Drug repurposing for COVID-19: the problem of excessive hypothesis testing

The problem of excessive hypothesis testing

Mariusz Maziarz (corresponding author)
Interdisciplinary Centre for Ethics, Jagiellonian University, Kraków (Poland)
Institute of Philosophy, Jagiellonian University, Kraków (Poland)
mariusz.maziarz@uj.edu.pl
ORCiD: 0000-0003-1979-0746
Grodzka 52, Kraków, Poland

Adrian Stencel
Institute of Philosophy, Jagiellonian University, Kraków
adrian.stencel@uj.edu.pl
ORCiD: 0000-0002-3933-2059

Abstract

Rationale, aims, and objectives

The current strategy of searching for an effective drug to treat COVID-19 relies mainly on repurposing existing therapies developed to target other diseases. There are currently more than four thousand active studies assessing the efficacy of existing drugs as therapies for COVID-19. The number of ongoing trials and the urgent need for a treatment poses the risk that false-positive results will be incorrectly interpreted as evidence for treatments' efficacy and a ground for drug approval. Our purpose is to assess the risk of false-positive outcomes by analyzing the mechanistic evidence for the efficacy of exemplary candidates for repurposing, estimate false discovery rate, and discuss solutions to the problem of excessive hypothesis testing.

Methods

We estimate the expected number of false-positive results and probability of at least one false-positive result under the assumption that all tested compounds have no effect on the course of the disease. Later, we relax this assumption and analyze the sensitivity of the expected number of true-positive results to changes in the prior probability (π ¿ that tested compounds are effective. Finally, we calculate False Positive Report Probability and expected numbers of false-positive and true-positive results for different thresholds of statistical significance, power of studies, and ratios of effective to non-effective compounds. We also review mechanistic evidence for the efficacy of two

exemplary repurposing candidates (hydroxychloroquine and ACE2 inhibitors) and assess its quality to choose the plausible values of the prior probability (π) that tested compounds are effective against COVID-19.

Results

Our analysis shows that, due to the excessive number of statistical tests in the field of drug repurposing for COVID-19 and low prior probability (π ¿ of the efficacy of tested compounds, positive results are far more likely to result from type-I error than reflect the effects of pharmaceutical interventions.

Key-words: prior probability, mechanistic evidence, EBM+, excessive hypothesis testing, false positives, covid-19

Conflict of Interest

No conflict of interest is reported by the authors.

Acknowledgement

The authors thank for insightful comments received during the Interdisciplinary Centre for Ethics research seminar.

Funding

The work of Mariusz Maziarz has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 805498). Mariusz Maziarz has received scholarship from the Foundation for Polish Science (FNP). Adrian Stencel has received scholarship from the Foundation for Polish Science (FNP).

Data Availability Statement

Data available in article supplementary material.

Ethical Approval Statement

Ethical approval is not required for the study.

Drug repurposing for COVID-19: the problem of excessive hypothesis testing

1. Introduction

SARS-CoV-2 belongs to a family of human coronaviruses that cause common cold and more severe conditions such as breathing difficulty and acute respiratory distress syndrome (ARDS)[1]. Direct mortality of COVID-19, the disease caused by SARS-CoV-2, is not the only reason for why the pandemic creates an unprecedented threat to the healthcare systems. Some patients require hospitalization (oxygen therapy and Intensive care, in particular) [2] and the patients who recovered from COVID-19 might experience long-term cardiovascular and neurological consequences[3]. As SARS-CoV-2 pandemic is a global threat that affects millions of people, medical researchers are trying to rapidly develop a therapy for COVID-19. Their efforts rely primarily on drug repurposing, i.e., identifying those existing or investigational drugs that are effective for COVID-19[4,5]. The primary advantages of drug repurposing (as opposed to developing new compounds) are shorter time of the process and known risk profile of the existing drugs[6].

The hope of this strategy is to find a therapy lowering mortality, alleviating symptoms, and/or shortening the course of the disease. Within the first hundred days of the pandemic, 689 randomized controlled trials (RCTs) have been conducted, ongoing, or in preparation (of which more than a hundred tested the efficacy of antimalarial drugs)[7]. The number of the trials is steadily growing. At the time when the manuscript is being prepared (December 23rd, 2020), Clinicaltrials.gov lists as much as 4'118 active studies targeting COVID-19, including 2'302 randomized controlled trials (RCTs). Among over two thousand RCTs, 168 active or planned trials tests the efficacy of hydroxychloroquine (HXQ) and 55 – remdesivir. We excluded the studies that are terminated, suspended or withdrawn.

The current strategy of drug repurposing is clearly very tempting as it might accelerate the process of getting the cure. However, we are concerned that, as we delineate below, it is likely to result in many false-positive results (i.e., reports showing statistically significant difference between the treatment and control arms due to chance alone). In effect, several small studies of the same compound are likely to deliver inconsistent results. Conflicting evidence has already emerged in regard to the efficacy of remdesivir[8], hydroxychloroquine[9], and convalescent plasma[10]. The chronological order of results, where (false) positive results were followed by negative outcomes (e.g., WHO Solidarity Trial Consortium 2020), has driven changes in treatment recommendations [11,12].

We argue that the false-positive results emerge due to the problem of multiple comparisons. It denotes a situation when one conducts more than one statistical inference from the same or dependent datasets. It is well known that, in such situations, p-values are underestimated and type-I errors (rejecting the null hypothesis of no effect when it is true) happen more often than it could be expected[13,14]. Still, it has only recently been recognized that false positive results can emerge in cases when several research teams address the same research question[15,16] . In unfavorable circumstances, multiple comparisons can drastically impede true inferences. For example, Colhoun et al.[17] estimate that as much as 95% of genome-wide research report false-positive results. False-positive results are also more frequent than expected in clinical trials[18] what impedes the process of drug approval[19]. Ioannidis[20] shows that false discovery rate (the ratio of false-positive to all positive results) depends

positively on the number of tested relationships and negatively on the prior probability of tested hypotheses and the number of effective therapies. Unfortunately, the field of drug repurposing for COVID-19 is populated by numerous trials (¼ include less than 1'000 participants) that are too small to report significant differences in mortality[21]. Underpowered RCTs are more likely to report false positive findings[22].

Our purpose is to show that false positive results are very likely to be obtained in the field of drug repurposing for COVID-19 and discuss some approaches regarding how to evaluate the results. First, we show that conducting that many clinical trials negatively influences false discovery rate and impedes making reliable inferences (section 2). Second, we argue, using hydroxychloroquine and ACE2 inhibitors as examples, that the mechanistic evidence for the efficacy of repurposing candidates is weak what justifies the assumptions endorsed in the analysis (section 3). We end up with discussing the implications of high false discovery rate for evaluating the results of drug-repurposing studies (section 4).

2. The problem of excessive hypothesis testing

False positive findings can emerge due to random allocation of patients to treatment and control arms and individual differences in disease progression. Unwarranted claims regarding treatment efficacy are most likely to emerge in cases when the natural course of disease is such that the health of a majority of patients improves over time[23]. In that cases, observing the improvements of patients in the treatment group can erroneously be ascribed to treatment if the control group accidently includes a larger proportion of cases that deteriorate. A majority of COVID-19 cases are moderate. About 10-20% develop into a severe disease requiring hospitalization[2]. Therefore, researchers conducting RCTs can ascribe random differences in outcomes between treatment and control arms arising from sampling to the intervention despite the treatment has no effect on the course of the disease.

To account for the random changes in outcomes, only statistically significant results (i.e., unlikely to emerge by chance alone) are taken as evidence for treatment's efficacy. The threshold of statistical significance α is usually set at the level of 0.05 (more conservative thresholds of 0.01 or 0.001 are sometimes employed). Under normal circumstances, not only statistical significance but also clinical relevance of the size of treatment effect is taken into account[24]. However, the case of COVID-19 is exceptional in the lack of any treatments with the exception for drugs that are known to alleviate certain symptoms (e.g., fever or inflammation). Therefore, any reduction in mortality or disease progression would be considered as a clinically useful effect and hence we can focus exclusively on the problem of statistical hypothesis testing.

The threshold of statistical significance α denotes also the probability of type-I error (false-positive result). In case $\alpha = 0.05$, (on average) one in twenty RCTs can be expected to report a false-positive result. This allows for estimating the mathematical expectancy of the number of false-positive results for a group of studies. Assuming that the drug repurposing for COVID-19 is the null field (i.e., no tested compounds make any difference in comparison to their controls), the number of false positives (FP) is given by[16]:

$$FP = n * \alpha$$

Where:

n - number of null studies;

α – the threshold of statistical significance/type-I error probability $\alpha = Pr\left(rejecting\ H_0 \middle| H_0\ is\ true\right)$.

If this pessimistic scenario is true and none of the candidates for repurposing tested in more than two thousand active RCTs is effective, then the number of false positives can be expected to be as high as 100 (for $\alpha = 0.05$). In a similar vein, one can estimate the expected number of false-positive results for the studies of individual drugs and for different thresholds α, see Fig. 1.

[Figure 1 should be put somewhere here]

In addition, one can calculate the probability of obtaining at least one false-positive result[14]:

$$P(FP \geq 1) = 1 - (1 - \alpha)^n$$

The formula shows that for little as 283 trials (in comparison to the number of studies in the field of drug repurposing for COVID-19) the chances of *not* obtaining at least one false positive result are one to billion (1:1'000'000) when $\alpha = 0.05$, see Fig. 2.

[Figure 2 should be put somewhere here]

Still, it is possible that some treatments will turn out to be genuinely effective and some RCTs will report true positive (TP) results. In that case, to calculate the False Positive Probability Report ($FPRP$), i.e., the probability that a positive result is a false positive, the number of true positive results needs to be extracted from the number of all studies:

$$n = N - N * \pi$$

Where:

N – number of all studies;

π – the ratio of genuinely effective therapies to all tested therapies.

Assuming that the power of each study equals 1, i.e., the probability of accepting the null hypothesis when it is in fact true (type-II error) equals zero (β=0), one can calculate $FPRP$ (this idealizing assumption will be lifted later):

$$FPRP = n * \alpha / N * \pi$$

Assessing the number of true positive results proves difficult, but a range of plausible values for π can be indicated. As we argue below, it is rational to expect that the number of existing drugs effective for COVID-19 is low. We estimate the expected number of true-positive results for the several plausible values of π: 0.00025; 0.001; 0.005; 0.03, see Fig. 3. Considering the overall number of clinical studies N, these values can be interpreted as assumptions that there are, respectively, 0-1; 2-4; 10-20; 60-120 effective treatments under investigation. The lower bounds of the intervals are calculated by multiplying the number of RCTs assessing the efficacy of repurposed drugs against COVID-19 by π and the upper bounds are estimated for all clinical trials, including non-interventional studies.

[Figure 3 should be put somewhere here]

Finally, we one analyze $FPRP$ under the assumption that some studies testing genuinely effective drugs will report (false) negative results. In that case, the probability that a positive result has been obtained despite ineffective treatment is higher and given by the following formula[25]:

$$FPRP = \frac{\alpha(1-\pi)}{\alpha(1-\pi)+\pi(1-\beta)}$$

Where:

π – prior probability that H_1 is true/the ratio of effective to non-effective drugs;

$1-\beta$ – statistical power; $1-\beta = Pr\left(rejecting\, H_0 \middle| H_1\, is\, true\right)$;

$FPRP$ – False Positive Report Probability $Pr\left(H_0\, is\, true \middle| H_0\, was\, rejected\right)$.

The sensitivity analysis (Fig. 4) shows that for the values of π meaningful in the context of drug repurposing for COVID-19, the statistical-significance threshold values standardly used in clinical research, and the range of expected power of studies, false-positive results dominate positive outcomes. In particular, the results show that the studies in the field of drug repurposing for COVID-19 are more likely to report false-positive results than true-positive findings. This conclusion has been obtained even though we exclude from our analysis such factors as poor research design, bias, and questionable research practices that may additionally raise the number of false-positive reports.

[Figure 4 should be put somewhere here]

3.The poor quality of mechanistic evidence for drug repurposing candidates

Above, we have shown that false discovery rate of the field of studies repurposing drugs for COVID-19 is worryingly high due to the excessive number of ongoing trials and the low prior probability of the candidates' efficacy for COVID-19. In this section, we delineate the reasons for why it is rational to expect that there are only few (if any) effective drugs among the candidates for repurposing. In other words, we argue that the prior probability (π ¿ of tested drugs' efficacy is low.

Despite some spectacular examples of successful repurposing attempts such as Viagra[26,27], most compounds target very specific biological processes and are ineffective beyond their domain. This makes the process of drug repurposing marked with failure. For example, amantadine targets only influenzavirus A and is ineffective for influenzavirus B[28] despite a high degree of similarity between the two pathogens. Neuberg et al.[26] analyzed the complete clinical development history of 834 drug candidates that entered clinical trials between 1980 and 2012. They discovered that less than 2% of them were ultimately launched in a therapeutic area other than the one for which they were developed. The success rate is higher for drugs repurposed within the same therapeutic area, e.g., the drugs developed for breast cancer have been successfully repurposed for ovarian cancer. The low success rate of the repurposing studies has been observed under the ordinary circumstances, when the process of selecting candidates for repurposing lasts, on average, about two years[27]. In the case of drug repurposing for COVID-19, this process has been accelerated and many trials had been started before convincing mechanistic evidence for drugs' efficacy was available.

Limiting the efforts to gather extensive evidence from in vitro and animal studies seems justified from the perspective of evidence-based medicine (EBM), the leading approach to assessing the

quality of evidence for clinical decisions[29,30]. This view has led to developing evidence hierarchies (e.g., The Oxford Centre for Evidence Based Medicine[31]) that prioritize RCTs and systematic reviews of RCTs over observational human studies, animal and in vitro research, and theories and expert opinion. Despite mechanistic evidence enters the EBM hierarchy informally, at the stage of developing new drugs and designing clinical trials[32,33], mechanisms are not considered explicitly when efficacy claims are evaluated and if they are, mechanistic evidence is believed to be of lower quality in comparison to associational studies. This view results from the prioritization of those research methods that deliver evidence less susceptible to bias or confounding[34,35]. From that perspective, limiting the efforts to produce high-quality mechanistic evidence are warranted considering the extraordinary context of the pandemic. According to EBM, RCTs and, subsequently, non-interventional studies deliver more trustworthy evidence and mechanistic research may seem to be excessively time consuming in case of a public health emergency such as the pandemic of SARS-CoV-2.

The voices advising expanding the evidentiary base of the EBM movement[36–39] and developing the epistemic theory of causality requiring both difference-making and mechanistic evidence for establishing causal claims[40] have motivated the emergence of the EBM+ movement[41]. According to EBM+, "[a] well-established mechanism of action can support the efficacy claim, while a hypothesized mechanism that has little evidence or contrary evidence (ie, lack of biological plausibility) can undermine the efficacy claim"[42]. The EBM+ position can be fruitfully applied to assessing the plausibility of efficacy claims arising from drug repurposing clinical trials. We analyze the mechanistic evidence for the efficacy of hydroxychloroquine (HXC) and angiotensin-converting enzyme 2 (ACE2) inhibitors, to assess the chances for repurposing candidates to be effective treatments for COVID-19.

We start with HXC, which was suggested as one of the early candidates for treating COVID-19 patients [43,44]. HXC is a widely used and relatively safe anti-malaria drug. Hydroxychloroquine is an analog of chloroquine that is safe and more popular because it is less likely to interact with other drugs. In recent years, Chloroquine and HCX has been shown in vitro to have antiviral, anticancer, and antifungal properties [45,46]. Therefore, it is not surprising that COVID-19 was suggested as another potential target. The suggestion results from a laboratory research whereby HXC and chloroquine was shown to inhibit the ability of SARS-CoV-2 to infect African green monkey kidney Vero cells [43,44]. These results have been used as a reason for starting more than 200 clinical studies, some of which have been prematurely terminated.

Despite some positive outcomes that, in the light of our analysis, can be interpreted as false positives, the larger and more conclusive studies have reported insignificant effect of hydroxychloroquine on the course of COVID-19[47,48]. This might be surprising considering that the mechanistic evidence from the in vitro research supporting the efficacy of HXC is considerably well justified. However, new negative mechanistic evidence has emerged recently. Hoffman et al.[49] discovered the exact mechanism blocking the replication of SARS-CoV-2 in African green monkey kidney Vero cells. This mechanism remained unknown at the time when clinical trials of HXC were started. Sars-CoV-2 can enter cells by two different mechanisms. First, SARS-CoV-2 spike protein attaches to the ACE2 receptor and inserts its genetic material into the cell. Second, the virus is absorbed into endosomes (a part of endocytic membrane transport pathway). Depending on the cell type, the enzymes involved in these mechanisms might be different. Some, like kidney cells, need an

enzyme called cathepsin L for the virus to successfully infect them. Others, like lung cells, need an enzyme called TMPRSS2 (on the cell surface). Cathepsin L requires an acidic environment to function and allow the virus to infect the cell while TMPRSS2 does not. HXC increases the endosomal pH of cells and inhibits viruses that depend on low pH for cell entry [45]. Hoffman et al. [49] showed that in the green monkey kidney cells, HXC decreases the acidity, what disables the cathepsin L enzyme, blocking the virus from infecting the kidney cells. In human lung cells, which have very low levels of cathepsin L enzyme, the virus uses the enzyme TMPRSS2 to infect the cells. Given that the enzyme is not controlled by acidity, HXC is unable to block SARS-CoV-2 from infecting the lung cells or stop the virus from replicating.

This shows that the clinical studies testing the efficacy of HXC for COVID-19 had been started without sufficient evidence for how the compound interferes with SARS-CoV-2 replication process in African green monkey kidney Vero cells. The in vitro research was used as difference-making evidence and was directly extrapolated into humans. HXC is a perfect example that the endeavor of drug repurposing for COVID-19 was not relying on high-quality mechanistic evidence. Clinical trials had been started because some compounds were effective in the tube with disregard for why they were effective. If mechanistic evidence that explains why HXC was that effective in the case of green African green monkey kidney Vero cells were available, then the extraordinary number of clinical trials of HXC would not be started and the emergence of false-positive findings and subsequent hype for HXC as an effective treatment could be prevented. Hoffman et al.[49] delivered negative mechanistic evidence that explains why HXC is unlikely to be effective for COVID-19. As one of the authors, Stefan Pöhlmann, pointed out in the press release discussing their results[50]: “[t]his means that in future tests of potential COVID-19 drugs, care should be taken that relevant cell lines are used for the investigations in order not to waste unnecessary time and resources in our search for effective COVID-19 therapeutics”

Another promising approach is to utilize the knowledge on the cell receptors that SARS-COV-2 uses to enter the host cells. SARS-CoV-2 uses the membrane protein angiotensin two converting enzyme (ACE2) as an entry receptor similarly to SARS-CoV[49,51]. It was quickly realized that a potential way to cure or prevent people from contracting SARS-COV-2 is to focus on the drugs that target ACE2 receptors such as ACE2 inhibitors (a group of antihypertensive drugs)[52,53]. However, these drugs can affect the course of COVID-19 in two opposite ways. On one hand, the increased expression of ACE2 that comes with taking these drugs protects against lung injury by regulating concentrations of angiotensin II, which is vasoconstrictive, pro−inflammatory, and pro−oxidative [52] . It has been suggested that increased expression of ACE2, as a consequence of ACE2 inhibitors, might reduce the intensity of COVID−19[53]. On the other hand, these antihypertensive drugs can possibly worsen the course of COVID−19 by providing a greater opportunity for SARS−CoV−2 to enter host cells by upregulating ACE2 receptors [52,54]. Furthermore, Aronson et al. [42] note these pros and cons of the antihypertensive drugs are not mutually exclusive and should be taken into account. In vivo, the interactions between the two mechanisms might differ from what is observed in separate experiments that model just one mechanism isolated from in vivo interactions. So far, the clinical studies provide conflicting results [55,56].

Additionally, it is not clear whether the protection against lung injury would be as well relevant in the case of SARS−CoV−2 as the mechanism that drives this process is poorly understood. For instance, Monteili et al. [57] show that clinical-grade human recombinant soluble ACE2 (hrsACE2) really

reduces the growth of SARS−CoV−2 in African green monkey kidney Vero cells and as well that it can significantly inhibit the infection of human blood vessel organoids and kidney organoids. At the same time, Monteili et al.[57] highlight that they have not demonstrated that hrsACE2 makes the same effect in the later stages of the disease or that it inhibits the growth of SARS−CoV−2 in lung organoids. Considering that lungs are the major target organ for SARS-CoV-2, a clear mechanistic rationale to extrapolate the results to lung cells is missing. And, as the example of HXC shows, one should be extraordinarily careful in regard to such extrapolations.

The two examples show that the quality of existing mechanistic evidence for the repurposing candidates was, at most, moderate when clinical trials had been started. Even though our discussion focuses on HXC and ACE2 inhibitors, the conclusion applies to other existing drugs that have been and are tested as potential treatments for COVID-19. These drugs have been developed biological processes that differ from those driving SARS-CoV-2 replication[58]. For instance, the attempt at repurposing lopinavir-ritonavir developed for HIV/AIDS target HIV-1 protease, missing from SARS-CoV-2[59]. Some clinical trials have been started without sufficient understanding of exact molecular mechanisms, based on direct extrapolation of in vitro results obtained on cell lines and pharmacokinetics analysis. All in all, mechanistic evidence for the efficacy of candidates for repurposing to treat COVID-19 is weak. This supports what is implicated by our analysis of the field of COVID-19 drug repurposing studies (section 2): considering the number of clinical trials and weak mechanistic evidence in favor of their efficacy, dozens of false-positive findings are likely to emerge. In the light of our analysis, new false-positive results can be expected.

4. Discussion and Recommendations

We have shown that False Positive Report Probability for studies repurposing existing drugs for COVID-19 can be expected to be extraordinarily high due to the number of clinical trials in the field and weak mechanistic evidence for the efficacy of tested compounds. In response to the public health emergency, the pace of research and policy decisions has sped up what resulted in concerns regarding the quality of evidence and policy decisions[60,61]. Drug repurposing is the strategy that, in comparison to developing new compounds, allows for speeding up the process of drug developing[4] and even limited mechanistic evidence may justify starting clinical trials considering the context of public health emergency. However, the number of potential candidates needs to be taken into account when the results are interpreted in order to limit the chances for repurposed drugs to enter clinical practice on the ground of false-positive results.

Unfortunately, drug agencies have endorsed treatments based on positive results of clinical trials that were subsequently contradicted by more decisive RCTs[11,12]. The emergence of conflicting evidence has led to reversals in clinical recommendations what may undermine trust in drug agencies. Food and Drug Administration (FDA) has revoked the emergency use authorization for HXC on June, 15[62]. Remdesivir alone and joined with baricitinib was endorsed by FDA[63] despite clinically-insignificant effect size: the combination therapy reduced the duration of hospitalization or oxygen therapy by just one day. These examples suggest that a change is needed in the way how evidence from the drug repurposing trials is evaluated.

Statisticians have developed many methods of controlling for multiple comparisons and false discovery rate in a field[64,65]. These approaches rely on adjusting the level of p-value threshold (α) to reduce the probability of type-I error. For instance, the simplest approach (Bonferroni correction)

relies on dividing the p-value threshold by the number of statistical tests ($\alpha = p/n$). They should be applied to limit the probability of type-I error and retrieve the interpretability of reported statistical inferences in all cases when multiple statistical tests are conducted on the same population or random samples of the same population. We believe that accounting for false discovery rate is especially needed in cases when compounds are tested without strong mechanistic evidence for their efficacy. Positive results need to be considered in the context of the field where they emerged, i.e., they need to be compared to the number of negative trials testing the same compounds or assessing the efficacy of different drugs for the same disease. Otherwise, obviously ineffective therapies can show statistically significant effects, e.g.[66]. Drug endorsement should not be based on positive results of limited clinical value just because no other therapies are available.

Applying the methods of correcting the level of statistical significance threshold (α) to account for multiple comparisons may be problematic due to the file drawer problem, the risk of p-hacking and other questionable research practices that may additionally influence false discovery rate. Considering vested interests related to finding new therapeutic areas for existing drugs[67], one can expect that the actual number of false-positive results in the field of drug repurposing for COVID-19 may exceed what is implicated by our analysis. Still, even underestimated correction (in the sense that the actual number of conducted statistical tests is higher than expected) improves the quality of conclusions in comparison to the situation when the reported p-values are taken as true likelihoods of type-I error. Alternatively, the field of drug repurposing studies can be considered as exploratory science that aims at generating hypotheses for further research. In that case, positive results should not be taken as evidence for approving COVID-19 treatments. Furthermore, false-positive results can be differentiated from true positive findings by analyzing effect sizes. For example, a reduction of hospitalization duration by just one day can be considered as a clinically insignificant effect that may have arisen by chance alone, especially if no change in mortality is observed. In contrast, if a tested compound would drastically reduce mortality, then it could be considered as genuinely positive result. Extraordinary effect sizes have previously been observed in cases when effective treatments targeting virus' molecular mechanisms were developed. For example, antiretroviral therapy is considered as one of the greatest achievements of modern medicine[68]. The notion of fragility [69] can also be useful to interpret the results of several small trials testing the same compound. According to their view, only results of clinical trials agreeing in effect size and direction should be considered as evidence for clinical decisions, while fragile outcomes that are not replicated by similar RCTs can be expected to emerge by chance.

All in all, the field of drug repurposing for COVID-19 clearly shows that mechanistic and difference-making evidence should go hand in hand particularly in cases when no high-quality RCT results are available. On one side, if false-positive results are expected because of the number of trials, negative mechanistic evidence for drugs' efficacy can impede hopes and influence the inferences drawn from positive results. On the other, strong mechanistic evidence for drug efficacy can motivate starting clinical trials or therapeutic decisions and drug approval in cases when difference-making evidence is of low quality because of multiple comparisons.

References:

1. Gaunt ER, Hardie A, Claas ECJ, Simmonds P, Templeton KE. Epidemiology and Clinical Presentations of the Four Human Coronaviruses 229E, HKU1, NL63, and OC43 Detected over 3 Years

Using a Novel Multiplex Real-Time PCR Method. *J Clin Microbiol*. 2010;48(8):2940-2947. doi:10.1128/JCM.00636-10

2. Alexandrova R, Beykov P, Vassilev D, Jukić M, Podlipnik Č. The virus that shook the world: questions and answers about SARS-CoV-2 and COVID-19. *Biotechnol Biotechnol Equip*. 2021;35(1):74-102. doi:10.1080/13102818.2020.1847683

3. Iadecola C, Anrather J, Kamel H. Effects of COVID-19 on the Nervous System. *Cell*. 2020;183(1):16-27.e1. doi:10.1016/j.cell.2020.08.028

4. Parvathaneni V, Gupta V. Utilizing drug repurposing against COVID-19 – Efficacy, limitations, and challenges. *Life Sci*. 2020;259:118275. doi:10.1016/j.lfs.2020.118275

5. Senanayake SL. Drug repurposing strategies for COVID-19. *Future Drug Discov*. 0(0). doi:10.4155/fdd-2020-0010

6. Pushpakom S, Iorio F, Eyers PA, et al. Drug repurposing: progress, challenges and recommendations. *Nat Rev Drug Discov*. 2019;18(1):41-58. doi:10.1038/nrd.2018.168

7. Janiaud P, Axfors C, van't Hooft J, et al. The worldwide clinical trial research response to the COVID-19 pandemic - the first 100 days. *F1000Research*. 2020;9:1193. doi:10.12688/f1000research.26707.2

8. Effectiveness of remdesivir for the treatment of hospitalized COVID−19 persons: A network meta−analysis - Jiang - - Journal of Medical Virology - Wiley Online Library. Accessed December 28, 2020. https://onlinelibrary.wiley.com/doi/full/10.1002/jmv.26443

9. Lauriola M, Pani A, Ippoliti G, et al. Effect of Combination Therapy of Hydroxychloroquine and Azithromycin on Mortality in Patients With COVID-19. *Clin Transl Sci*. 2020;13(6):1071-1076. doi:https://doi.org/10.1111/cts.12860

10. Valk SJ, Piechotta V, Chai KL, et al. Convalescent plasma or hyperimmune immunoglobulin for people with COVID−19: a rapid review. *Cochrane Database Syst Rev*. 2020;(5). doi:10.1002/14651858.CD013600

11. Godlee F. Covid-19: The lost lessons of Tamiflu. *BMJ*. 2020;371:m4701. doi:10.1136/bmj.m4701

12. Sanders JM, Monogue ML, Jodlowski TZ, Cutrell JB. Pharmacologic Treatments for Coronavirus Disease 2019 (COVID-19): A Review. *JAMA*. 2020;323(18):1824-1836. doi:10.1001/jama.2020.6019

13. Tukey J. Multiple comparisons. *J Am Stat Assoc*. 1953;48(264):624-625.

14. Dunn OJ. Multiple comparisons among means. *J Am Stat Assoc*. 1961;56(293):52-64.

15. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B Methodol*. 1995;57(1):289-300.

16. Tannock IF. False-Positive Results in Clinical Trials: Multiple Significance Tests and the Problem of Unreported Comparisons. *JNCI J Natl Cancer Inst*. 1996;88(3-4):206-207. doi:10.1093/jnci/88.3-4.206

17. Colhoun HM, McKeigue PM, Davey Smith G. Problems of reporting genetic associations with complex outcomes. *Lancet Lond Engl*. 2003;361(9360):865-872. doi:10.1016/s0140-6736(03)12715-8

18. Cleophas T, Zwinderman A. Clinical trials are often false positive: a review of simple methods to control this problem. *Curr Clin Pharmacol*. 2006;1(1):1-4. doi:10.2174/157488406775268228

19. van Ravenzwaaij D, Ioannidis JPA. True and false positive rates for different criteria of evaluating statistical evidence from clinical trials. *BMC Med Res Methodol*. 2019;19(1):218. doi:10.1186/s12874-019-0865-y

20. Ioannidis JPA. Why Most Published Research Findings Are False. *PLOS Med*. 2005;2(8):e124. doi:10.1371/journal.pmed.0020124

21. Kimmel SE, Califf RM, Dean NE, Goodman SN, Ogburn EL. COVID-19 Clinical Trials: A Teachable Moment for Improving Our Research Infrastructure and Relevance. *Ann Intern Med*. Published online June 16, 2020. doi:10.7326/M20-2959

22. Christley RM. Power and Error: Increased Risk of False Positive Results in Underpowered Studies. *Open Epidemiol J*. 2010;3(1). Accessed December 28, 2020. https://benthamopen.com/ABSTRACT/TOEPIJ-3-16

23. Medical Nihilism - Jacob Stegenga - Oxford University Press. Accessed December 28, 2020. https://global.oup.com/academic/product/medical-nihilism-9780198747048?cc=us&lang=en&

24. Kieser M, Friede T, Gondan M. Assessment of statistical significance and clinical relevance. *Stat Med*. 2013;32(10):1707-1719. doi:https://doi.org/10.1002/sim.5634

25. Wacholder S, Chanock S, Garcia-Closas M, El ghormli L, Rothman N. Assessing the Probability That a Positive Report is False: An Approach for Molecular Epidemiology Studies. *JNCI J Natl Cancer Inst*. 2004;96(6):434-442. doi:10.1093/jnci/djh075

26. Neuberger A, Oraiopoulos N, Drakeman DL. Renovation as innovation: is repurposing the future of drug discovery research? *Drug Discov Today*. 2019;24(1):1-3. doi:10.1016/j.drudis.2018.06.012

27. Ashburn TT, Thor KB. Drug repositioning: identifying and developing new uses for existing drugs. *Nat Rev Drug Discov*. 2004;3(8):673-683. doi:10.1038/nrd1468

28. Jackson RJ, Cooper KL, Tappenden P, et al. Oseltamivir, zanamivir and amantadine in the prevention of influenza: a systematic review. *J Infect*. 2011;62(1):14-25. doi:10.1016/j.jinf.2010.10.003

29. Sackett DL, MD SES, MD WSR, Rosenberg W, MD RBH. *Evidence-Based Medicine: How to Practice and Teach EBM*. 2nd edition. Churchill Livingstone; 2000.

30. Worrall J. Evidence: philosophy of science meets medicine. *J Eval Clin Pract*. 2010;16(2):356-362. doi:10.1111/j.1365-2753.2010.01400.x

31. Oxford Centre for Evidence-Based Medicine: Levels of Evidence (March 2009). Oxford Centre for Evidence-Based Medicine: Levels of Evidence (March 2009)

32. Andersen H. Mechanisms: what are they evidence for in evidence-based medicine? *J Eval Clin Pract*. 2012;18(5):992-999. doi:10.1111/j.1365-2753.2012.01906.x

33. Rocca E. The judgements that evidence-based medicine adopts. *J Eval Clin Pract*. 2018;24(5):1184-1190. doi:10.1111/jep.12994

34. Borgerson K. Valuing evidence: bias and the evidence hierarchy of evidence-based medicine. *Perspect Biol Med*. 2009;52(2):218-233. doi:10.1353/pbm.0.0086

35. La Caze A. *Evidence-Based Medicine Must Be …*. Social Science Research Network; 2009. doi:10.1093/jmp/jhp034

36. Buetow S, Kenealy T. Evidence-based medicine: the need for a new definition. *J Eval Clin Pract*. 2000;6(2):85-92. doi:10.1046/j.1365-2753.2000.00237.x

37. Clarke B, Gillies D, Illari P, Russo F, Williamson J. The evidence that evidence-based medicine omits. *Prev Med*. 2013;57(6):745-747. doi:10.1016/j.ypmed.2012.10.020

38. Clarke B, Gillies D, Illari P, Russo F, Williamson J. Mechanisms and the Evidence Hierarchy. *Topoi*. 2014;33(2):339-360. doi:10.1007/s11245-013-9220-9

39. Anjum RL, Copeland S, Rocca E. Medical scientists and philosophers worldwide appeal to EBM to expand the notion of 'evidence.' *BMJ Evid-Based Med*. 2020;25(1):6-8. doi:10.1136/bmjebm-2018-111092

40. Russo F, Williamson J. Interpreting Causality in the Health Sciences. *Int Stud Philos Sci*. 2007;21(2):157-170. doi:10.1080/02698590701498084

41. Parkkinen V-P, Wallmann C, Wilde M, et al. *Evaluating Evidence of Mechanisms in Medicine: Principles and Procedures*. Springer; 2018. Accessed December 28, 2020. http://www.ncbi.nlm.nih.gov/books/NBK543869/

42. Aronson JK, Auker–Howlett D, Ghiara V, Kelly MP, Williamson J. The use of mechanistic reasoning in assessing coronavirus interventions. *J Eval Clin Pract*. n/a(n/a). doi:https://doi.org/10.1111/jep.13438

43. Yao X, Ye F, Zhang M, et al. In Vitro Antiviral Activity and Projection of Optimized Dosing Design of Hydroxychloroquine for the Treatment of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2). *Clin Infect Dis*. 2020;71(15):732-739. doi:10.1093/cid/ciaa237

44. Liu J, Cao R, Xu M, et al. Hydroxychloroquine, a less toxic derivative of chloroquine, is effective in inhibiting SARS-CoV-2 infection in vitro. *Cell Discov*. 2020;6(1):1-4. doi:10.1038/s41421-020-0156-0

45. Rolain J-M, Colson P, Raoult D. Recycling of chloroquine and its hydroxyl analogue to face bacterial, fungal and viral infections in the 21st century. *Int J Antimicrob Agents*. 2007;30(4):297-308. doi:10.1016/j.ijantimicag.2007.05.015

46. Alani BG, Alwash AH, Ibrahim IT. Wide Applications of Chloroquine Other Than Antimalarial. *Pharmacol Amp Pharm*. 2020;11(10):251-281. doi:10.4236/pp.2020.1110022

47. Mitjà O, Corbacho-Monné M, Ubals M, et al. A Cluster-Randomized Trial of Hydroxychloroquine for Prevention of Covid-19. *N Engl J Med*. 2020;0(0):null. doi:10.1056/NEJMoa2021801

48. Boulware DR, Pullen MF, Bangdiwala AS, et al. A Randomized Trial of Hydroxychloroquine as Postexposure Prophylaxis for Covid-19. *N Engl J Med*. 2020;383(6):517-525. doi:10.1056/NEJMoa2016638

49. Hoffmann M, Mösbauer K, Hofmann-Winkler H, et al. Chloroquine does not inhibit infection of human lung cells with SARS-CoV-2. *Nature*. 2020;585(7826):588-590. doi:10.1038/s41586-020-2575-3

50. Malaria drug chloroquine does not inhibit SARS-CoV-2. https://www.dpz.eu/en/home/single-view/news/malaria-medikament-chloroquin-hemmt-sars-cov-2-nicht.html

51. Zamorano Cuervo N, Grandvaux N. ACE2: Evidence of role as entry receptor for SARS-CoV-2 and implications in comorbidities. van de Veerdonk FL, van der Meer JW, eds. *eLife*. 2020;9:e61390. doi:10.7554/eLife.61390

52. Fang L, Karakiulakis G, Roth M. Are patients with hypertension and diabetes mellitus at increased risk for COVID-19 infection? *Lancet Respir Med*. 2020;8(4):e21. doi:10.1016/S2213-2600(20)30116-8

53. Kuster GM, Pfister O, Burkard T, et al. SARS-CoV2: should inhibitors of the renin–angiotensin system be withdrawn in patients with COVID-19? *Eur Heart J*. 2020;41(19):1801-1803. doi:10.1093/eurheartj/ehaa235

54. Sanchis-Gomar F, Lavie CJ, Perez-Quilis C, Henry BM, Lippi G. Angiotensin-Converting Enzyme 2 and Antihypertensives (Angiotensin Receptor Blockers and Angiotensin-Converting Enzyme Inhibitors) in Coronavirus Disease 2019. *Mayo Clin Proc*. 2020;95(6):1222-1230. doi:10.1016/j.mayocp.2020.03.026

55. Mancia G, Rea F, Ludergnani M, Apolone G, Corrao G. Renin–Angiotensin–Aldosterone System Blockers and the Risk of Covid-19. *N Engl J Med*. Published online May 1, 2020. doi:10.1056/NEJMoa2006923

56. Meng J, Xiao G, Zhang J, et al. Renin-angiotensin system inhibitors improve the clinical outcomes of COVID-19 patients with hypertension. *Emerg Microbes Infect*. 2020;9(1):757-760. doi:10.1080/22221751.2020.1746200

57. Monteil V, Kwon H, Prado P, et al. Inhibition of SARS-CoV-2 Infections in Engineered Human Tissues Using Clinical-Grade Soluble Human ACE2. *Cell*. 2020;181(4):905-913.e7. doi:10.1016/j.cell.2020.04.004

58. Singh TU, Parida S, Lingaraju MC, Kesavan M, Kumar D, Singh RK. Drug repurposing approach to fight COVID-19. *Pharmacol Rep*. Published online September 5, 2020:1-30. doi:10.1007/s43440-020-00155-6

59. Harrison C. Coronavirus puts drug repurposing on the fast track. *Nat Biotechnol*. 2020;38(4):379-381. doi:10.1038/d41587-020-00003-1

60. PhD MM. Just follow the science: A government response to a pandemic. *J Eval Clin Pract*. 2020;26(6):1575-1578. doi:https://doi.org/10.1111/jep.13491

61. Hofmann B. The first casualty of an epidemic is evidence. *J Eval Clin Pract*. 2020;26(5):1344-1346. doi:https://doi.org/10.1111/jep.13443

62. FDA. Coronavirus (COVID-19) Update: FDA Revokes Emergency Use Authorization for Chloroquine and Hydroxychloroquine.Coronavirus (COVID-19) Update: FDA Revokes Emergency Use Authorization for Chloroquine and Hydroxychloroquine.

549 63. FDA. Coronavirus (COVID-19) Update: FDA Authorizes Drug Combination for Treatment of
550 COVID-19.https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-
551 fda-authorizes-drug-combination-treatment-covid-19.

552 64. Glickman ME, Rao SR, Schultz MR. False discovery rate control is a recommended alternative
553 to Bonferroni-type adjustments in health studies. *J Clin Epidemiol*. 2014;67(8):850-857.
554 doi:10.1016/j.jclinepi.2014.03.012

555 65. Sedgwick P. Confidence intervals and statistical significance: rules of thumb. *BMJ*.
556 2012;345:e4960. doi:10.1136/bmj.e4960

557 66. Leibovici L. Effects of remote, retroactive intercessory prayer on outcomes in patients with
558 bloodstream infection: randomised controlled trial. *BMJ*. 2001;323(7327):1450-1451.
559 doi:10.1136/bmj.323.7327.1450

560 67. Sunyoto T. Partnerships for better neglected disease drug discovery and development: how
561 have we fared? *Expert Opin Drug Discov*. 2020;15(5):531-537. doi:10.1080/17460441.2020.1736550

562 68. Laskey SB, Siliciano RF. A mechanistic theory to explain the efficacy of antiretroviral therapy.
563 *Nat Rev Microbiol*. 2014;12(11):772-780. doi:10.1038/nrmicro3351

564 69. Walter SD, Thabane L, Briel M. The fragility of trial results involves more than statistical
565 significance alone. *J Clin Epidemiol*. 2020;124:34-41. doi:10.1016/j.jclinepi.2020.02.011

566

567

568

569

570