Has CDC's COVID19 Death Ascertainment and Diagnosis Protocol Condemned Public Health and Medicine to a Sysiphian Task?

James Lyons-Weiler*
The Institute for Pure and Applied Knowledge
*jim@ipaknowledge.org

Abstract
Constructivist representations of the outlays of risk in public health have proven to be notoriously misleading. General comprehension of the impacts of population-wide testing, treatment and vaccination is based on simplified assumptions about the degree of heterogeneity of the population, and a lack of appreciation of unexpected, counter-intuitive effects. For example, in the COVID-19 pandemic, the assumption that more testing is always better carries an implicit assumption that the cost of the false positive test result is far less than the cost of the false negative test result, and that the effect of increasing the number of tests is universal improvement on the diagnostic process. This actually depends on the prevalence of the infection contagious pathogen at any given point in time. Here we present a decision-theoretic framework within which expected whole- or sub-population response outcomes can be studied, and given empirical parameter values, optimized. We show that it is possible, even with rough real-world parameter estimates, to conduct exhaustive conditioned hierarchical parameter space mapping of the framework to assess which variables in a chain of decisions will likely have the largest overall impact. The specific framework models the US CDC's published framework on cases and deaths ascertainment published in March, 2020, under which cases of death that test positive for COVID19 are presumed to be related to the virus, and under which also cases of death that are untested for the presence of COVID19 can be presumed to be related to COVID19 based on physicians' findings. CDC's death ascertainment protocol results in one pathway to true positives, one pathway to false negatives, two pathways to true negatives and six pathways to false positives. More testing under CDC's paradigm (diagnostic workflow) leads to increased false positive rates. We found that the possible general sources of variation fall into two categories (major and minor importance), each with four parameters. While numerous measures appear to point to the value of increased testing, the parameter with the largest singular impact on overall positive predictive value and the false positive rate included the number of patients who have been tested for COVID19. In spite of superficial expectations that more testing will always be helpful, the response functions of increased testing under CDC's framework includes a counterintuitive increase in false positives. Increased testing with non-zero false positives leads to the counterintuitive result of increased false positive ascertainment of COVID19 causality.

Introduction

Accuracy in the case determination of COVID-19 cases is essential for the formation of tractable, effective and realistic public health measures. However, public health policy decisions

made in the absence of knowledge of specific important features of molecular-assay guided diagnoses can lead to unexpected suboptimal effects on population health. For example, the positive and predictive value of a diagnostic framework cannot be estimated when the prevalence of an infection or disease is not known (Chen et al., 2013), and increased mass testing or screening can be problematic when the prevalence of a disease is a low (Basile et al., 2020). This fact, and the fact that SARS-CoV-2 diagnosis based on qRT-PCR status is not free from false positive results, has been acknowledged by the World Health Organization (WHO, 2021a, 2021b).

Early in the arrival of SARS-CoV-2 to the United States, CDC acknowledged that false positives could be a serious problem (CDC, 2020):

"In the event of a false positive result, risks to patients could include the following: a recommendation for isolation of the patient, monitoring of household or other close contacts for symptoms, patient isolation that might limit contact with family or friends and may increase contact with other potentially COVID-19 patients, limits in the ability to work, delayed diagnosis and treatment for the true infection causing the symptoms, unnecessary prescription of a treatment or therapy, or other unintended adverse effects." (CDC, 2020).

These concerns have been seconded, for example, by surgeons concerned with delayed life-saving surgeries due to false positive diagnoses of COVID19 (Katz et al., 2020). For clarity, we distinguish between "test accuracy" and "diagnosis accuracy" because while CDC allows case determination of cases from positive PCR results, it allows case determination with negative test results or in the absence of test results. Case reporting leads to public health data on the number of cases, number of hospitalizations, number of serious illnesses, and number of deaths is of course entirely dependent on case definition. Unfortunately, the rates of false negatives and false positives is not zero. The first reported false positive rate for qRT-PCR COVID19 case determination came from Basile et al. (2020). Their empirical false positive rate was 11%, and they correctly surmised that when prevalence of an infectious agent is low, the number of false positives can outnumber the number of true positives and false negatives.

Most public health policies centered on COVID19 case report trends are based on the assumption that "diagnosis by RT-PCR" has zero, or close to zero false positives. The technical, or test-kit false positive rate, however, is only one source false positives for diagnosis because COVID19 cases are often ascertained without testing, or even when a PCR-test is negative, but specific symptomatology even partly matches to known COVID-19 symptoms: fever with dry cough, loss of sense of smell and glass-like opacities with tissue damage to the periphery of the alveoli (radiology). In this study, we demonstrate the implementation of a decision-theoretic approach to evaluating the expected consequences of health decisions related to outbreaks, epidemics and pandemics when diagnoses are diffuse and complex. Having a thorough understanding of the consequences of public health decisions on matters such as mass quarantine and mass testing is essential to the development of public health policies that maximize benefit and minimize costs of myriad types of diagnostic inferential errors.

Wernicke warned that if the specificity of tests for SARS-CoV-2 were not increased, that most positive test results will be false positives – with reduction in prevalence leading to an increase in the false positive rate.

"If we assume a best-case scenario for specificity based on these results for the A-3 or B/E-Sarbeco setting, the diagnostic specificity was calculated as 0.9756 (40/41…). In calendar week 14 of 2020, 36,885 out of 408,348 samples (9.0%) tested positive in Germany (Robert-Koch-Institut, 2020). Under these conditions, the positive predictive value of the test system was 0.802, that is almost 20% of the positive results would have been false-positive. In calendar week 19, 10,187 out of 382,154 samples (2.7%) tested positive. In this scenario, a test system with a diagnostic specificitty(sic) of 0.9756 had resulted in a positive predictive value of 0.5319, that is almost half of the positive results would have been false-positive. Obviously, any further reduction of the prevalence of SARS-CoV-2 infections will result in decrease of the positive predictive value if the specificity of the employed assays is not dramatically increased."

Under such conditions, increased testing can lead to a numerical increase in the number of positive test results and to gross exaggeration of the number of cases if a positive PCR test result remains dependent on PCR test results. Lee (2020) found 30% false positives and 20% false negatives in a set of reference samples. Lee also argued for sequencing to validate PCR test results, or to replace qRT-PCR with sequencing to avoid false positives. The World Health Organization recently acknowledged that use of standard, high cycle threshold values for case determination leads to unacceptably high false positive values (WHO, 2021). False positive diagnoses can lead to disruptive public health measures including unwarranted quarantine, and societal lockdowns with attendant consequences such as permanent business closure and job loss. This problem has been identified as a form of cost externalization by medicine and public health on the rest of society (Lyons-Weiler, 2020).

The relationships among variables that impact the consequences of public health decisions are complex and context-dependent but can nevertheless be readily modeled as a hierarchical sorting protocol. Some relationships are well-established; for example, it is well known that the positive predictive value (PPV) of a screening test will be diminished at low prevalence, including for COVID19; Basile et al. (2020) reported 11% false positive rates in PCR-only based COVID19 diagnosis. Other relationships are not as well established, can vary over time, and can be impacted by the way public health strategies are structured and communicated. The relationships among variables can themselves be modified by relationships among other variables in way that might not be intuitive. Costs of decisions to favor specific outcomes are too often ignored and as thus externalized; this can lead to societal discord and unrest (Lyons-Weiler, 2020). In frustration, public health servants or even elected officials may feel compelled to leverage authority based on a specific set of "facts" that might not be supported by empirical evidence, such as assumption-by-assertion that there are zero false positives in qRT-PCR-based case determination.

Most of the discussion to date on "diagnostic accuracy" in COVID19 has been limited to the technical performance of qRT-PCR and antibody-based assays. One well-known exception is

the analysis by Ealy et al. (2020), based on the CDC's own reports that they changed the criteria for molecular diagnosis of respiratory illnesses treating COVID19 as a special case in which the presence of virus is considered sufficient evidence for a clinical diagnosis. The 2020 language from CDC also indicated that patients who test negative or untested patients can also, in the same framework, be diagnosed as presumed COVID19. CDC's current paradigm (announced in April, 2020; Ealy, 2020) implies a specific diagnostic workflow, that includes that patients who have died that had tested positive for SARS-CoV-2, the virus that causes COVID19, will generally be considered to have died from COVID19. Because diagnosis and death ascertainment goes beyond molecular assays, other factors must be fully considered in the characterization of the expected performance of the overall clinical diagnosis paradigm. Given the rapid onset and spread of the COVID19 epidemic in the US, transparent, accurate fully parameterized analyses of likely outcomes could not have possibly been considered.

The resulting false positives have led to chaos. A number of ad-hoc solutions have been proposed in response to shortcomings of case determination by PCR. These include repeating the test a few days following the first. Other solutions Use of >1 test with different primers (Peñarrubia et al., 2020), and sequencing part of the viral genome (Lee, 2020). None of these solutions solve other problems associated with the use of nucleic acid assays to diagnose COVID-19. Repeated testing, however, has low return-on-investment; Yamamoto et al. (2020) conducted clinical follow-up on patients with repeated PCR testing and found very low rates of clinical evidence of SARS-CoV-2 infection via computed tomography.

Here we introduce a "plug and play" decision theoretic approach toward the evaluation and characterization of the CDC's full case determination paradigm (Fig 1) and alternative paradigms that facilitates parameter mapping. This framework not only can be structurally updated; updated parameter values can be applied to refine the evaluation as more information on each parameter is learned using real-world data.

In explicit terms, if C is "COVID19", D is "death), "->" is "caused", and T+ is "tested positive", CDC's paradigm since April 2020 has been

$Pr(C \rightarrow D|T+) = 1.0$

This assertion of course had many critics, who argue that other factors such as the false positive rate of the test, and whether the virus that was present was viable, and whether the virus if viable actually contributed to the (ill) health of the patient and hastened death. This concern leads to debates over causality with respect to comorbid conditions; the counterexample of deaths from suicides, from car accidents and from non-self-inflicted gunshot wounds can nearly always be attributed to causes other that SARS-CoV-2 infection, adding ultimately to mistrust in public health.

Deaths from highly morbid comorbid conditions, however, are also often pointed to as not likely to have been caused by the SARS-CoV-2 infection, leaving an unknown percentage of deaths falsely attributed to COVID19 (See Note 1). This is an important unresolved issue because

CDC's death ascertainment protocol results in one pathway to true positives, one pathway to false negatives, two pathways to true negatives and six pathways to false positives (eight if the cycles are traced independently; Figure 1). Knowing how many persons fall into these outcome classes depends on the accuracy of each of the input parameters, which can be improved over time with real-world data.

The CDC's Semi-Hierarchical Case Determination Process

The overall CDC case determination paradigm (Fig 1) can be studied for death attribution using a decision theoretic exploration that begins with a death with a given immediate medical history that is being considered for possible causal ascertainment of "due to COVID19". A certain portion of people who die will have been tested (and some cadavers will have been tested, *a*). Of those tested, a certain portion will test "positive for COVID19" given successful PCR amplification (*b*). Of those that tested positive, only a certain portion will truly have the virus present (*c*). Similarly, some portion of individuals NOT presumed to have died from COVID19 will nevertheless have had the virus present (*d*). In our rendering, (c) and (d) can be combined (VPa).

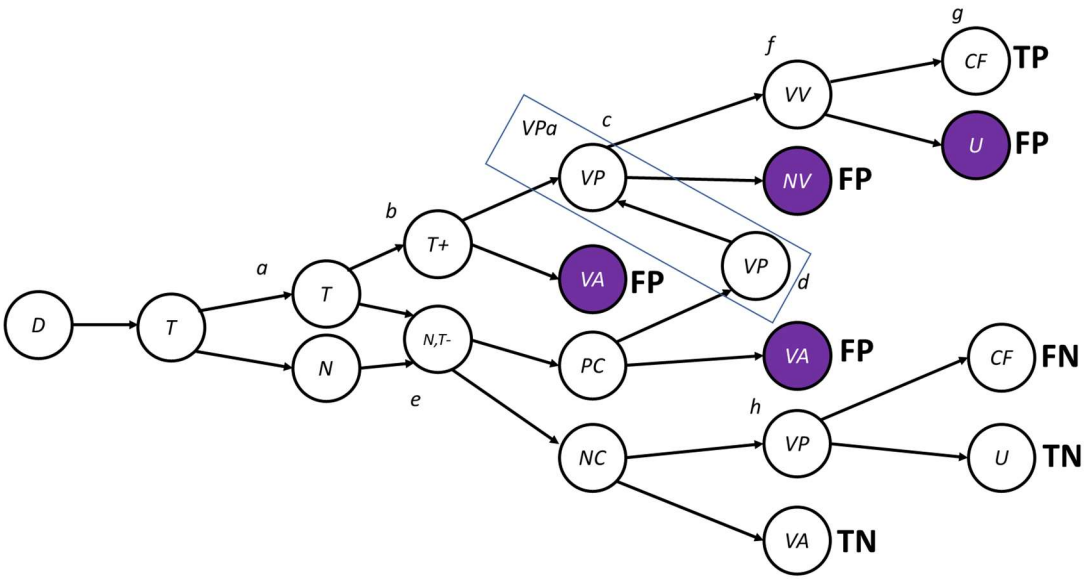

Figure 1. The Semi-Hierarchical model of the CDC's Case Reporting Paradigm for COVID19

Some patients are not tested and other will test negative. Together, these two categories make a composite group with rate (*e*).

Two final categories finish the possibilities. The penultimate category exists because the virus will be viable in only a portion of those individuals who either tested positive or were given a diagnosis with virus present (*f*). The final consideration is that the virus will have been a causal factor in a portion of patients in which the virus was present and viable (*g*).

All of these factors, arranged as a decision theoretic model, can be studied to characterize response curves under any number of conditions. In fact, even without realistic parameter value estimates for each category (which can be seen either as rates or probabilities), the framework allows a comprehensive evaluation of the expected performance of any number of possible configurations - and, importantly, allows us to ask the question: what is the relative contribution of variation in parameters values in each of the testing and diagnosis factors?

Information flow in semi-hierarchical systems with dynamic inputs can lead to unanticipated results, including trade-offs that lead to anti-correlations; propagation of errors through such systems can lead to undesirable overall diagnostic performance and unforeseen consequences. Parameter relational mapping is an important first step to creating improved general comprehension of the effects of decisions leading to overall structure of such systems.

Methods

We chose the public health decision framework adopted by the US CDC for case and death attribution to COVID19 as the basis of our analysis. In March, 2020, CDC issued both a new guidance and follow-up document that favored the diagnosis of COVID19 in cases in which comorbid conditions might also, or better explain a particular health outcome or death. The language differed from prior years in a number of important ways, including the use of the presence of the virus as sufficient evidence for a diagnosis of COVID19, and the allowance of the presumption of death from COVID19 as a primary cause of death indicated by a positive molecular test (Ealy et al., 2020). Under CDC's framework, deaths can be assumed to be due to respiratory illnesses in spite of lack of direct evidence of causality even when evidence of SARS-CoV-2 exposure, such as positive molecular assay result, are not available.

The semi-hierarchical model is a realistic overall representation of the CDC's case determination process paradigm; however, precise modeling would require parameter values require empirical data for each input parameter. Accurate empirical data leading to "generally true" estimates for each of the parameter values at each node are not available. Further, the correct input values for each parameter (Table 1) can change with context (e. g. , patient source). We therefore adopted the approach of varying each parameter value of interest to assess the scale and direction of the resulting functional response curves to specific overall performance evaluation measures. No cost considerations are built into the model; future modeling can be adapted using a utility function.

Symbol		**Description (Default Heuristic value)**
A.	α	testing effort (85%)
B.	β	% of tests that are "positive for COVID" (95%)
C.	χ	% tested pos w/virus present (60%)
D.	δ	% presumed not COVID with virus present (25%)
E.	ε	% negative test or no test presumed COVID (25%)
F.	ϕ	% presumed COVID with virus present (75%)
G.	γ	% of positive test or diagnosis with virus present where virus is viable (75%)
H.	η	% patients with virus present and viable where virus is causal factor (95%)

The performance measure responses are, importantly, relevant to the entire public health paradigm reflected in Figure 1, not merely the qRT-PCR testing conducted now as a matter of routine, which is most directly related to the parameter χ.

In addition to SN and SP, a variety of canonical performance evaluation measures were also calculated to provide additional key indicators on the outcome of policy decisions reflecting the parameter set configurations. The battery of standard performance evaluation measures is calculated for each configuration of parameter settings; these are defined in the Appendix.

Our parameter mapping numerical experiments involved holding all default parameter values constant and varying each parameter across the range of possible values.

Analysis 1. Effect of Percentage of Test Positive Samples that Contain the Virus (χ): on PPV, FPR and FDR

The most commonly considered variable that would impact a diagnostic outcome is the influence of the percentage of test positive samples that actually contain the virus (χ). This is related to disease prevalence, although the presence of a positive test may not actually indicate a viable, transmissible infection. We characterized these responses by holding all default parameter values constant and varying χ. It can be expected that increases in this percentage will increase the positive predictive value (PPV; Appendix), a reduction in the false positive rate (FPR), and a reduction in the false discovery rate (FDR).

Analysis 2. Effect of χ on PPV, SP and PLR
The increase in χ should not have any effect on the specificity (SP), but it should have a linear increase in positive predictive value (PPV). It should also have an increase in the positive likelihood ratio (PLR; Appendix).

Analysis 3. Effect of Increased Testing on the False Positive Rate (FPR)

It is known that at a low prevalence, increased testing can increase the total number of false positives; this is why the preferred test for screening would have zero false positives. Unfortunately, due to technical variation, false positives of qRT-PCR as currently implemented are not zero. Our consideration extends beyond the molecular test at point *a* (Fig 1); the calculation of the performance evaluations extend over the entire decision framework. Thus, the increase in the number of tests at a fixed non-zero false positive rate will lead to more presumed cases, thus populating the lower half of the hierarchy of events. Even if all individuals were tested, some would test negative, leading to individuals with test-negative presumed COVID-19 - only some of which will be true COVID-19 cases. Thus, concern over the decision framework false negative rate exists. However, balance between FN and FP has not been previously sought. With all assumptions on the table for evaluation, varied testing effort - the number of tests applied overall - to map the FPR response curve, keeping all other parameters at their default values.

Analysis 4. Effect of the Number of COVID-19 Test Positive Cases that Test Positive on FPR

We similarly varied the number of COVID-19 test positive cases that were “condition positive”, whether they actually contain the virus or not (β), also mapping the FPR response curve, keeping all other parameters at their default values.

Further Analyses

To study the impact of the other variables, each of the other input parameters (Table 1) were similarly varied, holding the other heuristic parameter values constant. *The response curves for SN and SP were determined* as each of the independent variables were changed.

Relative Importance Analysis

To help inform on the relative importance of securing accurate input parameters related to making public health decisions. the relative criticality of each of the parameters, each demonstrative parameter values were determined as follows: were varied away from its representative settings (reduced and increased) by incremental amounts (-20% to +20% of each individual input variable). The effect of suboptimal parameter inputs on the combined SN+SP allows the determination of the relative contribution of each factor to variation in overall accuracy. To determine relative importance, the absolute value of the difference of sum of SN and SP at +20 and -20% of each variable was calculated. These were expressed as relativized values compared to the minimum value of the absolute difference (ε). The provides a measure of the general scale of impact of change (improvements or deterioration) of each factor in the context of the heuristic parameter settings.

The multifactorial nature of the complex diagnostic paradigm adopted by CDC can be challenging to understand via pairwise comparisons of variables. To expedite, we created a graph called CHIRP Diagrams. CHIRP diagrams, which resemble songbird sound waveforms, can be view while the input variable settings are adjusted to study, at a glance, the impact of various or individual independent variables.

CHIRP Diagrams

The multifactorial nature of the complex diagnostic paradigm adopted by CDC can be challenging to understand via pairwise comparisons of variables. To expedite, we created a graph called CHIRP Diagrams. CHIRP diagrams, which resemble songbird sound waveforms, can be view while the input variable settings are adjusted to study, at a glance, the impact of various or individual independent variables.

CHIRP diagrams plot successive response curves reflective multiple parameter sweeps, with each swept variable holding the same position in the plot. Our swept variables of interest are labelled A-H, reflecting their reference symbols. As each variable is swept from minimum to maximum, a given response variable being studied can be examined considering swept variables. CHIRP diagrams provide both a tool for understanding relationships among possible

input and expected outcome parameters as well as a situational dashboard to public health can know what to expect given a particular diagnostic workflow to help ascertain when population-wide screening or indiscriminate testing makes sense.

CHIRP diagrams plot successive response curves reflective multiple parameter sweeps, with each swept variable holding the same position in the plot. Our swept variables of interest are labelled A-H, reflecting their reference symbols. As each variable is swept from minimum to maximum, a given response variable being studied can be examined considering swept variables.

To provide examples for study, we created CHIRP diagrams for Positive Likelihood Ratio (LH+), the Negative Likelihood Ratio (LH-), the Diagnostic Odds Ratio (the ratio of the odds of the test being positive if the subject has the disease relative to the odds of the test being positive if the subject does not have the disease), and F1-score, a measure of the accuracy of a diagnostic test or workflow.

Results

It is important to re-iterate that the specific response curves embodied in these results are relevant in the context of the overall diagnostic/death ascertainment paradigm (Fig 1), not merely for the molecular test δ. Other than immediate dependencies and anti-correlations enforced by the paradigm, and the exact heuristic configuration of the non-varying parameters, the response curves resulting from variation in each input parameter characterizes the relationship in a manner independent of variation in any of the other variables.

Analysis 1

In the context of the overall diagnostic/death ascertainment paradigm, as the % of Test Positive Samples w/Virus Present (δ) increases, the response function PPV increases, whereas both the false positive rate (FPR) and false discovery rate (FDR) decrease (Fig 2A). The false discovery rate drops much faster than the false positive rate.

As χ increases, PPV, accuracy and the overall positive likelihood ratio (PLR) increases; specificity remains unchanged (Fig 2B).

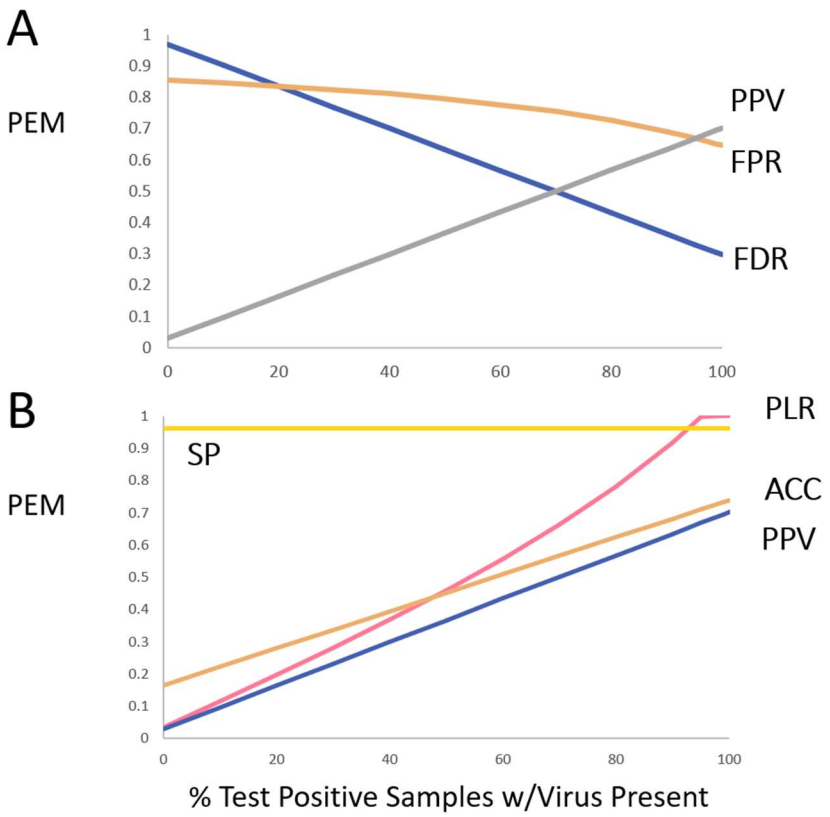

Fig 2. Response curve functions of the percentage of test positive samples that have virus material present (χ). PEM = Performance Evaluation Measure; PPV = Positive Predictive Value; FPR = false positive rate; FDR = false discovery rate; ACC = accuracy; PLR = positive likelihood ratio (terms all defined in Appendix).

Analysis 2.
Paradoxically, testing effort and percent COVID19 positive test results has a major increasing effect on FPR and a minor *reduction* on overall SN (Fig 2A and 2B, respectively).

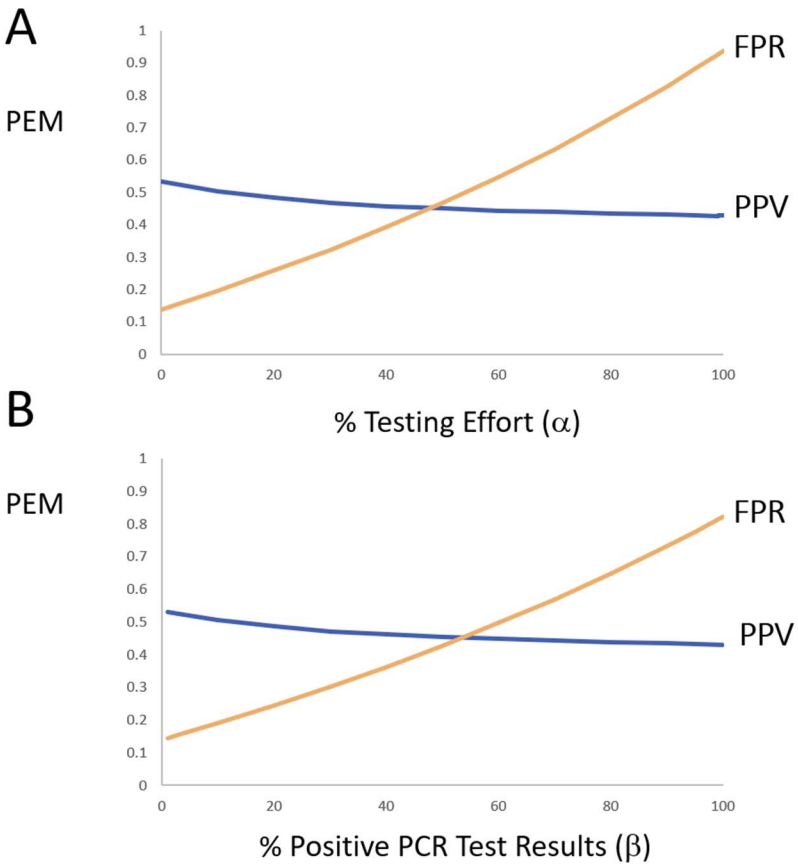

Figure 3. Response curves of the FPR (orange) and SN (blue) to the χ parameter (percentage of test positive patients with viral material present). The response curves are relevant to the overall case determination process and may therefore seem paradoxical to expectation of a molecular test used in isolation. Increase in % Positive PCR Test Results (β) has the same effect on SN (slight decrease) and a slightly delayed effect on the increase in the false positive rate.

Additional Analyses
Analysis 3.

Percentage of positive tests with virus present, virus viable (γ)
The effect of increased viral present in the sample, which would be related to prevalence, is a major linear increase in PPV, but only a minor decrease in the false positive rate at the default parameter values (Fig. 5).

Percentage of positive tests with virus present, virus viable (γ), and causal (η) The impact of variation in the percentage of positive tests with the virus present, viable and causal is nearly identical to that of γ (% of positive tests with virus present, virus viable) (Fig. 6).

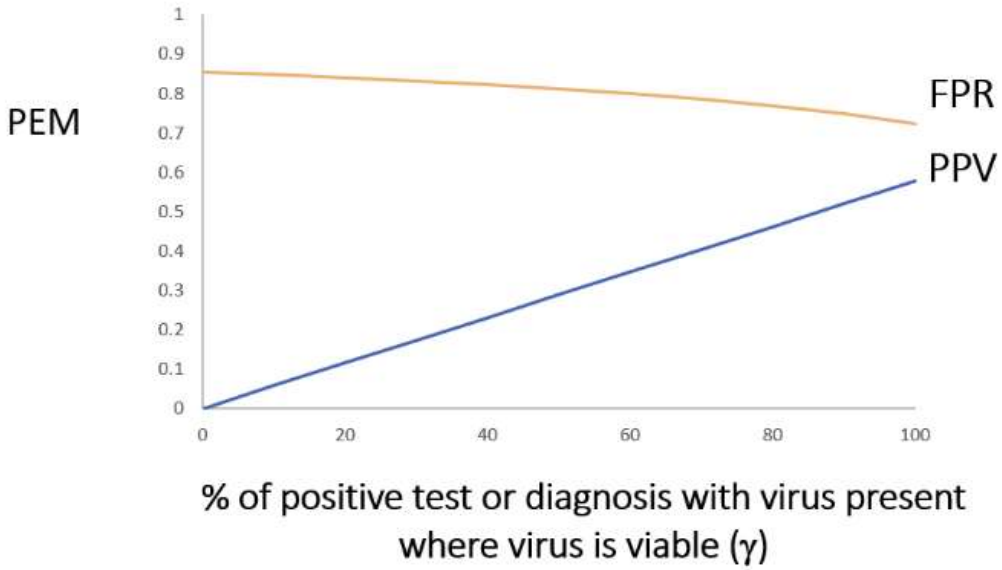

Figure 5. The percentage of patients in which the virus is present and viable increases PPV and has a small decreasing response in FPR *($and FDR, not shown*). Variation in the causality of the has the same individual singular parameter impact on the PEMs

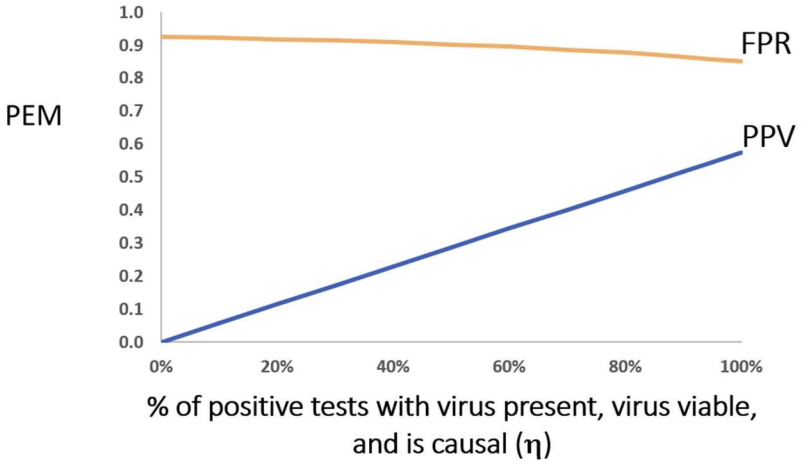

Figure 6. The percentage of patients in which the virus is present and viable and causal has fairly the same singular effect on SN and FPR *(and FDR, not shown*). Variation in the causality of the has the same individual singular parameter impact on the PEMs

Relative Impact Analysis – Impact Per Unit Change

Our findings imply that the CDC's decision to allow presumed COVID19 cases (δ) has the largest overall impact on the accuracy (or inaccuracy) of the case determination process (Figure 7), followed by χ (percentage of test positive patients with the virus present), γ (percentage of

test positive patients with the virus present and viable), η (% of positive tests in which the virus is present, virus viable, and is causal). These four represent the four major potential sources of variation in positive predictive value and specificity. The remaining four variables (ϕ, β, α, ε) each singularly contribute far less to the overall variation in SN and FPR (See Figures in Supplementary File S1). The functions between all variables and overall accuracy were linear.

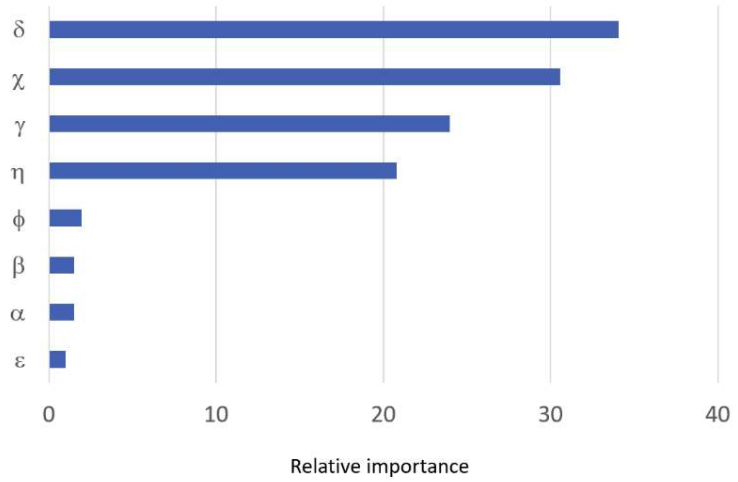

Figure 6. Relative Importance of Parameters in the CDC Diagnostic Paradigm

CHIRP Diagrams

To provide examples for study, we created CHIRP diagrams for Positive Likelihood Ratio (LH+), the Negative Likelihood Ratio (LH-), the Diagnostic Odds Ratio (the ratio of the odds of the test being positive if the subject has the disease relative to the odds of the test being positive if the subject does not have the disease), and F1-score, a measure of the accuracy of a diagnostic test or workflow.

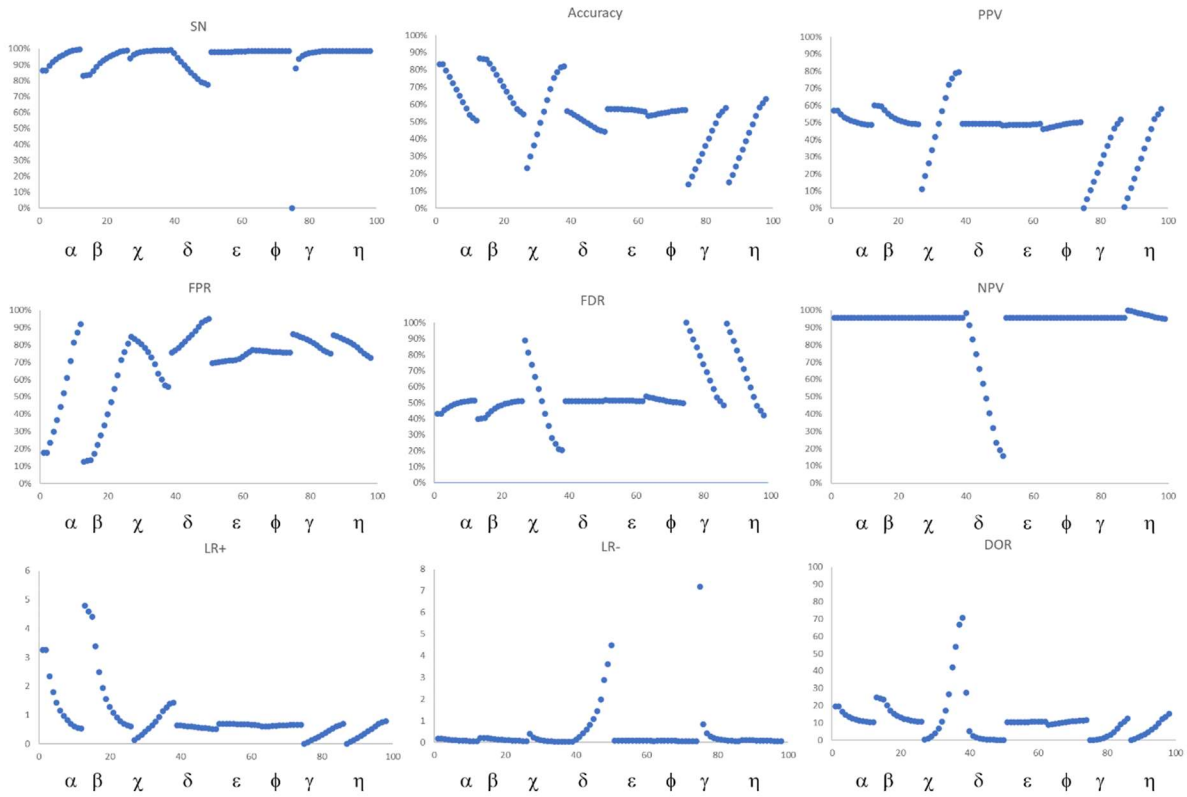

Fig 7. CHIRP Diagrams for Variables of Interest. Response variable (on y-axes) vary according to variation in the input variables (X-axis). Values within each input variable range from 0% to 100% per dectet.

The CHIRP diagrams (Fig 7) reveal a few surprises; relative to other variables, sensitivity (the ability to detect the virus when it is, in fact, present and causal) is less affective (i.e., is more robust) overall to variations in conditions than the other response variables. This means concern over false negatives should be lower than concern over false positives. Negative predictive value takes a nosedive as the % presumed not COVID with virus present (δ) increases; the other response variables show that this practice can be expected to have little beneficial impact in terms of capturing missed cases. In policy terms, assuming "died with" = "died from" comes at a far greater cost than benefit. Under the default parameter settings, this assumption becomes more problematic with greater use, to the extent that at the highest value of δ, the diagnostic odds ratio falls below 1.0, meaning that inverting the diagnosis would prove more accurate than following CDC's directive. This is seen also in the increased false positive rate (FPR) and the drop in overall accuracy. The cost is reflected both in decrease in the diagnostic odds ratio (DOR) and in the dramatic increase in the apparent prevalence (Fig 7) as δ is increased.

Note that the current analysis does not reflect variations on the default input values, but we provide the Excel spreadsheet so others may study the impact of variation on the default values of input parameters as increased empirical estimates are published, or as context changes. For

example, an informed analysis using default parameter values that reflect nursing homes would be informative.

Discussion

Our analyses have led to a greater understanding of the potential impact of various sources of variance in death attribution related to factors that impact differential diagnosis and to factors in the current SARS-CoV-2 diagnostic workflow that reveal it as problematic. This most important finding confirms the counter-intuitive negative attendant consequences of the “died with = died from” paradigm adopted by CDC in April, 2020; the PCR testing effort has leads to net increase in the false positive rate in overall context, creating a negative of return on investment (in terms of useful clinical information) using any test with non-zero FPR in case determination.

Public health policies that incorporate widespread testing requires understanding the intricacies of the expected relationships and responses in the performance evaluation measures in clinical diagnosis. Our analysis demonstrates how counterintuitive, self-defeating performance responses can occur in complex molecular test-aided diagnostic frameworks.

Our straightforward heuristic decision theoretic framework is first step toward parametric dynamic model that can aid in the prioritization of resource use in the ascertainment of deaths during the COVID19, or any other, pandemic if population-wide solutions are sought. We found that in terms of death ascertainment, the variables that can be controlled: testing effort, prevalence-based decision to switch to mass testing, test accuracy (technical SN+SP), assessment of virus viability and accuracy in determination of virus causal contribution to deaths are all impactful, albeit singularly decrease in the per-unit change of impact. This fact will have relevance if an empirically based public health policy on death ascertainment is seen as superior to constructivist public health policy.

Our results provide an idea of the relationship among and thereby the effects of the influence of specific factors. For example, the false positive rate (FPR) is also known as the “probability of a false alarm” and is the sum of the false positive outcomes divided by the number of true condition negatives, whereas the discovery rate (FDR) is the probability that a positive call is false (also the portion of the number of positive calls that are false). The FPR is the risk in a given population of making a false diagnosis, whereas the FDR is the risk that any of the positive diagnoses that have been made are false. Both of these rates are important to consider in widespread testing; the FDR, however, tells us out of the number “cases”, which one might be false. As in multiple hypothesis testing, the total number of false diagnoses will increase as a function of both the number of people tested and the stringency (or laxness) of the test. In qRT-PCR case determination, the use of high cycle thresholds has led to much confusion over the total number of cases and deaths, a fact acknowledge by the World Health Organization (WHO, 2021a,b).

It is now known how many cases of “asymptomatic transmission” of SARS-CoV-2 and “repeat infection” have been attributed to false positive diagnoses, but the number of actual asymptomatic transmissions must be low. An immense study by Cao et al., (2020) with

complete contact tracing found zero new cases that could be attributed to individuals without symptoms. This lends support toward consideration of false positive rates in the estimation of public health data, including numbers of cases and deaths attributed to COVID-19.

CDC updated their guidance for HCPs returning to work following SARS-CoV-2 infection prioritizing symptom clearance over health care providers who are not immunocompromised (CDC, 2020). This came following assessments of the limited utility and complexity of repeat PCR testing (e.g., Greene et al., 2020).

Emerging variants pose a threat to the longevity of PCR testing kits (e.g.,Ramírez et al., 2020; Peñarrubia et al., 2020). Our study also lends further support for the urgency to move nucleic acid assay technology being applied to SARS-CoV-2 detection and case determination to include viral sequence determination. A stunning 8.5% of variants in accumulating “new variants” are found in known PCR test kit sites (Peñarrubia et al., 2020). Sanger sequencing with overlapping, nested primers can detect test escape variants. Full genomic sequences can be recovered from viable virion populations; Letizia et al. (2020) were unable to produce full sequences from 37% of samples of closely tracked marine recruits who had tested positive for COVID19 via PCR-only case determination. How many of these represent false positives is unknown, but if whole-genome sequencing is used a gold standard to validate PCR, its false discovery rate could be unacceptably high. The ongoing escape of variants from PCR testing is the other side of the coin, leading also to unacceptably high false negative rates, making PCR testing not only futile, but grossly misleading and damaging to the task of differential diagnosis of respiratory viral infections. Sequencing for testing would allow far more straightforward diagnosis, but should be used as confirmatory of symptoms. Screening leads to diagnosis by molecular test, which fails to demonstrate infectivity or even determination of clinical disease.

Limitations

Clearly, this heuristic analysis can be modified by varying the representative heuristic parameter values; however, the point of this exercise is to demonstrate the feasibility and relative ease with which a decision theoretic paradigm can be used to help society understand the likely source of variation in the ascertainment of death in the complex setting of molecular testing influenced diagnoses. This analysis did not consider pairwise covariation of parameters’ effects on LOI, but the spreadsheet implementation is provided as Supplementary File 1 to allow others to explore its use in specific settings. For example, the implementation may vary substantially from setting to setting. Updates will be informative as testing strategies and clinical diagnoses improve.

General Considerations

Our results do not inform on the impact of the relative complexity or depth of options used to conduct case determination and thus does not by itself inform on the selection of alternative frameworks. Prior work indicates that if choosing among frameworks involving categorical variables with varying numbers of levels, information gain in decision trees tends to be biased in favor of those attributes with more levels (Deng et al., 2011). Selection among alternatives can be aided by clinical decision modeling (e.g., Shi & Lyons-Weiler et al., 2014). Although not

explicitly stated as such, our model is inherently Bayesian due to the effects of decisions made at earlier nodes on later nodes.

This approach is clearly useful for comparing competing paradigms and identifying the most important factors that influence emergent relationships and their expected bounds in the real world. It must be stressed that the approach presented is relevant only for heuristic estimation toward optimization of policies for ascertainment of death. It does not consider the prevalence of disease and is not configured to be used to try to learn about policies on overall diagnosis. That will require additional parameters. For instance, when prevalence is low, test settings skewed to avoid false negatives will tend to lead to overall testing workflows that lead to lowered predictive value (increased false positive). Our second study in this series will provide a similar heuristic assessment of qRT-PCR-based testing.

The relative importance analysis makes the implied assumption that each % change in all variables are equally feasible; clearly this is related to the investment cost of bringing about increases or decreases in each variable, as well as the real-world costs of false negative and false positive diagnoses and deaths ascertained to COVID19. Essential decisions in public health must consider the balance of risks of costs of inferential errors (Lyons-Weiler, 2020). We hope our analysis is useful in putting to bed the question of ascertainment bias under CDC's protocol; the bias is necessarily present in any world in which their paradigm is used, and especially so when the molecular tests in use result in false positives. More sophisticated models in the future can be derived that give due consideration to real-world data on investment costs.

Funding.
This study was funded by donations from the public to the Institute for Pure and Applied Knowledge. We are grateful to Andre Angelantoni for error-checking an early version of our analyses.

Literature Cited

(CDC, 2021). Return to Work Criteria for Healthcare Personnel with SARS-CoV-2 Infection (Interim Guidance) CDC https://www.cdc.gov/coronavirus/2019-ncov/hcp/return-to-work.html#TestBasedStrategy

Chen FY, Yang CP, Chen PY. Comment on the correct use of predictive values for evaluating diagnostic tests. Radiology. 2013 Jan;266(1):364-5; discussion 365-6. doi: 10.1148/radiol.12121346. PMID: 23264531.

Basile K, Maddocks S, Kok J, Dwyer DE. Accuracy amidst ambiguity: false positive SARS-CoV-2 nucleic acid tests when COVID-19 prevalence is low. Pathology. 2020 Dec;52(7):809-811. doi: 10.1016/j.pathol.2020.09.009. Epub 2020 Sep 30. PMID: 33087255; PMCID: PMC7524665. https://pubmed.ncbi.nlm.nih.gov/33087255/

Cao S, Gan Y, Wang C, Bachmann M, Wei S, Gong J, Huang Y, Wang T, Li L, Lu K, Jiang H, Gong Y, Xu H, Shen X, Tian Q, Lv C, Song F, Yin X, Lu Z. Post-lockdown SARS-CoV-2 nucleic acid screening in nearly ten million residents of Wuhan, China. Nat Commun. 2020 Nov 20;11(1):5917. doi: 10.1038/s41467-020-19802-w. PMID: 33219229; PMCID: PMC7679396.

Deng, H.; Runger, G.; Tuv, E. (2011). Bias of importance measures for multi-valued attributes and solutions. Proceedings of the 21st International Conference on Artificial Neural Networks (ICANN).

Ealy, H, M McEvoy, D Chong, J Nowicki, M Sava, S Gupta, D White, J Jordan, D Simon and P Anderson. COVID-19 Data Collection, Comorbidity & Federal Law: A Historical Retrospective. Science, Public Health Policy & the Law Oct 2020 2:4-22.

Greene DN, Dickerson JA, Greninger AL, Schmidt RL. When To Retest: an Examination of Repeat COVID-19 PCR Patterns in an Ambulatory Population. J Clin Microbiol. 2020 Aug 24;58(9):e01179-20. doi: 10.1128/JCM.01179-20. PMID: 32554480; PMCID: PMC7448670.

Katz AP, Civantos FJ, Sargi Z, Leibowitz JM, Nicolli EA, Weed D, Moskovitz AE, Civantos AM, Andrews DM, Martinez O, Thomas GR. False-positive reverse transcriptase polymerase chain reaction screening for SARS-CoV-2 in the setting of urgent head and neck surgery and otolaryngologic emergencies during the pandemic: Clinical implications. Head Neck. 2020 Jul;42(7):1621-1628. doi: 10.1002/hed.26317. Epub 2020 Jun 12. PMID: 32530131; PMCID: PMC7307014.

Lee, SH. 2020. Testing for SARS-CoV-2 in cellular components by routine nested RT-PCR followed by DNA sequencing International Journal of Geriatrics and Rehabilitation 2(1) 69-96.

Letizia AG, Ramos I, Obla A, Goforth C, Weir DL, Ge Y, Bamman MM, Dutta J, Ellis E, Estrella L, George MC, Gonzalez-Reiche AS, Graham WD, van de Guchte A, Gutierrez R, Jones F, Kalomoiri A, Lizewski R, Lizewski S, Marayag J, Marjanovic N, Millar EV, Nair VD, Nudelman G, Nunez E, Pike BL, Porter C, Regeimbal J, Rirak S, Santa Ana E, Sealfon RSG, Sebra R, Simons MP, Soares-Schanoski A, Sugiharto V, Termini M, Vangeti S, Williams C, Troyanskaya OG, van Bakel H, Sealfon SC. SARS-CoV-2 Transmission among Marine Recruits during Quarantine. N Engl J Med. 2020 Dec 17;383(25):2407-2416. doi: 10.1056/NEJMoa2029717. Epub 2020 Nov 11. PMID: 33176093; PMCID: PMC7675690.

Peñarrubia L, Ruiz M, Porco R, Rao SN, Juanola-Falgarona M, Manissero D, López-Fontanals M, Pareja J. Multiple assays in a real-time RT-PCR SARS-CoV-2 panel can mitigate the risk of loss of sensitivity by new genomic variants during the COVID-19 outbreak. Int J Infect Dis. 2020 Aug;97:225-229. doi: 10.1016/j.ijid.2020.06.027. Epub 2020 Jun 12. PMID: 32535302; PMCID: PMC7289722.

Ramírez JD, Muñoz M, Hernández C, Flórez C, Gomez S, Rico A, Pardo L, Barros EC, Paniz-Mondolfi AE. Genetic Diversity Among SARS-CoV2 Strains in South America may Impact

Performance of Molecular Detection. Pathogens. 2020 Jul 17;9(7):580. doi: 10.3390/pathogens9070580. PMID: 32708840; PMCID: PMC7400710.

Shi H, Lyons-Weiler J. Clinical decision modeling system. BMC Med Inform Decis Mak. 2007 Aug 13;7:23. doi: 10.1186/1472-6947-7-23. PMID: 17697328; PMCID: PMC2131745.

US CDC. FACT SHEET FOR HEALTHCARE PROVIDERS Updated: December 1, 2020 https://www. cdc. gov/coronavirus/2019-ncov/downloads/Factsheet-for-Healthcare-Providers-2019-nCoV. pdf

Wernike K, Keller M, Conraths FJ, Mettenleiter TC, Groschup MH, Beer M. Pitfalls in SARS-CoV-2 PCR diagnostics. Transbound Emerg Dis. 2020 Jun 14:10.1111/tbed.13684. doi: 10.1111/tbed.13684. Epub ahead of print. PMID: 32536002; PMCID: PMC7323359

World Health Organization, 2021a. WHO Information Notice for IVD Users 2021/05: Nucleic acid testing (NAT) technologies that use polymerase chain reaction (PCR) for detection of SARS-CoV-2 https://www.who.int/news/item/20-01-2021-who-information-notice-for-ivd-users-2020-05
Accessed 3/7/2021

World Health Organization 2021b. COVID-19 Weekly Epidemiological Update Weekly epidemiological update - 16 February 2021. https://www.who.int/publications/m/item/weekly-epidemiological-update---16-february-2021

Yamamoto K, Saito S, Hayakawa K, Hashimoto M, Takasaki J, Ohmagari N. When should clinicians repeat SARS-CoV-2 RT-PCR?: Repeat PCR testing targeting patients with pulmonary CT findings suggestive of COVID-19. Jpn J Infect Dis. 2020 Aug 31. doi: 10.7883/yoken.JJID.2020.531. Epub ahead of print. PMID: 32863359.

--
References

Lyons-Weiler, 2020a. Balance of Risk in COVID-19 Reveals the extreme cost of the false positives. Intern J Vacc Theor, Pract, Research 1(2):209-222.

Lyons-Weiler, J. 2020. Lyons-Weiler, J. 2020b. Plan B public health infrastructure and operations oversight reform for America. Intl J Vacc Theor, Pract, Research 1(2):283-294.

Madewell ZJ, Yang Y, Longini IM Jr, Halloran ME, Dean NE. Household Transmission of SARS-CoV-2: A Systematic Review and Meta-analysis. JAMA Netw Open. 2020 Dec 1;3(12):e2031756. doi: 10.1001/jamanetworkopen.2020.31756. PMID: 33315116; PMCID: PMC7737089. https://pubmed.ncbi.nlm.nih.gov/33315116/

US Food and Drug Administration. Memorandum. Guidelines for the Validation of Analytical Methods for Nucleic Acid Sequence-Based Analysis of Food, Feed, Cosmetics and Veterinary Products. Accessed 1/30/2021 https://www.fda.gov/media/121751/download

US Food and Drug Administration Emergency Use Authorization (EUA) information, and list of all current EUAs. https://www.fda.gov/emergency-preparedness-and-response/mcm-legal-regulatory-and-policy-framework/emergency-use-authorization