The epistemology of the SARS-CoV-2 test

Rainer J. Klement[1], Prasanta S. Bandyopadhyay[2]

1 Department of Radiotherapy and Radiation Oncology, Leopoldina Hospital Schweinfurt, Germany

2 Department of History & Philosophy, Montana State University, Bozeman, MT, USA

* Correspondence to:
Dr. Rainer Johannes Klement
Department of Radiotherapy and Radiation Oncology
Leopoldina Hospital Schweinfurt
Robert-Koch-Straße 10
97422 Schweinfurt
Email: rainer_klement@gmx.de
Phone: +49 9721 720 2761
Homepage: www.rainerklement.com
ORCID ID: https://orcid.org/0000-0003-1401-4270

Abstract

We investigate the epistemological consequences of a positive SARS-CoV-2 test for two relevant hypotheses: (i) V is the hypothesis that an individual has been infected with SARS-CoV-2; (ii) C is the hypothesis that SARS-CoV-2 is the sole cause of flu-like symptoms in a given patient. We ask two fundamental epistemological questions regarding each hypothesis: First, given a positive SARS-CoV-2 test, what should we believe about the hypothesis and to what degree? Second, how much evidence does a positive test provide for a hypothesis against its negation? We respond to each question within a formal Bayesian framework. We construe degree of confirmation as the difference between the posterior probability of the hypothesis and its prior, and the strength of evidence for a hypothesis against its alternative in terms of their likelihood ratio. We find that for realistic assumptions about the base rate of infected individuals, $P(V) \lesssim 20\%$, positive tests having low specificity (75%) would not raise the posterior probability for V to more than 50%. Furthermore, if the test specificity is less than 88.1%, even a positive test having 95% sensitivity would only yield weak to moderate evidence for V against ¬V. We also find that in plausible scenarios, positive tests would only provide weak to moderate evidence for C unless the tests have a high specificity. One has thus to be careful in ascribing the symptoms or death of a positively tested patient to SARS-CoV-2, if the possibility exists that the disease has been caused by another pathogen.

Keywords

Bayesianism; Confirmation; COVID-19; Evidence; Medical tests

Declarations

Funding: This research received no external funding.

Conflicts of interest/Competing interests: The authors declare that they have no competing interests relevant to this research.

Availability of data and material: NA

Code availability: NA

Authors' contributions: RJK conceived of the presented idea and wrote the manuscript; PSB developed the theoretical conceptions and supervised the project. Both authors discussed the results and contributed to the final manuscript

Acknowledgments

We thank Prof. Ulrike Kämmerer for helpful discussions of the pRT-PCR method and its principles as well as Prof. Gordon Brittan Jr. for helpful comments on a previous version of this paper.

Introduction

> *Usually a question like this is theoretical: What would it be like to find your town, your state, your country, shut off from the rest of the world, its citizens confined to their homes, as a contagion spreads, infecting thousands, and subjecting thousands more to quarantine? How would you cope if an epidemic disrupted daily life, closing schools, packing hospitals, and putting social gatherings, sporting events and concerts, conferences, festivals and travel plans on indefinite hold?*
>
> (Albert Campus, *The Plague*, 1947)

It seems that Albert Campus' fictitious scenario has become reality. "The plague" affecting large parts of the World is caused by a tiny organism, a new virus called severe acute respiratory syndrome-corona virus-2, or short SARS-CoV-2. COVID-19 (coronavirus disease 2019) is the term used to describe the disease symptoms caused by the SARS-CoV-2. Recent estimates from China reveal that approximately 80% of infected individuals may be asymptomatic (Day 2020). However, symptoms of COVID-19 are typically flu-like symptoms such as fever and cough and in severe cases pneumonia, which however predominantly occurs in frail patients with other comorbidities. In general, SARS-CoV-2 appears to cause similar symptoms as other coronaviruses or influenza strains (Guan et al. 2020; Shi et al. 2020).

To prevent the virus from spreading and causing symptoms and deaths, policy makers at the helm of affairs have taken dramatic measures which include the closure of kindergartens, schools and universities, restaurants, museums and shops, prohibition of gatherings, cancellation of public events and the prohibition to leave the house without good reason. These measures are justified under the premise that SARS-CoV-2 is extremely virulent. However, the validity of this premise itself is rarely investigated. In particular, which data justify this premise? It appears that the positive testing of patients and its reporting in the media play a central role in sustaining the belief in a high virulence of SARS-CoV-2 and the notion of a pandemic. Our goal here is therefore to conduct a critical analysis of the SARS-CoV-2 testing and the conclusions that can be drawn from it.

Evidence, confirmation, and diagnostic testing

Philosophy of science is, among other things, a careful reflection on the scientific methodology that informs both discovering and justifying theories or hypotheses. The tenability of scientific hypotheses turns crucially on making inferences based on data, well-supported hypotheses, and auxiliaries. A comprehensive understanding of scientific hypotheses thus requires an understanding of scientific inference, broadly construed. However, several epistemological issues need to be distinguished in

order to appreciate the proper relationship between the tenability of scientific hypotheses and inference. We will discuss the significance of these issues/questions by borrowing an insight from Richard Royall (Royall 1997; Royall 2004). Our approach provides a unified Bayesian response to three questions posed by Royall.

Consider two hypotheses: V, stating that a patient is infected with the SARS-CoV-2, and ¬V, its denial. Assume that a SARS-CoV-2 test comes out positive. Based on this simple scenario, one could pose at least three types of question that underline the epistemological issues at stake, following an insight from Royall (1997):

(i) Given the datum, what should we *believe* about V and to what degree?
(ii) Does the datum provide strong *evidence* for V against its alternative ¬V?
(iii) Given the datum, what should we do?

We call the first question the belief or confirmation question, the second the evidence question and the third the decision question. These three questions are pre-theoretical and statistical paradigm-neutral; yet they require some statistical/probabilistic tools for their articulation. Here, we will confine ourselves to the first two questions and just briefly touch upon the decision question in the Discussion section.

We have developed two distinct accounts to answer the first two types of questions (Bandyopadhyay and Brittan 2006; Bandyopadhyay et al. 2016). The first is an account of belief/confirmation, the second of evidence. Bayesians interpret confirmation relations in various ways. For us, an account of confirmation explicates a relation, C(D,H,B) among data D, hypothesis H, and the agent's background knowledge B. For Bayesians, degrees of belief need to be fine-grained. A satisfactory Bayesian account of confirmation, according to us, should be able to capture this notion of degree of belief. In formal terms:

$$\text{D confirms H to a greater degree iff } P(H|D) > P(H) \quad (1)$$

The posterior/prior probability of H could vary between 0 and 1. Confirmation becomes strong or weak depending on how great the difference is between the posterior probability, P(H|D), and the prior probability of the hypothesis, P(H). P(H|D) represents an agent's degree of belief in the hypothesis after the data are accumulated. P(H) stands for an agent's degree of belief in the hypothesis before the data for the hypothesis have been acquired. The likelihood function, P(D|H), provides an answer to the question "how likely are the data given the hypothesis"? P(D) is the marginal probability of the data averaged over the hypothesis being true or false. The relationships between these terms, P(H|D), P(H), P(D|H) and P(D) are succinctly captured in Bayes' theorem:

$$P(H|D) = [P(D|H) \times P(H)]/P(D) > 0 \quad (2)$$

While the account of confirmation is concerned with belief in a single hypothesis, our account of evidence compares the merits of two simple statistical hypotheses, H1 and H2 (or ¬H1) relative to the data D, auxiliaries A, and background information B. However, because an evidence relation is not a belief relation, but a likelihood ratio, it need not satisfy the probability calculus. Bayesians use the Bayes factor (BF) to make this comparison, while others use the likelihood ratio (LR) or other functions designed to measure evidence. The Bayes factor in favor of H1, given A1 and B is

$$BF=P(D|H1,A1\&B)/\ P(D|H2,A2\&B) \quad (3)$$

For hypotheses under which there are unknown parameters θ, the densities[1] P(D|H,A&B) are obtained by integrating over the parameter space (Kass and Raftery 1995), so that

$$P(D|H,A\&B) = \int P(D|\theta,H,A\&B)\ \pi(\theta|H,A\&B)\ d\theta \quad (4)$$

For simple statistical hypotheses with no free parameters[2], the Bayes factor and the likelihood ratio are identical, and capture the bare essentials of an account of evidence without any appeal to prior probability. The data D constitute evidence for H1&A1& B against H2&A2&B if and only if their likelihood ratio is greater than one. An immediate corollary of the BF(LR) equation is that there is equal evidential support for both hypotheses only when BF(LR)=1. Note that in this equation if 1< BF(LR) $\lesssim$ 8, then D is often said to provide weak to moderate evidence for H1 against H2, while when BF(LR) $\gtrsim$ 8, D provides strong evidence (Kass and Raftery 1995; Bandyopadhyay et al. 2016). This cut-off point is otherwise determined contextually and may vary depending on the nature of the problem. Some data may provide very strong evidence for a hypothesis over its negation without necessarily confirming it strongly enough to make it more probable than its prior probability. It is also possible that confirmation can be very strong, but the evidence is weak (Bandyopadhyay and Brittan 2006; Bandyopadhyay et al. 2016).

In diagnostic testing, such as when testing for the new corona virus, we assume hypotheses to be simple statistical hypotheses. Let V denote the hypothesis "the patient is infected", ¬V its denial and T the proposition "the test is positive". To answer the belief and evidence questions given a positive test result, one then has to know three quantities: P(T|V), called the sensitivity of the test, P(T|¬V), the false-positive rate or one minus sensitivity, and P(V), the so-called base rate. The base rate is the proportion of infected persons within the representative population, or in other words the

[1] These probabilities of the data are also known as marginal or integrated likelihoods; some authors also denote them as "evidence" (e.g., MacKay 2004; Bailer-Jones 2017), which must not be confused with our account of evidence that always implies a comparison between *two* competing simple statistical hypotheses.

[2] See Bandyopadhyay et al. (1996) for a general model selection case when models have adjustable parameters.

probability that a randomly chosen individual is infected without a test being conducted. In the following, we summarize what is known about these crucial test statistics for the new SARS-CoV-2 tests used in clinical practice.

Sensitivity and specificity

Testing for SARS-CoV-2 is based on real-time quantitative reverse transcription polymerase chain reaction (qRT-PCR), a biomedical testing procedure that is routinely applied for detecting and quantifying RNA. Such tests are able to detect very few numbers of nucleic acid molecules – in this case RNA – by amplifying a target nucleic acid sequence several million fold. Because RNA cannot serve as a template for PCR, the RNA template is first reversely transcribed into cDNA (RT) which is then exponentially amplified in a PCR over many cycles (Bustin 2000). In quantitative (also called real time) PCR "the amount of product formed is monitored during the course of the reaction by monitoring the fluorescence of dyes or probes introduced into the reaction that is proportional to the amount of product formed, and the number of amplification cycles required to obtain a particular amount of DNA molecules is registered. Assuming a certain amplification efficiency, which typically is close to a doubling of the number of molecules per amplification cycle, it is possible to calculate the number of DNA molecules of the amplified sequence that were initially present in the sample" (Kubista et al. 2006).

Upon the rapid spread of the SARS-CoV-2 and its disease (COVID-19) in Wuhan, China, a qRT-PCR test was developed by an international group (Corman et al. 2020). The test was developed without original patient specimens, but with the help of synthetic biology based on the closely related SARS-CoV from 2003:

> "We report here on the establishment and validation of a diagnostic workflow for 2019-nCoV screening and specific confirmation, designed in absence of available virus isolates or original patient specimens. Design and validation were enabled by the close genetic relatedness to the 2003 SARS-CoV, and aided by the use of synthetic nucleic acid technology."

In their report the authors show that the SARS-CoV-2 test did not yield a positive result when applied to 297 samples of other viruses[3], indicating 100% specificity. These 297 samples also included 67 samples of human coronaviruses other than SARS-CoV-2. On the other hand, the test was able to detect other SARS-related viruses stemming from European rhinolophid bats (Corman et al. 2020), indicating cross-reactions with such virus variants and hence a certain non-specificity towards the

[3] The viruses used in this negative control validation set were adenovirus, bocavirus, several influenza A viruses, influenza B virus, metapneumovirus, parainfluenzae virus 1-4, respiratory syncytial virus (A/B), rhinovirus/enterovirus and the coronaviruses HCoV-HKU1, HCoV-OC43, HCoV-NL63, HCoV-229E and MERS-CoV.

SARS-CoV-2. Furthermore, some non-specific positive signals were detected which the authors explained by "handling issues caused by the rapid introduction of new diagnostic tests and controls during this evaluation study" (Corman et al. 2020). In an initial application of this test protocol in Bavaria, Germany, 60.7% unspecific signals were detected, but could be reduced to 5% and lower using different RT-PCR systems (Konrad et al. 2020). A Chinese validation study of the SARS-CoV-2 tests used in China revealed a false-positive rate of almost 50% or higher (Zhuang et al. 2020) – however, the study was retracted for unknown reasons soon after ahead-of-print publication. In clinical practice, false negative signals can be introduced by sample contamination and cross-reactions with other nucleic acids if the primer pairs selected are not highly specific for the target nucleic acid sequence. As an example, a study from 2006 reported on a respiratory infection outbreak in a Canadian hospital due to the human coronavirus HCoV-OC43 causing a fatality rate of 8% (Patrick et al. 2006). However, nine of 40 patient specimens were subsequently and falsely tested positive for the 2003 SARS-CoV by the National Microbiology Laboratory using RT-PCR, indicating a false-positive rate of 22.5%. The authors attributed these false-positive results to intrinsic test performance and contamination (Patrick et al. 2006).

Concerns about the sensitivity of the SARS-CoV-2 test have also been raised recently. In a laboratory validation study, sensitivity has been shown to depend on the primer/probe sets and commercial qRT-PCR kits used; good tests appear to have sensitivities in the range 90-100% (Casto et al. 2020). However, in clinical routine application, the test could have lower sensitivity as indicated by some recent publications. For example, among 4084 samples tested in a large French laboratory between January and mid-February 2020, no single case of SARS-C0V-2 infection was detected despite the samples partly stemming from 32 suspected SARS-CoV-2 cases and 337 people repatriated at the beginning of February 2020 from China (Colson et al. 2020). Xie et al. (2020) reported on five patients presenting with flu-like symptoms and radiological diagnosis of COVID-19 pneumonia, but negative qRT-PCR test results; in these patients it took multiple re-tests to obtain a positive SARS-CoV-2 detection. Finally, Li et al. (2020) reported highly variable SARS-CoV-2 test results in hospitalized patients from Wuhan, China, when multiple tests were conducted over a few days. Overall, these findings from clinical practice indicate a potentially higher false-negative rate (lower sensitivity) of the SARS-CoV-2 test than obtained in laboratory validation studies.

We thus conclude that both the sensitivity and specificity of SARS-CoV-2 tests are context-dependent[4], not well known in general and therefore unknown for a particular clinical case under consideration. Sensitivity is dependent on the initial RNA copy number concentration and

[4] Context-dependence of test performances is not restricted to the new SARS-CoV-2 tests. The reason is that such tests are not experiments within closed systems, but are conducted on open systems that are highly responsive to their environment or the given context, respectively (Klement and Bandyopadhyay 2019).

primer/probe sets used; according to the published laboratory validations so far sensitivity may be assumed high (90-100%), but is probably lower in clinical routine testing (Li et al. 2020). The specificity appears susceptible towards sample handling, contamination and cross-reactions with related viral RNA, also resulting in a potentially high false-positive rate in clinical routine testing. In the following, we therefore test different scenarios by assuming a sensitivity and specificity of 95% and 75%, respectively.

The base rate

Due to the novelty of the SARS-CoV-19 and the fact that most infections remain undetected little is known about the base rate, i.e., the probability that a randomly chosen individual is infected with the virus. The base rate is population-specific and space-time-dependent. An estimation of the base rate for a given population is difficult because most cases remain asymptomatic or only mildly symptomatic and therefore are not recorded as cases (Verity et al. 2020). To estimate the base rate, one would have to systematically test a representative sample of the target population and then use Bayes' theorem (Equation 2) together with the sensitivity and specificity of the test; this, however, has not yet been done. We therefore decide to discuss the inferences that can be made from a positive test result for a range of base rates up to 80%; 80% represents the upper limit expected for a Western country in a worst-case scenario in which no control measures or spontaneous changes in individual behavior are undertaken (Ferguson et al. 2020). However, even much more modest estimates such as "20-60% of adults" are probably still substantially exaggerated (Ioannidis 2020), placing the probable base rate somewhere in the range 5-20%.

The two questions revisited

Basic inferences from a positive SARS-CoV-2 test

Let us now assume that a patient has been tested positive for SARS-CoV-2. We are interested in the following hypotheses:

V: The patient is infected with the SARS-CoV-2

¬V: The patient is not infected with the SARS-CoV-2

Given the positive test result, what should we believe about the infection status of the patient and to which degree? This question can be answered by calculating the posterior probability of V and comparing it to its prior probability.

Given T, the positive test result, we can derive the posterior probability of V from Bayes' theorem:

$$P(V|T) = [P(T|V) \times P(V)]/P(T) \quad (5)$$

$$P(T) = P(T|V) \times P(V) + P(T|\neg V) \times P(\neg V) \quad (6)$$

The left panel of Figure 1 shows the degree of confirmation of the hypothesis that the patient is infected as a function of its prior probability (the base rate) and four combinations of assumed sensitivities and specificity of the SARS-CoV-2 test. One can see that high specificity results in a higher degree of confirmation than high sensitivity (red and blue lines). Even for low sensitivity and specificity, a positive test result would increase the probability of V by more than 20% if the base rate would be in the range 20-60%, which is probably exaggerated (Ioannidis 2020). For more reasonable estimates of P(V)<20%, confirmation would be weaker, and the posterior P(V|T) would not exceed 50% if test specificity is low. We therefore conclude that a positive test only confirms the hypothesis that an individual has been infected with the SARS-CoV-2 to a high degree if either the base rate is substantial (which is unlikely), or the test has very high specificity (which is currently unknown but can be debated).

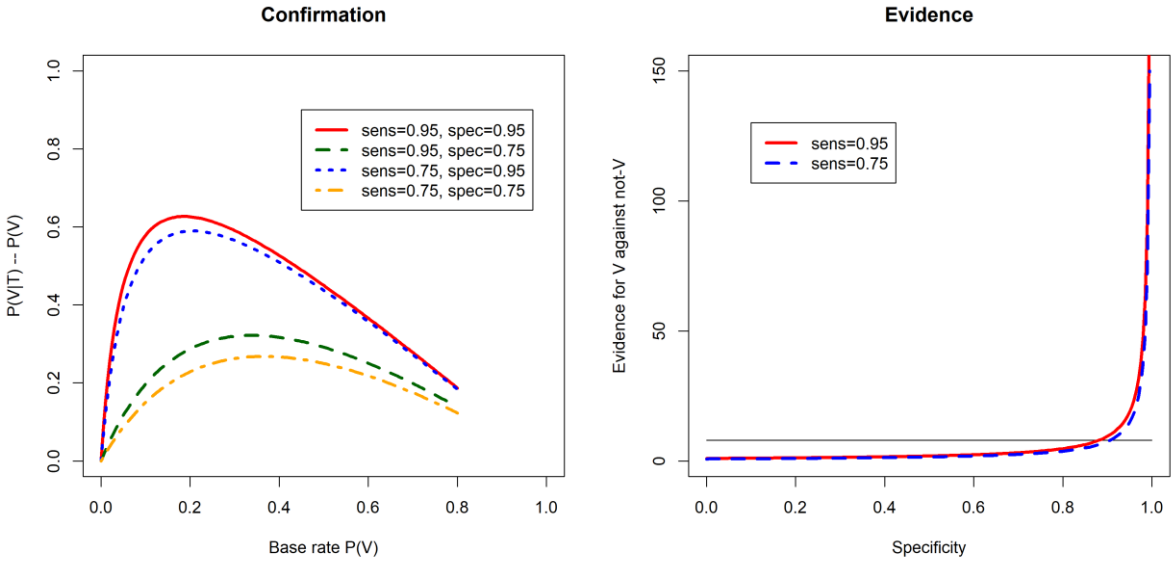

Figure 1: Left: Difference between posterior and prior probabilities for the hypothesis V "The tested patient is infected with the SARS-CoV-2" as a function of the prior probability or base rate P(V). The difference P(V|T)-P(V) is the degree of confirmation of V. Right: Evidence measured by the LR as a function of the test specificity with sensitivity fixed at 0.95 and 0.75, respectively. The black straight line denotes the threshold of a LR of 8 above which we speak of strong evidence for a hypothesis.

We next ask: how much does a positive test provide evidence for V against its alternative ¬V? The evidence for one hypothesis against its alternative is usually measured by the LR which is independent of prior probabilities and hence objective. The LR is P(T|V)/P(T|¬V). The right panel of

Figure 1 shows the LR as a function of the test specificity (=1−P(T|¬V)) for two assumed test sensitivities. For 95% or 75% sensitivity, the test would need to have a specificity of at least 88.1% or 90.6%, respectively, in order to provide strong evidence in favor of V against its alternative.

COVID-19 and the SARS-CoV-2 test

Imagine now a patient presenting with flu-like symptoms. We can then pose the following hypothesis:

C: Only SARS-CoV-2 infection has caused the flu-like symptoms of the patient.[5]

Given that COVID-19 has many symptoms in common with other corona- or influenza viruses as well as multidrug resistant bacteria common in many hospitals, the negation of C could be conceived as the catch-all hypothesis for all these other possible causes:

¬C: A pathogen other than SARS-CoV-2 has caused the flu-like symptoms of the patient.

Therefore:

¬C &V: SARS-CoV-2 infection is associated with but has not caused the symptoms of the patient.

Now imagine a SARS-CoV-2 test is conducted and turns out positive.

We can then ask to what degree the positive test result confirms hypothesis C. To this end, we construct the simple Bayesian network model depicted in Figure 2. The SARS-CoV-2 test assesses the truth of the hypothesis V, that the patient has the SARS-CoV-2 infection. As above, T stands for a positive test result. We conceive of V as a testable consequence of the hypothesis C the truth of which is both a necessary and sufficient condition for the truth of V, because if C is true for a given patient it deductively follows that he/she must be infected with SARS-CoV-2:

$$P(V|C) = 1 > P(V|\neg C) \equiv v \quad (7)$$

$$P(C|\neg V) = P(\neg V|C) = 0 \quad (8)$$

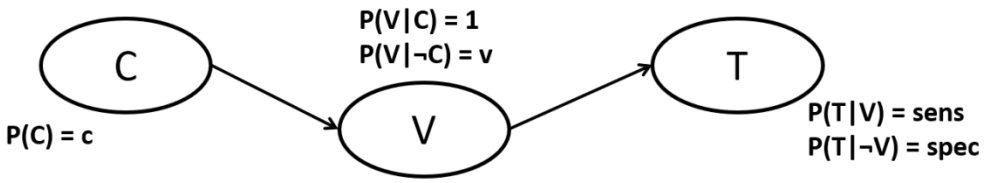

Figure 2: Bayesian network model of the relationship between a positive test T and hypothesis C.

[5] This is what most media reports implicitly assume.

According to the Bayesian network structure shown in Figure 2, C and T are independent when conditioned on V, so that P(T|C&V)=P(T|V). This can be used along with the definition of conditional probability to derive the product rule (Pearl et al. 2016) for computing the total probability of C, V and T according to Figure 2:

$$P(C\&V\&T)=P(T|C\&V)P(C\&V)=P(T|V)P(V|C)P(C) \quad (9)$$

To calculate the degree of confirmation given to C by a positive test report, we need to calculate

$$P(C|T) = P(T|C) \times P(C)/P(T) = \frac{P(T|C)\,P(C)}{P(T|C)\,P(C)+P(T|\neg C)P(\neg C)} = \frac{c}{c+(1-c)\,\frac{P(T|\neg C)}{P(T|C)}} \quad (10)$$

To calculate the LR in the denominator of Equation (10), we re-express the terms P(T|¬C) and P(T|C) using the definition of conditional probability:

$$P(T|C)=\frac{P(T\&C)}{P(C)} \quad (11)$$

$$P(T|\neg C)=\frac{P(T\&\neg C)}{P(\neg C)} \quad (12)$$

Furthermore, according to the law of total probability (Pearl et al. 2016):

$$P(T\&C) = P(T\&C\&V) + P(T\&C\&\neg V) \quad (13)$$

Using Equation (9), equation (13) can be written as

$$P(T\&C) = P(C)\,P(V|C)\,P(T|V) + P(C)\,P(\neg V|C)\,P(T|\neg V) \quad (14)$$

Using $P(V|C) = 1$ (Eq.7) and $P(\neg V|C) = 0$ (Eq.8) in Equation (14) and inserting P(T&C) into Equation (11), we obtain

$$P(T|C)=P(T|V)\equiv sens \quad (15)$$

By an analogous calculation, we obtain

$$P(T|\neg C) = P(V|\neg C)P(T|V) + P(\neg V|\neg C)P(T|\neg V) \quad (16),$$

or, using the notations given in Figure 2:

$$P(T|\neg C) = v\cdot sens + (1-v)(1-spec) \quad (17)$$

We define the LR≡x and use (15) and (17):

$$x=\frac{P(T|\neg C)}{P(T|C)} = \frac{v\cdot sens + (1-v)(1-spec)}{sens} \quad (18)$$

Inserting (18) into (10) we get

$$P(C|T) = \frac{c}{c+(1-c)x} \quad (19)$$

In Figure 3 we plot the posterior P(C|T) and the degree of confirmation P(C|T)−P(C) against the prior probability P(C). Because v≡P(V|¬C) is unknown, we adopted four different values for it: 1%, 5%, 10% and 20%. Note that v is the proportion among symptomatic cases not caused by SARS-CoV-2 who in addition to the symptom-causing pathogen carry the SARS-CoV-2. For example, Nickbakhsh et al. (2019) found that 8% of respiratory virus-positive patients from a United Kingdom population had coinfection with at least one other virus. Our choices of v reflect this finding and also cover a putatively higher coinfection rate (20%) in case that SARS-CoV-2 is highly viral. It can be seen that the confirmation we gain in our hypothesis C depends on its prior, the value for v and more on the specificity of the test than its sensitivity. In scenarios in which we would know that only a small percentage (1-10%) of symptom-bearing patients have asymptomatic SARS-CoV-2 coinfection in addition to the actual symptom-causing pathogen, a positive SARS-CoV-test in a patient would provide strong confirmation for the hypothesis that in the given case SARS-CoV-2 is the cause of the symptoms, even for non-optimal test performances. However, for larger v (20%), confirmation provided by a positive test result becomes weaker and less dependent on the test performance, in particular specificity.

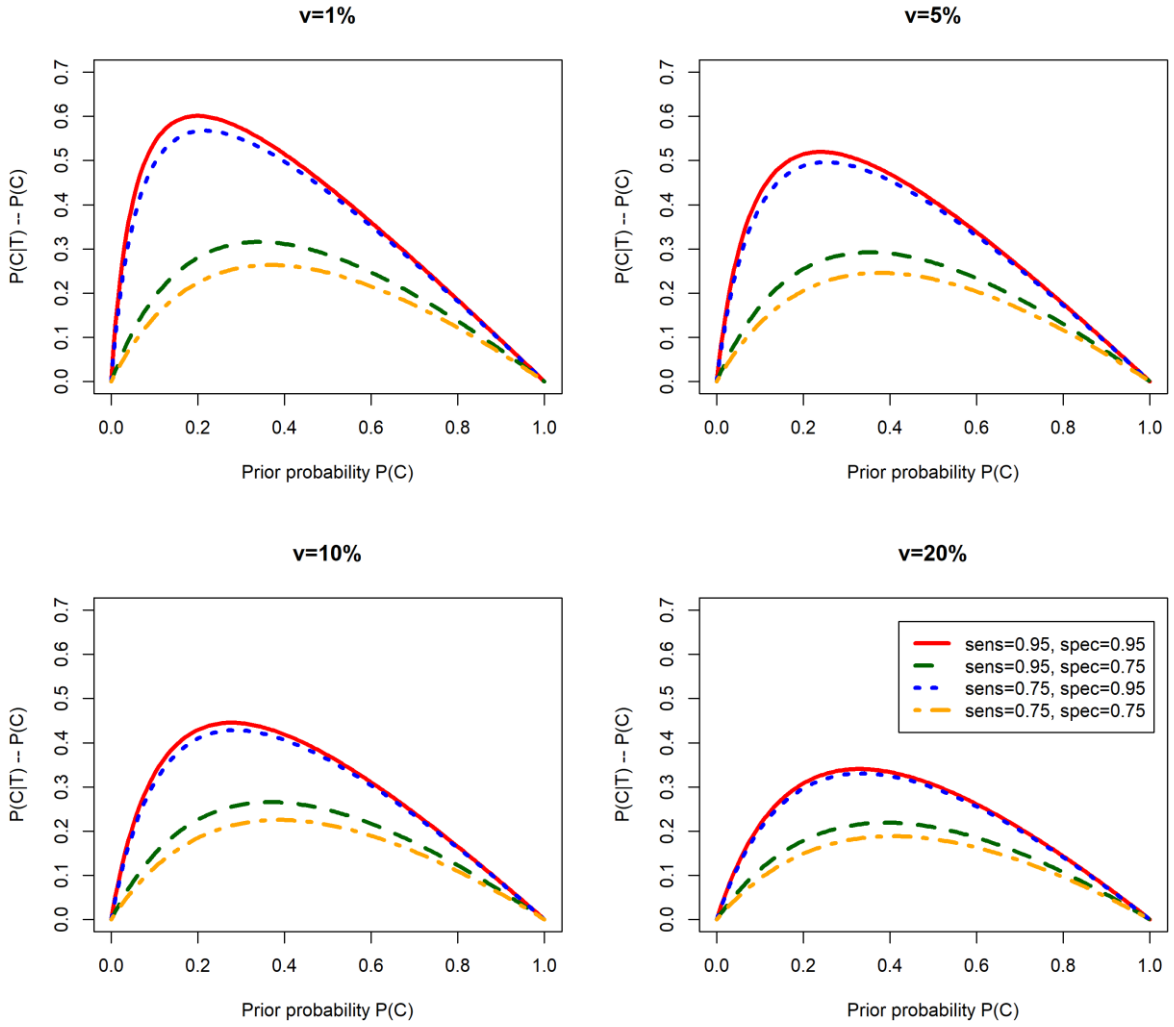

Figure 3: Confirmation of the hypothesis C (that SARS-CoV-2 caused the COVID-19-like symptoms of a patient) by a positive test as a function of the prior probability for C, v≡P(V|¬C) and different test performances.

In Figure 4 we have plotted the evidence provided by a positive test result for the hypothesis C against ¬C as a function of the test specificity for different fixed values of sensitivity and v. Note that in the depicted situations a positive test always constitutes evidence for C (likelihood ratio >1). This is consistent with the theorem that data constitute evidence for a hypothesis against its mutually exclusive and jointly exhaustive alternative iff the data confirm the hypothesis to a certain degree (Bandyopadhyay et al. 2016). Figure 4 shows that the evidence for C against ¬C becomes stronger as v decreases. In other words, as the probability for asymptomatic coinfection with SARS-CoV-2 decreases, a positive SARS-CoV-2 test would yield more evidence that these symptoms can be attributed to COVID-19. As coinfection with asymptomatic SARS-CoV-2 becomes more likely, a positive SARS-CoV-2 test constitutes less evidence for C against ¬C. Note, however, that even for low

values of v the SARS-CoV-2 test would need high specificity to constitute strong evidence for C against ¬C.

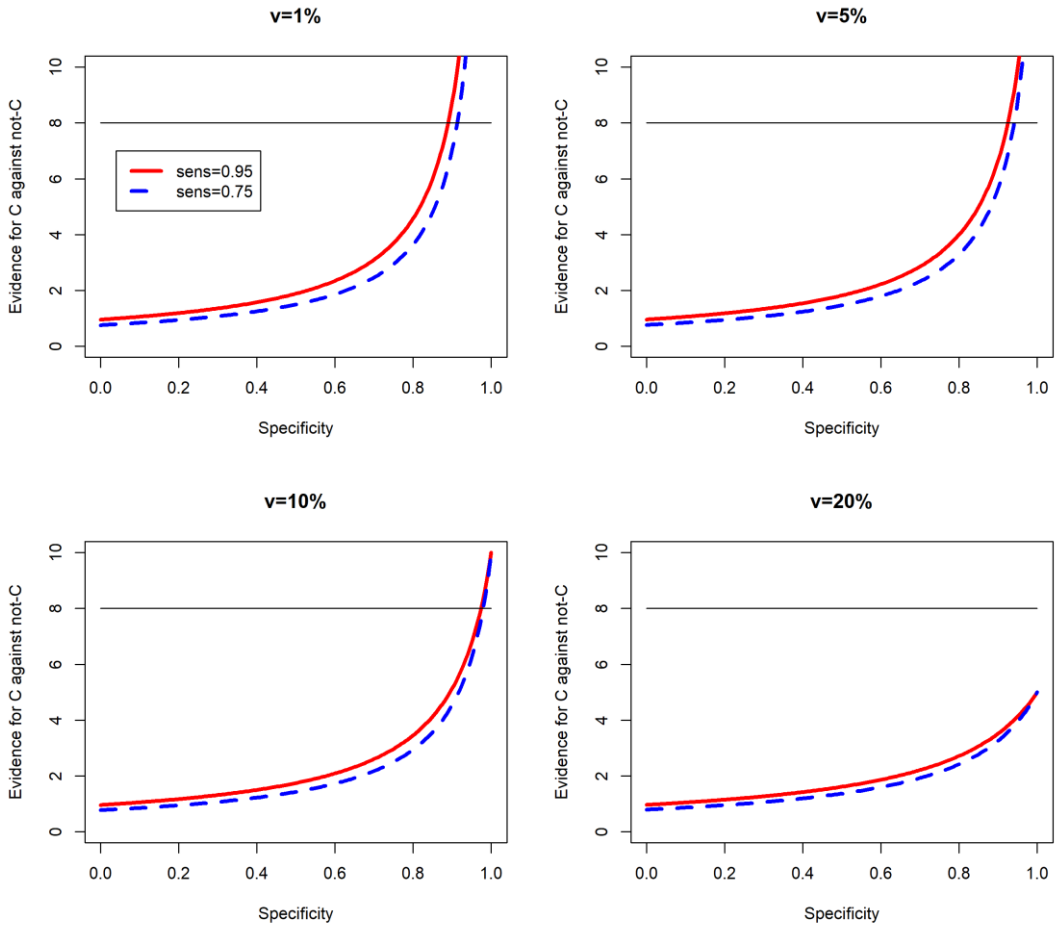

Figure 4: Evidence for the hypothesis C as a function of the test specificity for a fixed sensitivity of 75% and 95%, respectively, and different values of v, the probability that a patient who has symptoms caused by a pathogen other than SARS-CoV-2 additionally has SARS-CoV-2 coinfection. The black sold line denotes the threshold of strong evidence.

COVID-19 and a positive test for influenza A

Influenza A is one of the dominating respiratory viruses responsible for causing flu-like symptoms (Nickbakhsh et al. 2019). Imagine now that instead of a SARS-CoV-2 test, a test for influenza A has been performed in a patient with flu-like symptoms and came out positive. How does this affect an agent's degree of belief in C and the evidence for C against its alternative? Using a model with the same structure as Figure 2 and similar notation as above, let T' denote the positive test result (now for influenza A) and V' the hypothesis that an influenza A infection is present. However, now we do not have a deductive relationship between C and V' as was the case in the SARS-CoV-2 test example.

Instead, V' can be conceived as a testable consequence of C in the sense that P(V'|C) < P(V'|¬C).[6] Let us define p≡P(V'|¬C) and q≡P(V'|C). We assume p>q because if C is true, it follows deductively that the patient must have SARS-CoV-2, so that q is the probability of a coinfection with influenza A. In contrast, if ¬C is true, the symptoms are caused by a pathogen other than SARS-CoV-2, so p represents the probability of both a single-virus infection and coinfection with influenza A. Given that the probability of coinfections is considerably lower than the probability of a single-virus infection (Nickbakhsh et al. 2019), our assumption p>q is justified.

Using p>q and adopting the likelihood formula from Equation (16), the LR for the case of a positive influenza A test is

$$x=\frac{P(T'|\neg C)}{P(T'|C)}=\frac{p\cdot sens'+(1-p)(1-spec')}{q\cdot sens'+(1-q)(1-spec')}=\frac{(sens'+spec'-1)p+1-spec'}{(sens'+spec'-1)q+1-spec'}>1 \quad (20)$$

Hereby, sens' and spec' denote the sensitivity and specificity, respectively, of the influenza A test. These are better known than for the new SARS-CoV-2 test. We adopt the values derived by López Roa et al. (2011) for the influenza A qRT-PCR test which were obtained by comparison to conventional cell culture as the gold standard. These are sens'=95.6%, spec'=82.3%. Influenza A is a very common virus associated with flu-like symptoms and was present in 794 out of 3380 cases with respiratory viruses (23.5%) that were tested negatively for SARS-CoV-2 (Colson et al. 2020), so that we assume that ¬C was true in these cases. Accordingly, we set p=0.235 and vary q from zero to 0.235 (because p>q). Figure 5 plots the disconfirmation of C constituted by a positive influenza A test for four different values of q (left panel) as well as the evidence for ¬C against C as a function of q (right panel). Two results can be read off the graph: First, the smaller the chance of having coinfection with influenza A given that the symptoms are in fact caused by SARS-CoV-2, the stronger a positive influenza A test disconfirms the hypothesis C. Second, for all assumed values of q, a positive test for influenza A would only constitute weak evidence for the hypothesis that the symptoms are caused by a pathogen other than SARS-CoV-2 (¬C) against C.

[6] We thereby apply a similar definition of the testable consequence as Bovens & Hartmann (2003)in their chapter on confirmation (page 90), the difference being that the probability of the consequence given that the hypothesis is false is greater than the probability of the consequence given that the hypothesis is true in our case, while it is the other way around in Bovens & Hartmann (2003).

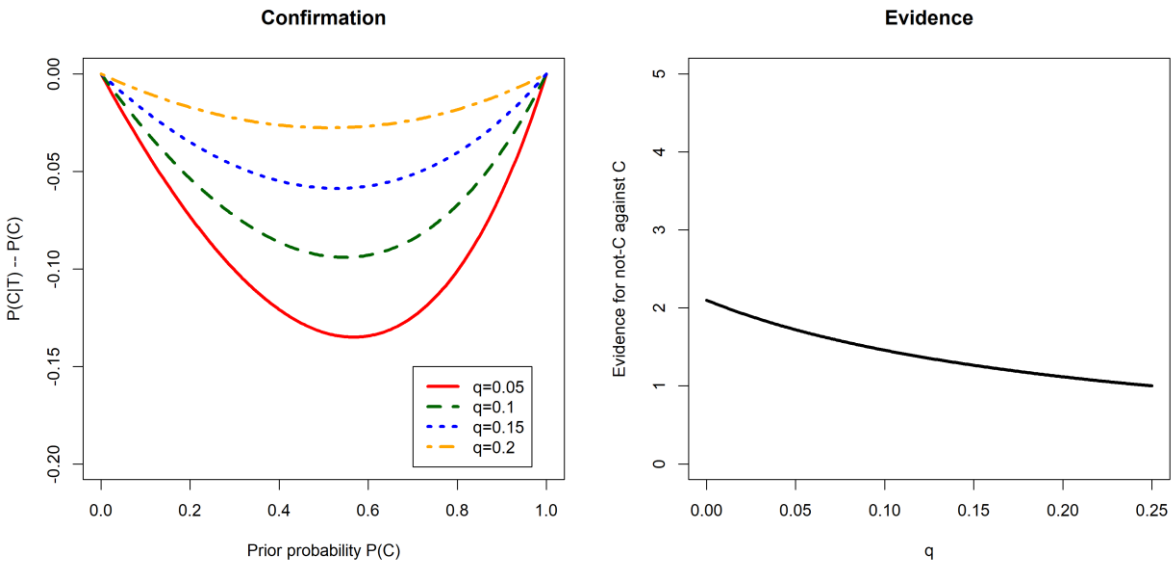

Figure 5: (Dis-)confirmation of C and evidence of ¬C against C (Eq.20) constituted by a positive influenza A test in dependence of q, the probability of having an influenza A infection when in fact the symptoms are caused by SARS-CoV-2.

Discussion

We here provided a critical investigation of the inferences that can be drawn from a positive SARS-CoV-2 test result for the two main hypotheses: One is that a patient is infected with SARS-CoV-2 (V), the second that the flu-like symptoms of a patient are caused by this virus (C) and not any other pathogen (¬C). The two epistemological questions we posed are: Given a positive SARS-CoV-2 test result (i) what should we believe about each hypothesis to what degree? (ii) What is the evidence for each hypothesis compared to its negation?

Our main result is that a positive test confirms hypothesis V to a certain degree that depends on the base rate and test performance, while only providing weak to moderate evidence for it if the test specificities are less than 88.1% or 91.6% (assuming a sensitivity of 95% or 75%, respectively). Since the base rate is probably below 20% (Ioannidis 2020), the positive test would only result in a more than 50% posterior probability that the tested person is indeed infected if the test also has high specificity (95%). A positive SARS-CoV-2 test in a patient with flu-like symptoms also confirms the hypothesis that these symptoms are caused by the SARS-CoV-2 virus, and again the degree of confirmation was found to highly depend on the test specificity. In addition, the degree of confirmation of C depends on the probability that a patient has asymptomatic SARS-CoV-2 coinfection when in fact her symptoms are caused by another pathogen. If this probability (that we denoted as v) is low, the degree of confirmation is high, and vice versa, *ceteris paribus*. The problem is that neither a reliable estimate for the SARS-CoV-2 test performance nor for v is available at the time of writing. We found that for plausible assumptions about v and test sensitivities, positive tests

would only provide weak to moderate evidence for the hypothesis that the symptoms in a given patient are caused solely be SARS-CoV-2 unless the tests have a high specificity. Unfortunately, the latter are currently unknown and may be insufficient (Zhuang et al. 2020).

One has thus to be careful in ascribing the symptoms or death of a positively tested patient to COVID-19, if the possibility exists that the disease has been caused by another pathogen. To rule out the second possibility, one would have to test for all other possible symptom causes which in practice is neither feasible nor attempted.[7] We have briefly investigated a scenario in which one positive test for a competing pathogen, the influenza A virus, is obtained. In this case, such a result would disconfirm the hypothesis that the symptoms are solely caused by SARS-CoV-2, and provide weak evidence for its alternative. This is consistent with the theorem that there is evidence for ¬C against C iff there is confirmation for ¬C (and disconfirmation of C) because we have taken C and ¬C to be mutually exclusive and jointly exhaustive (Bandyopadhyay et al. 2016). The reason that the disconfirmation and the evidence for ¬C against C are only weak is that we adopted a realistic assumption about the prevalence of influenza A infections that, despite being a very frequent pathogen causing flu-like symptoms, is only less than 25%. In principle, assuming influenza A remains similarly frequent and infectious even during the SARS-CoV-2 spread, this indicates that realistic estimates for the prior probability of C are less than ≈20%, and hence that the degree of belief in C after a positive SARS-CoV-2 test would still be below 50% for a test specificity of 75% (Fig. 3).

Since COVID-19 is a novel disease, there are still many uncertainties associated with basic statistical quantities such as the test sensitivity and specificity, the space-, time- and population-dependent base rate of infected individuals and the probability of having coinfection with SARS-CoV-2 when in fact the symptoms are caused by another pathogen (what we called v). We tried to account for these uncertainties by assuming several plausible values for these quantities. Nevertheless, empirical estimates are urgently needed in order to make the inferences that could be drawn from the current test campaigns more precise and useful.

We began with the *Plague* which has an intriguing story line pertinent to the current crisis. One of the things in the early part of the novel, Albert Camus discusses is that the narrative is based on three kinds of data: (i) what he saw, (ii) the accounts of the other eye-witnesses, and (iii) the documents that subsequently reached his hands. Unlike the narrative which is imaginary, we are right now very much inside the COVID-19 crisis. We advance an epistemological analysis of the situation based on data we have. However, what we point out in addition is that empirical estimates would also be needed in order to provide an adequate answer to the third epistemological question

[7] Nickbakhsh et al. (2019) demonstrate that the presence of multiple pathogens can be tested by multiplex qRT-PCR methods, so more research in this direction could be feasible.

that we posed in the beginning, but so far have neglected: Given a positive test result, what should we do? We can conclude that unless one is certain that the test has a high specificity, clinical decision making should not be based on such tests. Therefore, clinical evaluations and independent validation of these tests against other SARS-CoV-2 detection methodologies are urgently needed. Given that the rise of COVID-19 incidence coincides with the annual peak of other respiratory viruses (Nickbakhsh et al. 2019), the possibility of other virus infections should also be considered and at least a test for another common respiratory virus that might cause similar symptoms be performed. In general, policy makers and the media should recognize the limitations of the new SARS-CoV-2 tests and consider the possibility that deceased patients who were tested positive for this virus might only have died with but not because of it.

References

Bailer-Jones, Coryn A.L. 2017. *Practical Bayesian Inference. A Primer for Physical Scientists.* Cambridge, United Kingdom: Cambridge University Press.

Bandyopadhyay, Prasanta S, Robert J. Boik, and Prasun Basu. 1996. The curve fitting problem: A Bayesian approach. *Philosophy of Science*. 63 (Supplement): S264-S272

Bandyopadhyay, Prasanta S., and Gordon G. Brittan. 2006. Acceptibility, evidence, and severity. *Synthese* 148: 259–293. https://doi.org/10.1007/s11229-004-6222-6.

Bandyopadhyay, Prasanta S, Gordon Brittan Jr., and Mark L Taper. 2016. *Belief, Evidence, and Uncertainty: Problems of Epistemic Inference*. 1st ed. Basel: Springer International Publishing. https://doi.org/10.1007/978-3-319-27772-1.

Bovens, Luc, and Stephan Hartmann. 2003. *Bayesian Epistemology*. 1st ed. New York: Oxford University Press.

Bustin, S. A. 2000. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *Journal of Molecular Endocrinology* 25: 169–193. https://doi.org/10.1677/jme.0.0250169.

Casto, Amanda M., Meei-Li Huang, Arun Nalla, Garrett Perchetti, Reigran Sampoleo, Lasata Shrestha, Yulun Wei, Haiying Zhu, Alexander L. Greninger, and Keith R. Jerome. 2020. Comparative Performance of SARS-CoV-2 Detection Assays using Seven Different Primer/Probe Sets and One Assay Kit. *medRxiv*: 2020.03.13.20035618. https://doi.org/https://doi.org/10.1101/2020.03.13.20035618.

Colson, Philippe, Bernard La Scola, Vera Esteves-Vieira, Laetitia Ninove, Christine Zandotti, Marie-Thérèse Jimeno, Céline Gazin, et al. 2020. Letter to the editor: Plenty of coronaviruses but no SARS-CoV-2. *Eurosurveillance* 25: 6–7. https://doi.org/https://doi.org/10.2807/1560-7917.ES.2020.25.8.2000171.

Corman, Victor M., Olfert Landt, Marco Kaiser, Richard Molenkamp, Adam Meijer, Daniel Kw Chu, Tobias Bleicker, et al. 2020. Detection of 2019 novel coronavirus (2019-nCoV) by real-time RT-PCR. *Eurosurveillance* 25: 1–8. https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000045.

Day, Michael. 2020. Covid-19: four fifths of cases are asymptomatic, China figures indicate. *BMJ* 369: m1375. https://doi.org/10.1136/bmj.m1375.

Ferguson, Neil M, Daniel Laydon, Gemma Nedjati-Gilani, Natsuko Imai, Kylie Ainslie, Marc Baguelin, Sangeeta Bhatia, et al. 2020. *Report 9: Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19 mortality and healthcare demand*. London. https://doi.org/10.25561/77482.

Guan, Wei-Jie, Zheng-Yi Ni, Yu Hu, Wen-Hua Liang, Chun-Quan Ou, Jian-Xing He, Lei Liu, et al. 2020. Clinical Characteristics of Coronavirus Disease 2019 in China. *The New England journal of medicine*. https://doi.org/10.1056/NEJMoa2002032.

Ioannidis, John P A. 2020. Coronavirus disease 2019: the harms of exaggerated information and non-evidence-based measures. *European Journal of Clinical Investigation*: e13223. https://doi.org/10.1111/eci.13223.

Kass, Robert E, and Adrian E Raftery. 1995. Bayes Factors. *Journal of the American Statistical Association* 90: 773–795.

Klement, Rainer J., and Prasanta S. Bandyopadhyay. 2019. Emergence and Evidence: A Close Look at Bunge's Philosophy of Medicine. *Philosophies* 4: 50.

Konrad, Regina, Ute Eberle, Alexandra Dangel, Bianca Treis, Anja Berger, Katja Bengs, Volker Fingerle, Bernhard Liebl, Nikolaus Ackermann, and Andreas Sing. 2020. Rapid establishment of laboratory diagnostics for the novel coronavirus SARS-CoV-2 in Bavaria, Germany, February 2020. *Eurosurveillance* 25: 2000173. https://doi.org/doi:https://doi.org/10.2807/1560-7917.ES.2020.25.9.2000173.

Kubista, Mikael, José Manuel Andrade, Martin Bengtsson, Amin Forootan, Jiri Jonák, Kristina Lind, Radek Sindelka, et al. 2006. The real-time polymerase chain reaction. *Molecular Aspects of Medicine* 27: 95–125. https://doi.org/10.1016/j.mam.2005.12.007.

Li, Yafang, Lin Yao, Jiawei Li, Lei Chen, Yiyan Song, Zhifang Cai, and Chunhua Yang. 2020. Stability

Issues of RT-PCR Testing of SARS-CoV-2 for Hospitalized Patients Clinically Diagnosed with COVID-19. *Journal of Medical Virology*. https://doi.org/10.1002/jmv.25786.

López Roa, Paula, Pilar Catalán, Maddalena Giannella, Darío García de Viedma, Virginia Sandonis, and Emilio Bouza. 2011. Comparison of real-time RT-PCR, shell vial culture, and conventional cell culture for the detection of the pandemic influenza A (H1N1) in hospitalized patients. *Diagnostic Microbiology and Infectious Disease* 69. Elsevier Inc.: 428–431. https://doi.org/10.1016/j.diagmicrobio.2010.11.007.

MacKay, David J. C. 2004. *Information Theory, Inference, and Learning Algorithms*. 3rd ed. Cambridge, United Kingdom: Cambridge University Press.

Nickbakhsh, Sema, Colette Mair, Louise Matthews, Richard Reeve, Paul C.D. Johnson, Fiona Thorburn, Beatrix Von Wissmann, et al. 2019. Virus-virus interactions impact the population dynamics of influenza and the common cold. *Proceedings of the National Academy of Sciences of the United States of America* 116: 27142–27150. https://doi.org/10.1073/pnas.1911083116.

Patrick, David M., Martin Petric, Danuta M. Skowronski, Roland Guasparini, Timothy F. Booth, Mel Krajden, Patrick McGeer, et al. 2006. An outbreak of human coronavirus OC43 infection and serological cross-reactivity with SARS coronavirus. *Canadian Journal of Infectious Diseases and Medical Microbiology* 17: 330–336. https://doi.org/10.1155/2006/152612.

Pearl, Judea, Madelyn Glymour, and Nicholas P. Jewell. 2016. *Causal Inference in Statistics - A Primer*. 1st ed. Chichester, West Sussex: Wiley.

Royall, Richard. 1997. *Statistical Evidence: A Likelihood Paradigm*. New York: Chapman and Hall/CRC.

Royall, Richard. 2004. The likelihood paradigm for statistical evidence. In *The Nature of Scientific Evidence*, ed. Mark L. Taper and Subhash R. Lele, 1st ed., 119-152 (with discussion). Chicago & London: The University of Chicago Press.

Shi, Heshui, Xiaoyu Han, Nanchuan Jiang, Yukun Cao, Osamah Alwalid, Jin Gu, Yanqing Fan, and Chuansheng Zheng. 2020. Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. *The Lancet. Infectious diseases* 20. Elsevier Ltd: 425–434. https://doi.org/10.1016/S1473-3099(20)30086-4.

Verity, Robert, Lucy C Okell, Ilaria Dorigatti, Peter Winskill, Charles Whittaker, Natsuko Imai, Gina Cuomo-Dannenburg, et al. 2020. Estimates of the severity of COVID-19 disease. *medRxiv*: 2020.03.09.20033357. https://doi.org/10.1101/2020.03.09.20033357.

Xie, Xingzhi, Zheng Zhong, Wei Zhao, Chao Zheng, Fei Wang, and Jun Liu. 2020. Chest CT for Typical

2019-nCoV Pneumonia: Relationship to Negative RT-PCR Testing. *Radiology*. https://doi.org/10.1148/radiol.2020200343.

Zhuang, GH, MW Shen, LX Zeng, BB Mi, FY Chen, WJ Liu, LL Pei, X Qi, and C Li. 2020. WITHDRAWN: Potential false-positive rate among the “asymptomatic infected individuals” in close contacts of COVID-19 patients. *Zhonghua Liu Xing Bing Xue Za Zhi* 41: 485–488. https://doi.org/10.3760/cma.j.cn112338-20200221-00144.