of 3 minutes duration during a two hour event, where no social distancing rules were obeyed and almost no face coverings were worn.

The numbers in these two examples may be hypothetical (despite being based on reasonable assumptions) but the mathematical approach shows that calculating the densities of infectious persons and vulnerable persons per area and assessing the likelihood of contacts within a 6ft radius may be a way to assess the infection risk in small locations or at events where people congregate.

Regulatory actions on the level of large geographic or legislative areas (counties, cities) are less successful in fighting the epidemic than actions which are focused and targeted on smaller locations. In a study in 37 OECD countries, travel restrictions and public transport restrictions had no effect, but school and workplace closures, the introduction of a mask requirement and the restriction of events led to a statistically significant decrease in the number of infections.[22]

Balance Health and Economy

It must not become a binary question between health and economy. Both – if neglected – can have disastrous effects on humanity. A prolonged lock-down will have (has already) detrimental effects on businesses, jobs, minorities, private and public budgets and is not sustainable much longer without significant damaging effect on food supply, public safety, and eventually the health care system itself. A prolonged lock-down or stay-at-home policy itself has negative impact on physical and mental health of people due to neglected chronic diseases, reduction of mobility, increased risk of cardiovascular or thrombotic events.[23]

The Need to Protect Risk Populations Identified by Real-World Data

As many health care systems in the world are fully digitalized, all the data needed for natural history studies, observational trials or to learn about risk profiles are available in electronic format.

Who is at risk for a more severe or even lethal outcome? A risk score based on characteristics of COVID-19 patients at the time of admission to the hospital does already exist.[24]

A similar score needs to be developed, based on real world data (RWD), for profiling of patients who may be more likely to survive an infection with no or low symptoms versus those who may become more severely sick.

An Italian study[25] found correlations of susceptibility and severity of COVID-19 with HLA haplotypes but could only do this by geographically overlapping two different data sources. This tells us the huge opportunity in analyzing electronic health records (EHR) – as far as they are complete or linked on an individual patient basis.

It has already been shown that artificial neural network (ANN) analyses of socio-medical data from insurances can be used to search for predictors of undiagnosed HCV infections.[26] Why not apply these techniques to COVID-19?

As the initial outbreak was in the Wuhan province in China, many of the studies about clinical outcomes and phenotype risk factors are based on Chinese patients. Now the primary impacted country is the US with broad diversity of races and ethnicities. Therefore, it is imperative to use data sources which include information on race, ethnicity and, ideally, genetic information.

Currently it is unclear which role previous vaccinations[27], medical history, genetic factors, or constellation of laboratory parameters play. All these records are part of a well-documented EHR system and can be analyzed by artificial intelligence (AI) to find patterns which no human brain has ever been thinking about.

Beyond the well-known comorbidities like obesity, diabetes, hypertension and asthma, other patterns of factors have turned out to pose an increased risk to develop COVID-19 symptoms or a more severe or lethal course. South-East Asian descent seems to increase, and blood type 0 to decrease[28] the risk for COVID-19. Down Syndrome is associated with immune dysfunction, congenital heart disease, pulmonary pathology and may be a relevant albeit unconfirmed risk factor for severe COVID-19.[29] RWD will help identify clinical and phenotype patterns which pose an increased risk to develop clinical symptoms and a more serious course of the infection.

As clichéd as it may sound, decision makers in public health need more “thinking out of the box”. The following should make us think:

Meteorology has better prediction models than were used during the COVID-19 pandemic

Smartphone apps with Bluetooth and geo-data work faster and more efficiently than an army of human tracers

Real World Data from electronic medical records in combination with artificial intelligence may provide more promising hypotheses about clinical risk than individual experts

Deep analyses of real world data, preferably using artificial intelligence rather than pre-specified (and potentially biased) hypotheses, should be used to identify previously unknown risk patterns which then can be translated into a COVID-19 risk score and eventually be used to protect vulnerable individuals.

This would allow a much more targeted approach in the fight against the pandemic than an "all-or-nothing", "lock-down-or-open-all" method which is still applied in most countries.

Implications for Policy & Practice

Interpretations of epidemiologic data about SARS—CoV-2 infections, e.g. to designate an area as a "risk area", vary by state and country, but should be harmonized.

Currently reported data used for regulatory decisions should include important parameters like population density and local likelihood of potentially infectious contacts.

The currently often used "all or nothing" approach of shut-down orders needs to be replaced, or at least complemented, by more sophisticated tactics quantifying individual local exposure risks and need to be balanced towards metrics on economic short term and long-term impact.

"Big data", by analyzing electronic health records and wearables, coupled with artificial intelligence, can help identify risk areas and better protect vulnerable populations.

Such targeted approaches may have less economic impact than current "lock-down-or-open-all" methods.

Competing Interests

The authors have declared that no competing interests exist.

References

1. The New York Times. 8 Million Have Slipped Into Poverty Since May as Federal Aid Has Dried Up. https://www.nytimes.com/2020/10/15/us/politics/federal-aid-poverty-levels.html?campaign_id=2&emc=edit_th_20201015&instance_id=23180&nl=todaysheadlines&regi_id=43808518&segment_id=41232&user_id=dab810be00c61b173be76f91142ba1ca
2. *CDC COVID Data Tracker. https://covid.cdc.gov/covid-data-tracker/#cases_casesinlast7days*
3. Johns Hopkins Coronavirus Resource Center. https://coronavirus.jhu.edu/map.html. Accessed 11/09/ 2020
4. Buitrago-Garcia D. Occurrence and transmission potential of asymptomatic and presymptomatic SARSCoV-2 infections: A living systematic review and meta-analysis. *PLoS Med*, 17(9):1-25. https://doi.org/10.1371/journal.pmed.1003346
5. Böhmer M, Buchholz U, Corman V M, Hoch M, Katz K, Marosevic D V. Investigation of a COVID-19 outbreak in Germany resulting from a single travel-associated primary case: a case series. Lancet online. 2020:920-928. https://doi.org/10.1016/S1473-3099(20)30314-5
6. 2020. *As*, of:11-09
7. National Center for Immunization and Respiratory Diseases (NCIRD), Division of Viral Diseases. Interim Clinical Guidance for Management of Patients with Confirmed Coronavirus Disease (COVID-19) https://www.cdc.gov/coronavirus/2019-ncov/hcp/clinical-guidance-management-patients.html
8. The New York Times. New York Covid Map and Case Count. https://www.nytimes.com/interactive/2020/us/new-york-coronavirus-cases.html, accessed 11/03/2020
9. The New York Times. Minnesota Covid Map and Case Count. https://www.nytimes.com/interactive/2020/us/minnesota-coronavirus-cases.html, accessed 11/03/2020
10. Koch-Institut Robert. Bundesinstitut im Geschäftsbereich des Bundesministeriums für Gesundheit https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Risikogebiete_neu.html

11. Commonwealth of Massachusetts COVID-19 Travel Order. https://www.mass.gov/info-details/covid-19-travel-order#lower-risk-states-
12. Cambridge Public Health Department. 2020. COVID-19 Data Center. https://cityofcambridge.shinyapps.io/COVID19/?tab=new
13. The New York Times. New York City Covid Map and Case Count. https://www.nytimes.com/interactive/2020/nyregion/new-york-city-coronavirus-cases.html#cases, accessed 11/03/2020
14. Long Q X, Tang X J, Shi Q L. Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. 2020. *Nat Med*, 26 8 https://doi.org/10.1038/s41591-020-0965-6
15. Begley S. 2020. Trump said more Covid-19 testing 'creates more cases.' STAT News. 07/20/2020 https://www.statnews.com/2020/07/20/trump-said-more-covid19-testing-creates-more-cases-we-did-the-math/
16. U S Food, Administration Drug. Potential for False Positive Results with Antigen Tests. https://www.fda.gov/medical-devices/letters-health-care-providers/potential-false-positive-results-antigen-tests-rapid-detection-sars-cov-2-letter-clinical-laboratory?utm_medium=email&utm_source=govdelivery
17. www.thelancet.com/respiratory; Published online September 29, 2020 https://doi.org/10.1016/S2213-2600(20)30453-7
18. Nikolai L A, Meyer C G, Kremsner P G, Velavan T P (2020).Asymptomatic SARS Coronavirus 2 infection: Invisible yet invincible. International Journal of Infectious Diseases, 100 112-116. https://doi.org/10.1016/j.ijid.2020.08.076
19. 23 (2020).https://www.cdc.gov/coronavirus/2019-ncov/global-covid-19/operational-considerations-contact-tracing.html Updated
20. Lan F-Y, Suharlim C, Kales S N. Occup Environ Med Epub ahead of print 1-7. https://doi.org/10.1136/oemed-2020-106774
21. https://en.wikipedia.org/wiki/White_House_Rose_Garden
22. https://www.rki.de/DE/Content/InfAZ/N/Neuartiges_Coronavirus/Projekte_RKI/impact-control-measures-oecd-summary.pdf. Submitted to Lancet Public Health
23. Gäbel G, Kröger K. Risiken der „Stay at home"-Politik im Rahmen der COVID-19- Pandemie Gefässchirurgie https://doi.org/10.1007/s00772-020-00697-2
24. Liang W, Liang H, Ou L. Development and Validation of a Clinical Risk Score to Predict the Occurrence of Critical Illness. 2020. in Hospitalized Patients With COVID-19. JAMA Intern Med. Published online 80(8):1081-1089. https://doi.org/10.1001/jamainternmed.2020.2033
25. Al. Correlation of the two most frequent HLA haplotypes in the Italian population to the differential regional incidence of Covid-19. 2020. *J Transl Med*, 18(352):1-16. https://doi.org/10.1186/s12967-020-02515-5
26. Reiser M, Wiebner B, Hirsch J. Neural-network analysis of socio-medical data to identify predictors of undiagnosed hepatitis C virus infections in Germany (DETECT) 2019. *J Transl Med*, 17(94):1-7. https://doi.org/10.1186/s12967-019-1832-4
27. Hamiel U. SARS-CoV-2 Rates in BCG-Vaccinated and Unvaccinated Young Adults. JAMA online. 2020. https://doi.org/10.1001/jama.2020.8189
28. Hoiland R L, Fergusson N A, Mitra A R. The association of ABO blood group with indices of disease severity and multiorgan dysfunction in COVID-19. *Blood Advances*. 2020; 4: 4981-9. https://doi.org/10.1182/bloodadvances.2020002623
29. Ashley Kieran Clift. COVID-19 Mortality Risk in Down Syndrome: Results From a Cohort Study Of 8 Million Adults. https://doi.org/10.7326/M20-4986

Publisher's Note

All claims, opinions, data, and conclusions expressed in this article are solely those of the author(s) and contributor(s), and do not necessarily represent those of Open Access Pub, the journal, its editors, or reviewers. Open Access Pub does not guarantee or endorse any product, method, instruction, or claim discussed in the article, and disclaims responsibility for any injury to people or property resulting from ideas, methods, instructions, or products referred to in the content.