

# Binary Classification Tests, Imperfect Standards, and Ambiguous Information\*

Gabriel Ziegler<sup>†</sup>

Most recent version available [\[here\]](#)

November 15, 2022

## ABSTRACT

Diagnostic test performance is usually measured relative to a reference test, which almost always detects the underlying health status only imperfectly. In this paper, I argue that partial identification of the relevant performance measures can be described as producing ambiguous information when the reference test is imperfect. This allows for a phenomenon called dilation—an extreme form of non-informativeness. As an example, I present hypothetical test data satisfying the WHO’s minimum quality requirement for rapid Antigen tests which leads to dilation. The ambiguity arises because of a missing data problem: the joint distribution of health status and test results is not observed. I construct the (usually non-singleton) set of all these possible joint distributions, which allows me to assess the new test’s informativeness. This analysis leads to a simple sufficient condition to make sure that a new test is not a dilation. I illustrate my approach with applications to data from three COVID-19 related tests. Two rapid Antigen tests satisfy my sufficient condition easily and are therefore informative. However, less accurate procedures, like chest CT scans, may exhibit dilation.

KEYWORDS: Binary tests, ambiguity, information, dilation, SARS-CoV-2, COVID-19

---

\*Thanks to Dan Sacks, Charles Manski, and Jörg Stoye for literature pointers. I thank Filip Obradovic for very valuable comments. Christopher Staphenurst provided excellent research assistance. All errors are of course mine.

<sup>†</sup>University of Edinburgh (School of Economics) & University of Pittsburgh (Department of Economics); [ziegler@ed.ac.uk](mailto:ziegler@ed.ac.uk).

# 1 INTRODUCTION

An important aspect of evaluating a new diagnostic test is to assess its accuracy. Intuitively, a sensible binary test should have test results highly correlated with the underlying health condition. In other words, a positive test result should be likely if and only if the tested person is indeed infected or sick.<sup>1</sup> However, establishing whether a person is truly infected is often costly or even impossible. Therefore, a new test is analyzed relative to an established test. An established test is perfect when a positive test result occurs if and only if the person is truly infected. The medical literature calls these perfect tests a “gold standard” (Watson et al., 2020). In these situations the joint distribution of the new test’s outcomes and the underlying true health condition is the *same* as the joint distribution of test results from both tests. Thus, this observed joint distribution can be used to evaluate the new test’s accuracy.

In practice, however, a perfect reference test does not exist. In such a case, the researcher would need the joint distribution of the health conditions and the outcomes of both tests.<sup>2</sup> This overall joint distribution is not observable (or maybe only if the researcher incurs a high costs for obtaining the data). This missing data problem leads to two distinct problems: (i) the marginal distribution of the underlying health condition is missing and (ii) the *correlation* between new test’s outcome and health status is missing too.<sup>3</sup> The latter of these problems will introduce ambiguity in the information provided by the new test.

The first of these problems, missing data about the underlying health condition, is well-known. Recently, Manski and Molinari (2021) use methods known from the literature on partial identification to provide bounds on prevalence—the fraction of infected people in the population.<sup>4</sup> Measuring prevalence is different from the usual inference problem because the tested population might not be representative of the overall population. Here the data are observed selectively which corresponds to a *selection problem* as introduced by Manski (1989). Furthermore, Manski (2020) illustrates how this problem carries over to evaluating accuracy of new

---

<sup>1</sup>In the following, I will not differentiate between being infected and being sick.

<sup>2</sup>Depending on the question, it might suffice to consider the joint distribution of the new test’s outcomes and the underlying true health condition. For example, the analysis Subsection 2.2 requires only this bivariate joint distribution. The trivariate distribution is needed to evaluate the informativeness of performing both tests as discussed in Subsection 2.3.

<sup>3</sup>In the introduction, I use the word correlation loosely and informal.

<sup>4</sup>Similar approaches were used by Stoye (2021) and Sacks et al. (forthcoming).

tests in the context of COVID-19 Antibody tests maintaining the assumption of a perfect reference test.

The second problem of missing data about the correlation is different in nature and avoided when a perfect reference is available. Even if one would assume knowledge of prevalence, potentially multiple 'correlation structures' are consistent with the observed data. The reason for this multiplicity is well-known from copulas as studied in probability theory. Knowledge of prevalence provides the marginal distribution of the health condition, whereas the observed testing data provides (a bivariate) marginal distribution. In general, there are multiple (trivariate) joint distributions with these marginal distributions. Due to this multiplicity a simple, unambiguous interpretation of the new test is not possible. Without knowledge of prevalence, the problem identified before carries over and therefore exacerbates the overall multiplicity. However, as discussed in more detail later, the ambiguous information stems only from the missing data on correlation and therefore occurs whether or not the researcher has knowledge about prevalence.

In this paper, I provide a theoretic framework that combines insights from [Manski and Molinari \(2021\)](#) and [Stoye \(2021\)](#) about selective testing with the missing correlation data due to an imperfect reference test. Within this framework, it is possible to address informativeness of both tests. First, [Proposition 1](#) shows that the established test's negative predictive value<sup>5</sup> is usually not given by a unique number, but it always informative nevertheless. This multiplicity arises because of problem (i) only. Then, I analyze the new test's informativeness for the test population only. The focus on the tested population simplifies the algebra and furthermore shuts down the ambiguity about prevalence (cf. problem (i)) and therefore allows me to study the essence of ambiguous information for the new test in separation (problem (ii) only). Finally, I study the implications on informativeness if both effects are present.

Studying the informativeness of tests has a long tradition in probability theory, statistics, economics, and philosophy. [Blackwell \(1951, 1953\)](#) introduces a notion of "(more) informative" for (what is now called Blackwell) experiments.<sup>6</sup> An experiment is a mapping from states of the world to a distribution over signals.

---

<sup>5</sup>A test's negative predictive values is the probability of being healthy conditional on obtaining a negative test result. Another important informativeness measure is the positive predictive value, which is the probability of being infected conditional on a positive test result. I will assume throughout that the established has a perfect predictive value in line with the application to SARS-CoV-2 testing.

<sup>6</sup>[de Oliveira \(2018\)](#) provides a more recent treatment.

In the current setting, an experiment is a function that associates a distribution over test results to each of the possible health conditions, i.e. for being infected and for being healthy. In such a setting, the value of information is defined as the amount a Bayesian decision maker is willing to pay for the experiment. Since every experiment is more informative than an uninformative experiment,<sup>7</sup> Blackwell’s theorem shows that the value of information is (weakly) positive for every Bayesian decision maker.<sup>8</sup> Ideally a diagnostic test should satisfy Blackwell’s definition of an experiment in order to ensure that it is always informative. However, this is typically only true for the established test in my framework.

The new test fails to be a Blackwell experiment because it does not map each state to a unique distribution over test results. Rather, due to the multiplicity of joint distributions, there is a *set* of distributions over test results for a given health condition.<sup>9</sup> Therefore, Blackwell’s informativeness notion does not apply to the new test. Furthermore, the value of information needs to be adjusted because a Bayesian analysis does not readily apply with sets of probabilities. Such a situation is usually referred to as a situation of “ambiguity” and the literature has identified several extensions of Bayesian decision making to the realm of ambiguity.<sup>10</sup>

Instead of defining the value of information for a specific decision criterion in such a situation, I adopt a very weak notion of informativeness: the diagnostic test is informative if and only if it is not a *dilation*. Seidenfeld and Wasserman (1993) introduce the notation of dilation for situations with sets of probabilities. In the current context, a dilation occurs if, no matter what test result is obtained, the set of probabilities conditional on this information contains the original set of probabilities. Figure 1 illustrates an example of a dilation. Here, the set of probabilities indicating the infection likelihood before the test (black set) lies within both sets after the test result (blue for a positive result and red indicating the set after a negative result). Thus, in a sense, the decision maker is worse-off after taking the test than before taking the test *no matter* what the test result is. For this reason, Seidenfeld and Wasserman call a dilation a “counterintuitive phenomenon” and Gul and Pesendorfer (2018) refer to it as “all news is bad news”.

---

<sup>7</sup>An experiment is uninformative if the mapping mentioned above is a constant function?

<sup>8</sup>More generally, Blackwell characterizes his notion of “more informative” with the requirement that every Bayesian decision maker has a higher value of information for the more informative experiment.

<sup>9</sup>Formally, the new test can be seen as a correspondence or set-valued function.

<sup>10</sup>Machina and Siniscalchi (2014) provide a recent overview about this topic.

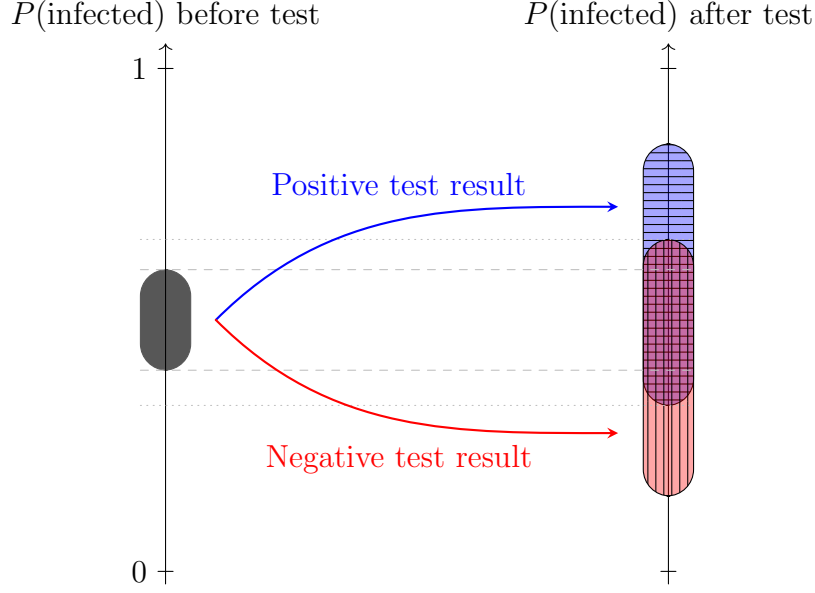


Figure 1: A diagnostic test as dilation

Table 1: Dilation Test Data

$z \backslash y$	$y = 0$	$y = 1$	Sum
$z = 0$	39.5%	11.5%	51%
$z = 1$	1.2%	47.8%	49%
Sum	40.7%	59.3%	

$y = 1$  indicates a positive PCR-test result (i.e. the established test).  $z = 1$  denotes a positive Antigen test.

My framework allows to fully characterize when a new diagnostic test is a dilation (cf. [Expression \(3\)](#)). Since the definition of informativeness for tests is rather weak, any reasonable test should satisfy this criterion. The characterization provides a method to verify whether the new test is informative.

The [WHO \(2020\)](#) recommends a minimum standard of accuracy for rapid Antigen tests.<sup>11</sup> Usually a PCR test is the established test used to evaluate these Antigen tests ([Esbin et al., 2020](#)). [Table 1](#) illustrates hypothetical test data, which fulfill the minimal requirements of the [WHO](#). However, as the analysis will reveal, this test is actually a dilation and therefore not informative.<sup>12</sup> For minimum required accuracy standards, the dilation characterization provides an easy to

<sup>11</sup>For the informed reader, the [WHO](#) recommends a sensitivity of at least 80% and a specificity of at least 97%. These measures will be formally introduced and defined later.

<sup>12</sup>This statement depends on the accuracy of the established PCR test. A dilation occurs only if the PCR has sensitivity at the lower range identified by the literature.

verify sufficient condition to avoid dilation. The new test is informative (in the population of tested people) if

$$\frac{\text{Fraction of people with a positive new test}}{\text{minimum sensitivity of new test} \times \text{sensitivity of established test.}} \leq 1$$

Besides theoretic applications of dilation, there is not much empirical evidence in the literature yet. Recently, economists started to investigate dilation and ambiguous information experimentally. The only experiment focusing on how decision makers react to dilation and relate the behavior to the value of information is conducted by [Shishkin and Ortoleva \(2021\)](#). To the best of my knowledge, the possible occurrence of dilation with diagnostic tests (or SARS-CoV-2 tests more specifically) is the first observation of this phenomenon ‘in the field.’<sup>13</sup>

Of course, researchers studying diagnostic tests are well aware of the general issues addressed here. The problem of selection leading to unobserved prevalence is known as *Verification Bias*, whereas the problem arising from unobserved correlation due to an imperfect reference test is descriptively named *Imperfect Gold Standard Bias*. ([Zhou et al., 2014](#), Chapters 10–11) This paper is not the first to document that either of these problem leads to non-identified models; rather the novelty of this paper comes in the approach. Diagnostic test research seeks to avoid non-identified models by introducing additional assumptions and then address resulting biases relative to a baseline assumption. Proposed methods include simply imputing missing data or considering more sophisticated correction methods. Moreover, the two problems are often addressed separately. By contrast, my framework requires minimal assumptions and addresses both problems simultaneously.<sup>14</sup>

## 2 MAIN ANALYSIS

I consider the following situation. Let  $x = 1$  denote that a person is infected and  $x = 0$  if the person is healthy. Initially, there is binary test available where  $y = 1$

<sup>13</sup>[Manski \(2018\)](#) mentions that a dilation might occur in a different medical context, but does not address this issue further.

<sup>14</sup>[Reitsma et al. \(2009\)](#) provide a flowchart as guidance for applied researchers to address several problems arising when establishing accuracy of diagnostic tests. The two problems addressed here are in two distinct branches of the flowchart.

indicates a positive test result and  $y = 0$  a negative result. Finally, a new test is introduced which again can be either positive ( $z = 1$ ) or negative ( $z = 0$ ).

Let  $P(x, y, z)$  denote the population distribution under consideration with  $p := P(x = 1)$  denoting prevalence. However, the population distribution is not directly observable. This is, of course, almost always the case, because a researcher usually only observes a sample from the population distribution. This leads to the usual inference problem. Throughout, I will abstract away from inference altogether. Instead, the data is given for people who were tested to obtain data on the new test. For this denote tested people with  $t = 1$  and  $t = 0$  otherwise. Then, the data are given by  $P(y, z|t = 1)$  and I assume that  $P(t = 1) > 0$ .<sup>15,16</sup>

Furthermore, since the established test is well-known, precise information about the sensitivity and specificity of this test is available as well. The following assumption ensures that both of these measures are well defined.

**Assumption 1** (Non-trivial prevalence). *The population satisfies  $p \in (0, 1)$ .*

With this assumption, sensitivity and specificity for the initial test are respectively defined as:

$$P(y = 1|x = 1) = \frac{P(x = 1, y = 1)}{p} =: \sigma, \quad \text{and} \quad (1)$$

$$P(y = 0|x = 0) = \frac{P(x = 0, y = 0)}{1 - p}. \quad (2)$$

As discussed in [Manski \(2020\)](#), for decision making sensitivity and specificity are not the relevant measures. The relevant measures are *positive predictive value (PPV)* and *negative predictive value (NPV)*. For the established test these measures can be obtained from specificity and sensitivity via Bayes' rule if prevalence  $p$  and  $P(y = 1)$  are known:

$$\begin{aligned} \text{PPV}_y &:= P(x = 1|y = 1) = \frac{p}{P(y = 1)} P(y = 1|x = 1) = \frac{p}{P(y = 1)} \sigma \\ \text{NPV}_y &:= P(x = 0|y = 0) = \frac{1 - p}{P(y = 0)} P(y = 0|x = 0). \end{aligned}$$

---

<sup>15</sup>Furthermore, the following logical implications of (not) being tested hold: (i)  $t = 0 \implies z = 0$  and (ii)  $z = 1 \implies t = 1$ . Note that  $y = 1$  is possible even if not tested, because the participation pool concerns only the new test.

<sup>16</sup>Equivalently, the data is given by sensitivity and specificity of the new test *relative* to the established test with the additional information about how many established or new tests had a positive result.

Since the tested people are usually not representative of the overall population,<sup>17</sup> even for the established test these two measures are not point-identified. (Manski and Molinari, 2021; Manski, 2020; Stoye, 2021)

To simplify the analysis and in-line with the application to SARS-CoV-2 testing, I also consider the following three baseline assumptions.

**Assumption 2** (No False-Positives for established test). *The population satisfies  $P(x = 0, y = 1) = 0$ .*

Assumption 2 implies that the established test achieves a maximum specificity and PPV<sub>y</sub> of 1.<sup>18</sup>

Additionally, I will assume *test-monotonicity* as in Manski and Molinari (2021), meaning conditional on being tested the probability of being infected is greater than if not being tested.<sup>19</sup>

**Assumption 3** (Test-monotonicity). *The population satisfies  $P(x = 1|t = 1) \geq P(x = 1|t = 0)$ .*

Lastly, I assume that the established test's sensitivity does depend on the underlying health status  $x$ , but not on whether the person is in the testing pool  $t = 1$ .<sup>20</sup>

**Assumption 4** (Health-sufficiency). *The population satisfies  $P(y = 1|x = 1, t = 1) = P(y = 1|x = 1) = \sigma$ .*

To reduce cumbersome lengthy notation in the following, I will use this simplified notation henceforth:

$$\begin{aligned} \gamma &:= P(y = 1|t = 1) && \dots \text{test yield for established test} \\ \zeta &:= P(z = 1|t = 1) && \dots \text{test yield for new test} \\ \tau &:= P(t = 1) && \dots \text{measure of data representativeness.} \end{aligned}$$

---

<sup>17</sup>For example, supposedly infected people may be oversampled in order to get meaningful results.

<sup>18</sup>This holds because  $P(x = 0, y = 0) = P(x = 0, y = 0) + P(x = 0, y = 1) = P(x = 0) = 1 - p$  and  $P(x = 1, y = 1) = P(x = 1, y = 1) + P(x = 0, y = 1) = P(y = 1)$ .

<sup>19</sup>This might not be true, if there is voluntary enrollment into the testing pool. However, for establishing the accuracy of new tests this assumptions seems to be applicable often. See Footnote 17.

<sup>20</sup>Recall that the testing pool is obtained for the new test. This assumption might be violated, if, for example, the medical staff performing the established test for the testing pool is extra careful. In this case, the established test might be more sensitive for the testing pool.



To avoid trivial cases, assume that  $\gamma, \zeta, \tau > 0$ . Note that  $\tau$  has a slightly different interpretation as in [Manski and Molinari \(2021\)](#) or [Stoye \(2021\)](#). Here,  $\tau = 1$  means the data  $P(y, z|t = 1)$  is perfectly representative of the overall population. In particular, such a parameter value implies no oversampling of infected participants.<sup>21</sup> In particular, even if the participation pool is small (as is often the case), this does not mean that  $\tau$  should be close to zero.<sup>22</sup>

With this notation, we have  $P(z = 1) = \tau\zeta$  since the new test is positive only if the person was tested. Furthermore, [Assumption 2](#) combined with [Assumption 4](#) gives  $P(x = 1|t = 1) = \gamma/\sigma$ . Then, the Law of Total Probability together with [Assumption 3](#) provides sharp bounds<sup>23</sup> on prevalence  $p \in [\tau\gamma/\sigma, \gamma/\sigma] =: [\underline{\chi}, \bar{\chi}]$  because

$$p := P(x = 1) = \underbrace{P(x = 1|t = 1)}_{=\frac{\gamma}{\sigma}} \tau + \underbrace{P(x = 1|t = 0)}_{\in [0, \frac{\gamma}{\sigma}] \text{ by Assumption 3}} (1 - \tau).$$

In turn, bounds on the established test's overall positivity rate are implied by sensitivity  $\sigma$  and [Assumption 2](#):  $P(y = 1) = p\sigma \in [\tau\gamma, \gamma]$ .

Since we consider the non-trivial case of  $p \in (0, 1)$ , consistency of the data with the maintained assumptions requires the established test's sensitivity to be sufficiency high, i.e.  $\gamma < \sigma \leq 1$ . In turn, the assumptions imply  $P(y = 1) \in (0, 1)$ .

## 2.1 THE ESTABLISHED TEST

[Assumption 2](#) implies a perfect positive predictive value for the established test ( $PPV_y = 1$ ). However, the negative predictive value is only partially identified and [Proposition 1](#) provides sharp bounds.

**Proposition 1.** *Under [Assumption 1](#)–[Assumption 4](#), the established test's negative predictive value is sharply bounded as follows:*

$$NPV_y \in \left[ \frac{1}{\sigma} \frac{\sigma - \gamma}{1 - \gamma}, \frac{1}{\sigma} \frac{\sigma - \tau\gamma}{1 - \tau\gamma} \right].$$

<sup>21</sup>See [Footnote 17](#) for why such an assumption might be problematic.

<sup>22</sup>A small participation pool might worsen the statistical inference problem: suppose the participation pool is perfectly representative but small. In this case  $\tau = 1$ , but inference usually relies on some sort of central limit theorem which would not be appropriate in this scenario. However, recall that I abstract away from inference problems as mentioned above.

<sup>23</sup>A bound for a given set is called *sharp* if the bound itself is a member of this set.

*Proof.* Fix  $\alpha = P(y = 1) \in [\tau\gamma, \gamma]$  and define prevalence as a function of  $\alpha$  by  $P_\alpha(x = 1) = \alpha/\sigma$ . Then

$$\text{NPV}_y(\alpha) = \frac{1 - P_\alpha(x = 1)}{1 - \alpha} \underbrace{P(y = 0|x = 0)}_{=1 \text{ by Assumption 2}} = \frac{1}{\sigma} \frac{\sigma - \alpha}{1 - \alpha}$$

Since  $\sigma - \alpha \leq 1 - \alpha$  for all  $\alpha \in [\tau\gamma, \gamma]$ ,  $\text{NPV}_y(\cdot)$  is decreasing. Therefore,  $\text{NPV}_y \in [\text{NPV}_y(\gamma), \text{NPV}_y(\tau\gamma)]$ . ■

With this in hand, the established test's informativeness can be analyzed. Table 2 summarizes the prevalence before and after observing a test result from the established test, which are the relevant measures for defining informativeness (cf. Figure 1). Formally, the established test is a *dilation* if every possible prevalence  $p \in [\underline{\chi}, \bar{\chi}]$  is a possible value of both  $P(x = 1|y = 1) = \text{PPV}_y$  and  $P(x = 1|y = 0) = 1 - \text{NPV}_y$ . Obviously, a positive test result gives perfect knowledge due to the maintained assumptions. Thus, the established test cannot be a dilation. On the other hand, a negative result lowers the lower and upper bound of prevalence conditional on a negative result because  $\tau\gamma \leq \gamma < \sigma$ .<sup>24</sup> Furthermore, the interval width for any test result shrinks the set of possible values for prevalence conditional on either test result.<sup>25</sup> In this sense, the established test is not just informative (i.e. not a dilation), but also strictly shrinks the size of the set of possible prevalence values after a negative test result.

Table 2: Informativeness of the established test

$P(x = 1 \cdot)$	Lower bound	Upper bound	Interval Width
Prior testing	$\tau \frac{\gamma}{\sigma}$	$\frac{\gamma}{\sigma}$	$\frac{\gamma}{\sigma} (1 - \tau)$
Positive result ( $y = 1$ )	1	1	0
Negative result ( $y = 0$ )	$\frac{\tau\gamma}{\sigma} \frac{1-\sigma}{1-\tau\gamma}$	$\frac{\gamma}{\sigma} \frac{1-\sigma}{1-\gamma}$	$\frac{\gamma}{\sigma} \frac{1-\tau}{1-\tau\gamma} \frac{1-\sigma}{1-\gamma}$

Remark: The second row corresponds to  $\text{PPV}_y$  and the third row is given by  $1 - \text{NPV}_y$ .

It is well known that knowledge of prevalence is needed in order to apply Bayes' rule to obtain NPV. Since in most applications prevalence is not known, a common practice is to assume a given prevalence level. For example the United States Food and Drug Administration (FDA, 2020b) assumes a prevalence of 5% to cal-

<sup>24</sup>Of course, this has to hold since the Law of Total Probability holds pointwise.

<sup>25</sup>It is obvious for a positive test result. For a negative result, note that the width strictly increases if and only if  $1 - \sigma > (1 - \gamma)(1 - \tau\gamma)$ , which is equivalent to  $\tau(1 + \gamma) > \frac{\sigma}{\gamma} + 1 \geq 2$  leading to a contradiction.

culate PPV and NPV. If such an assumption ( $p = \chi$ ) is added to the maintained assumptions, then  $P(y = 1) = \chi\sigma$  and furthermore  $P(y = 1|t = 0) = \frac{\chi\sigma - \gamma\tau}{1 - \tau}$ .<sup>26</sup> This additional assumption allows to exactly pin down the established test's NPV as  $\frac{1 - \chi}{1 - \chi\sigma}$  and therefore  $P(x = 1|y = 0) = \chi \frac{1 - \sigma}{1 - \chi\sigma}$ . Thus, this additional assumption not only assumes away the ambiguity about prevalence, but also illustrates that the established test does not provide ambiguous information itself.<sup>27</sup> The apparent ambiguity reflected in the non-trivial interval for values of prevalence after a negative test result (cf. Table 2) or NPV (cf. Proposition 1) is only a manifestation of the ambiguity about prevalence, but it is not due to the test itself.

## 2.2 THE NEW TEST

Next, the new test's informativeness is analyzed. First, I will discuss informativeness only based on the tested population. For this subpopulation the prevalence is given by  $\bar{\chi} = \gamma/\sigma$  and therefore the ambiguity about prevalence is muted. Subsection 2.4 extends the analysis then to the informativeness of the new test for the overall population. For the test population, the relevant measures are again positive-predictive value (PPV) and negative-predicative value (NPV), but now they are also conditional on being tested:

$$\begin{aligned} \text{PPV}_z &:= P(x = 1|z = 1, t = 1) = \frac{P(x = 1, z = 1|t = 1)}{P(z = 1|t = 1)} = \frac{P(x = 1, z = 1|t = 1)}{\zeta} \\ \text{NPV}_z &:= P(x = 0|z = 0, t = 1) = \frac{P(x = 0, z = 0|t = 1)}{P(z = 0|t = 1)} = \frac{P(x = 0, z = 0|t = 1)}{1 - \zeta}. \end{aligned}$$

To obtain these measures, the distribution  $P(x, z|t = 1)$  is needed. For a fixed  $\tau$ , I use a result from Joe (1997) that provides the set of all possible joint distributions  $P(x, y, z)$  compatible with the data  $P(x, y|t = 1)$  (cf. Appendix A). Setting  $\tau = 1$  in this construction gives the possible distributions  $P(x, y, z|t = 1)$ . Finally,  $P(x, z|t = 1)$  is obtained by marginalization.

<sup>26</sup>Alternatively, one could drop the assumption that  $P(t = 1) = \tau$  is known exactly. In this case (and allowing  $P(y = 1|t = 0) \in [0, \gamma]$  as in the general case) the assumed prevalence bounds  $\tau$ . Calculations show that  $\tau \in \left[0, \frac{\chi\sigma}{\gamma}\right]$ . Since the lower bound is always  $\tau_{\min} = 0$ , we do not find this case very interesting.

<sup>27</sup>Technically, the established test is an experiment á la Blackwell (1951), where sensitivity and specificity can be seen as functions mapping (health) states to distributions over signals (i.e. test results). As mentioned in the introduction, this implies that the established test's value of information is (weakly) positive under these assumptions.

To simplify the algebraic expressions it will be useful to differentiate between four cases defined in Table 3. Fixing the established test's sensitivity  $\sigma$ , the test data  $P(x, y|t = 1)$  immediately reveals the case the test belongs to. Figure 2 illustrates this for three real tests considered later (StQ, BiN, CT) and three hypothetical tests (including the dilation test from Table 1). When  $\sigma \rightarrow 1$ , then all but the informative case (I) cease to be relevant. For SARS-CoV-2 detecting Antigen test the WHO recommends a minimum specificity close to one. Tests close to the (top-right) frontier in Figure 2 satisfy this criterion.<sup>28</sup> Thus, for most applications, either the *confirmatory* (if  $\sigma < 1$ ) or the *informative* case (if  $\sigma \approx 1$ ) will be the relevant ones.

Table 3: Cases relating the two tests

Case name	Parameter restriction
<i>Confirmatory (C)</i>	$P(y = 0, z = 0 t = 1) \geq \max\{\bar{\chi}(1 - \sigma), 1 - \bar{\chi}\}$
<i>Informative (I)</i>	$1 - \bar{\chi} > P(y = 0, z = 0 t = 1) \geq \bar{\chi}(1 - \sigma)$
<i>Uninformative (U)</i>	$\bar{\chi}(1 - \sigma) > P(y = 0, z = 0 t = 1) \geq 1 - \bar{\chi}$
<i>Contradictory (X)</i>	$\min\{\bar{\chi}(1 - \sigma), 1 - \bar{\chi}\} > P(y = 0, z = 0 t = 1)$

Recall  $\bar{\chi} = \gamma/\sigma$  is the upper bound on prevalence and  $\gamma = P(y = 1|t = 1)$  is established test's yield.

In contrast to the established test, the new test's PPV could be less than one and is, in general, only set-identified. The reason for set-identification is that  $P(x, z|t = 1)$  is not directly observed. As explained above, there are multiple distributions  $P(x, z|t = 1)$  consistent with the data and each distribution leads to a potentially different PPV. Proposition 2 establishes the sharp identified set for values of PPV.

**Proposition 2 (PPV).** *Under Assumption 1, Assumption 2, and Assumption 4, the new test's positive predictive value is*

$$\text{PPV}_z \in \begin{cases} [P(y = 1|z = 1, t = 1), 1] & \text{in case (C)} \\ \left[ P(y = 1|z = 1, t = 1), P(y = 1|z = 1, t = 1) + \bar{\chi} \frac{1-\sigma}{\zeta} \right] & \text{in case (I)} \\ \left[ 1 - \frac{1-\bar{\chi}}{\zeta}, 1 \right] & \text{in case (U)} \\ \left[ 1 - \frac{1-\bar{\chi}}{\zeta}, P(y = 1|z = 1, t = 1) + \bar{\chi} \frac{1-\sigma}{\zeta} \right] & \text{in case (X)}. \end{cases}$$

<sup>28</sup>CT, Uni, and Anti do not satisfy the WHO minimum requirement of a 97% minimum specificity.

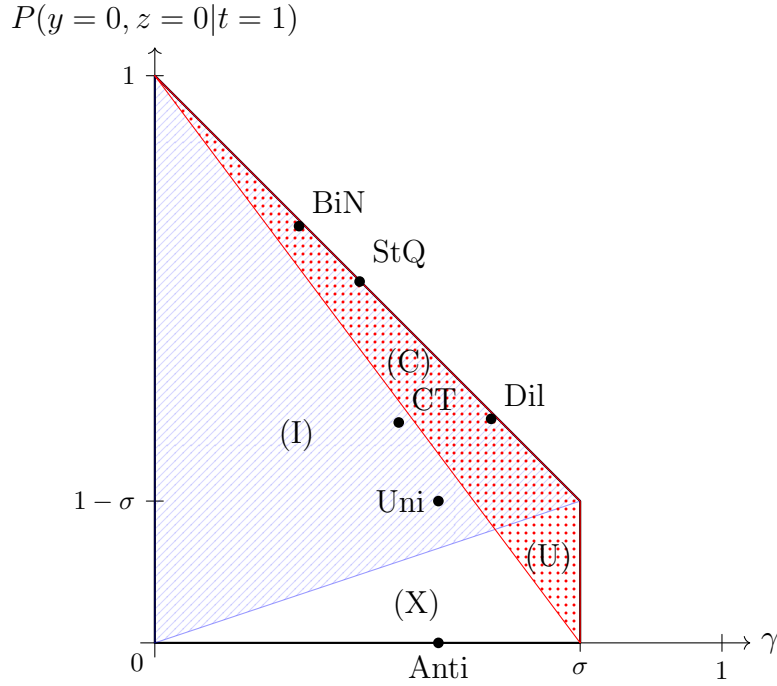


Figure 2: Illustrating the four cases with  $\sigma = 0.75$ :

*Uni* is a test corresponding to a uniform distribution  $P(y, z | t = 1) = 1/4$ . *Anti* always produces the opposite result of the established test  $P(y = 1, z = 0 | t = 1) = P(y = 0, z = 1 | t = 1) = 1/2$ . *StQ*, *BiN*, and *CT* are real tests studied later in Section 3. *Dil* is the dilation test given by Table 1.

*Proof.* Conditional on  $t = 1$  is the same as if  $\tau = 1$ . Thus, From Table 20 and Table 21, we obtain respectively:<sup>29</sup>

$$\begin{aligned} \underline{P}(x = 1, z = 1 | t = 1) &:= P(y = 1, z = 1 | t = 1) + \max \left\{ 0, \underbrace{P(y = 0, z = 1 | t = 1) - (1 - \bar{\chi})}_{=\bar{\chi} - \gamma - P(y=0, z=0|t=1)} \right\} \\ \bar{P}(x = 1, z = 1 | t = 1) &:= P(y = 1, z = 1 | t = 1) + \min \{ \bar{\chi}, 1 - P(y = 0, z = 0 | t = 1) \} - \gamma \\ &= \min \{ \bar{\chi}, 1 - P(y = 0, z = 0 | t = 1) \} - P(y = 1, z = 0 | t = 1) \\ &= \min \left\{ \underbrace{\bar{\chi} - P(y = 1, z = 0 | t = 1)}_{=\bar{\chi} - \gamma + P(y=1, z=1|t=1)}, \zeta \right\}. \end{aligned}$$

Now, note that  $\bar{\chi} - \gamma = \bar{\chi}(1 - \sigma)$  and divide by  $\zeta$  to obtain PPV. ■

<sup>29</sup>Here and in the following I will use  $\underline{P}$  to denote lower bound distributions and  $\bar{P}$  for upper bounds.

To avoid partially identified predictive values, these measures for the new tests are often reported *as if* the reference test is perfect. In this case, the data  $P(y, z)$  alone delivers a unique predictive value:

**Corollary 1** (Perfect Gold Standard - PPV). *Suppose [Assumption 1](#), [Assumption 2](#), and [Assumption 4](#) hold. If  $\sigma = 1$ , i.e. the established test has perfect sensitivity, then  $PPV_z = P(y = 1|z = 1, t = 1)$ .*

*Proof.* If  $\sigma = 1$ , then  $\bar{\chi} = \gamma = P(y = 1|t = 1)$ . Thus, the relevant case is (I).<sup>30</sup> Therefore the lower bound is  $P(y = 1|z = 1, t = 1)$  and the upper bound is  $P(y = 1|z = 1, t = 1) + \underbrace{\frac{\gamma}{\zeta} \frac{1 - \sigma}{\sigma}}_{=0}$ . ■

We saw before that the established test always achieves a maximal PPV of one and therefore provides a lot of information in case it delivers a positive result. How informative is a positive result of the new test? To answer this question, note that since we condition on being tested, there is no prior uncertainty as the prevalence in the testing pool is given by  $\bar{\chi}$ . Even without this prior ambiguity there remains ambiguity in the test result. For example, in the confirmatory case (C) the interval's width of possible values for  $PPV_z$  is  $1 - P(y = 1|z = 1, t = 1)$ , which is usually small—but non-zero—in applications. Thus, the information obtained from a new test is ambiguous at least after a positive test result.<sup>31</sup>

In contrast to the established test as discussed in [Subsection 2.1](#), the ambiguity arising from the new test allows for the occurrence of dilation. In the current setting a dilation occurs if  $\bar{\chi} = P(x = 1|t = 1)$  is contained in the intersection of the two sets with possible values for  $P(x = 1|y = i, t = 1)$  for each test result  $i \in \{0, 1\}$ . Is it possible that after a positive test result the set of possible values for  $P(x = 1|y = 1, t = 1)$  contain  $\bar{\chi} = P(x = 1|t = 1)$ ? [Corollary 2](#) provides a full characterization. The corresponding case after a negative test result will be discussed afterwards.

**Corollary 2** (Bounds Increase - PPV). *Suppose [Assumption 1](#), [Assumption 2](#), and [Assumption 4](#) hold. The new test's possible values for PPV contain  $\bar{\chi} =$*

<sup>30</sup>Technically, one could be on the border of case (C) or (U), but by continuity the resulting bounds do not change.

<sup>31</sup>This observation alone implies the test is *not* an experiment à la [Blackwell \(1951\)](#). Similarly to deriving PPV, it is possible to derive the new test's sensitivity. This will be a set in general too. Thus, there is a correspondence from (health) states to distributions over signals (test results).

$P(x = 1|t = 1)$  if and only if

$$\sigma \leq \min \left\{ \frac{\gamma(1 - \zeta)}{P(y = 1, z = 0|t = 1)}, \frac{\gamma\zeta}{P(y = 1, z = 1|t = 1)} \right\},$$

where the first entry corresponds to an increase in the upper bound, and the second condition ensures the lower bound decreases.

*Proof.* For the upper bounds, note that a strict decrease can only happen if and only if cases (I) or (X) occur (i.e.  $1 - \bar{\chi} > P(y = 0, z = 0|t = 1)$ ) and  $P(y = 1|z = 1, t = 1) + \bar{\chi}\frac{1-\sigma}{\zeta} < \bar{\chi}$ . The first is equivalent to  $\sigma > \frac{\gamma}{1-P(y=0, z=0|t=1)}$  and the second to  $\sigma > \frac{1-\zeta}{P(z=0|y=1, t=1)} = \frac{\gamma}{P(y=1|z=0, t=1)}$ . Since  $P(y = 0, z = 0|t = 1) \leq P(y = 0|z = 0, t = 1)$  we also have  $P(y = 1|z = 0, t = 1) \leq 1 - P(y = 0, z = 0|t = 1)$ . Thus, a strict decrease happens if and only if  $\sigma > \frac{\gamma}{P(y=1|z=0, t=1)}$ .

For the lower bound, note that  $1 - \frac{1-\bar{\chi}}{\zeta} \leq \bar{\chi}$  always holds. For the other cases (C and I, i.e.  $P(y = 0, z = 0|t = 1) \geq \bar{\chi}(1 - \sigma)$ ), a strict increase is  $P(y = 1|z = 1, t = 1) > \chi$ , which is equivalent to  $\sigma > \frac{\gamma}{P(y=1|z=1, t=1)}$ . Whereas the case condition is equal to  $\sigma \geq \frac{\gamma}{1-P(y=0, z=1|t=1)}$ . Similar to above,  $P(y = 1|z = 1, t = 1) \leq 1 - P(y = 0, z = 1|t = 1)$  and therefore a strict increase happens if and only if  $\sigma > \frac{\gamma}{P(y=1|z=1, t=1)}$ . ■

The inequality of [Corollary 2](#) becomes non-trivial in case the testing data does not correspond to an independent distribution, which will be the case for most applications. In these cases, a dilation cannot occur when the non-trivial inequality of [Corollary 2](#) is violated.

It remains to analyze the information contained in a negative test result. [Proposition 3](#) establishes sharp bounds for the negative predictive value of the new test. This is the relevant measure for analyzing how informative a negative test result is.

**Proposition 3 (NPV).** *Under [Assumption 1](#), [Assumption 2](#), and [Assumption 4](#), the new test's negative predictive value is*

$$\text{NPV}_z \in \begin{cases} \left[ P(y = 0|z = 0, t = 1) - \bar{\chi}\frac{1-\sigma}{1-\zeta}, \frac{1-\bar{\chi}}{1-\zeta} \right] & \text{in case (C)} \\ \left[ P(y = 0|z = 0, t = 1) - \bar{\chi}\frac{1-\sigma}{1-\zeta}, P(y = 0|z = 0, t = 1) \right] & \text{in case (I)} \\ \left[ 0, \frac{1-\bar{\chi}}{1-\zeta} \right] & \text{in case (U)} \\ [0, P(y = 0|z = 0, t = 1)] & \text{in case (X)}. \end{cases}$$

*Proof.* Conditional on  $t = 1$  is the same as if  $\tau = 1$ . Thus, From [Table 20](#) and [Table 21](#), we obtain respectively:

$$\begin{aligned}\underline{P}(x = 0, z = 0|t = 1) &= \max \{0, 1 - \bar{\chi} - P(y = 0, z = 1|t = 1)\} \\ &= \max \left\{ 0, \underbrace{\gamma - \bar{\chi}}_{=-\bar{\chi}(1-\sigma)} + P(y = 0, z = 0|t = 1) \right\} \\ \bar{P}(x = 0, z = 0|t = 1) &= \min \{1 - \bar{\chi}, P(y = 0, z = 0|t = 1)\}.\end{aligned}$$

Division by  $P(z = 0|t = 1) = 1 - \zeta$  gives NPV. ■

The uninformative (U) and contradictory (X) case seem problematic in light of [Proposition 3](#). In both cases, the lower bound is zero and also the width of the interval is rather large. This is another indication that any reasonable test should not fall in either of these two cases. However, even for the other cases—and like for PPV—the NPV is generally only set-identified. Therefore a negative test result also produces ambiguous information.

Avoiding this ambiguity can be achieved with a perfect reference test. [Corollary 3](#) verifies that if the reference test is perfect, [Proposition 3](#) reduces to the expression used in many applications and can be calculated directly from the data  $P(y, z)$ .

**Corollary 3** (Perfect Gold Standard - NPV). *Suppose [Assumption 1](#), [Assumption 2](#), and [Assumption 4](#) hold. If  $\sigma = 1$ , i.e. the established test has perfect sensitivity, then  $\text{NPV}_z = P(y = 0|z = 0, t = 1)$ .*

*Proof.* If  $\sigma = 1$ , then  $\frac{\gamma}{1-\zeta} \frac{1-\sigma}{\sigma} = 0$  and as in the proof of [Corollary 1](#) the relevant case is (I). ■

If there is no perfect reference test available, the negative new test's result leads to ambiguity. Similar to the case of a positive test result, this ambiguity allows for the occurrence of dilation. Using [Proposition 3](#), [Corollary 4](#) provides a characterization of when the set of possible values of  $P(x = 1|z = 0, t = 1) = 1 - \text{NPV}_z$  contains the prior information  $P(x = 1|t = 1) = \bar{\chi}$ .

**Corollary 4** (Bounds Increase - NPV). *Suppose [Assumption 1](#), [Assumption 2](#), and [Assumption 4](#) hold. The set of possible values for  $P(x = 1|z = 0, t = 1)$*



includes the prior information  $\bar{\chi} = P(x = 1|t = 1)$  if and only if

$$\sigma \leq \min \left\{ \frac{\gamma\zeta}{P(y = 1, z = 1|t = 1)}, \frac{\gamma(1 - \zeta)}{P(y = 1, z = 0|t = 1)} \right\},$$

where the first entry corresponds to an increase in the upper bound, and the second condition ensures the lower bound decreases.

*Proof.* For the upper bound (of  $1 - \text{NPV}_z$ ) a strict decrease can only happen in cases (C) and (I), i.e.  $P(y = 0, z = 0|t = 1) \geq \bar{\chi}(1 - \sigma)$ . In these cases, a strict decrease is equivalent to  $(1 - \zeta) - [P(y = 0, z = 0|t = 1) - \chi(1 - \sigma)] < \chi(1 - \zeta)$  or  $1 - \zeta + \bar{\chi}(\zeta - \sigma) < P(y = 0, z = 0|t = 1) = 1 - \zeta - P(y = 1, z = 0|t = 1)$ . Rearranging gives,

$$\sigma > \frac{\zeta}{P(z = 1|y = 1, t = 1)} = \frac{\gamma\zeta}{P(y = 1, z = 1|t = 1)}.$$

As in the proof of [Corollary 2](#) the condition for being in case (C) or (I) is implied by this condition.

For the lower bound, a decrease occurs if

$$\begin{aligned} \chi(1 - \zeta) &\geq \\ (1 - \zeta) - \min \{1 - \chi, P(y = 0, z = 0|t = 1)\} &= \max \{\chi - \zeta, P(y = 1, z = 0|t = 1)\}. \end{aligned}$$

First,  $\chi - \zeta \leq \chi(1 - \zeta)$  always holds as  $\chi \leq 1$ . Second, rearranging  $P(y = 1, z = 0|t = 1) \leq (1 - \zeta)\frac{\gamma}{\sigma}$  provides the condition.  $\blacksquare$

[Corollary 2](#) combined with [Corollary 4](#) provides an exact characterization for when the new test is a dilation. In fact, as the conditions are the same a dilation occurs if and only if

$$\sigma \leq \min \left\{ \frac{\gamma\zeta}{P(y = 1, z = 1|t = 1)}, \frac{\gamma(1 - \zeta)}{P(y = 1, z = 0|t = 1)} \right\}. \quad (3)$$

When evaluating a new test's accuracy it is important to make sure the data violates [Expression \(3\)](#). Otherwise, the test is uninformative in an extreme sense. In typical applications, the data often satisfies  $P(y = 1, z = 0|t = 1) \leq \gamma(1 - \zeta)$ .<sup>32</sup> In these cases, a dilation can only occur if  $\sigma \leq \frac{\gamma\zeta}{P(y=1, z=1|t=1)}$ .

<sup>32</sup>Even data that regards a test as inadequate as in [Cassaniti et al. \(2020\)](#) satisfies this inequality. I thank Filip Obradovic for making me aware of this report.

The WHO (2020) recommends minimum quality requirements using only information directly provided by the data  $P(y, z|t = 1)$ . In light of this analysis, an evaluation should also take  $\sigma$ , the established test's sensitivity, into account and with this also make sure that the test is not a dilation.  $\sigma \leq \frac{\gamma\zeta}{P(y=1, z=1|t=1)}$  combined with a given minimum standard provides an easy-to-verify sufficient condition to avoid dilation.

For this, let  $\underline{\Sigma}$  be a minimum (apparent) sensitivity threshold below which a test is deemed not reliable and denote the new test's apparent sensitivity with  $\Sigma = P(z = 1|y = 1, t = 1)$ , so that a test is reliable if  $\Sigma > \underline{\Sigma}$ .<sup>33</sup> Then, the application relevant case from Expression (3) to avoid a dilation can be expressed as  $\sigma > \zeta/\Sigma$  or equivalently as  $\Sigma > \zeta/\sigma$ . If  $\underline{\Sigma} \geq \zeta/\sigma$ , then any test meeting the minimum requirement cannot be a dilation. Thus, it suffices to make sure the new test's yield is not too high:<sup>34</sup>

$$\zeta := P(z = 1|t = 1) \leq \sigma \times \underline{\Sigma}. \quad (4)$$

If the established test is highly specific, i.e.  $\sigma \approx 1$ , then Expression (4) is satisfied unless the new test's yield is extremely high.

For SARS-CoV-2 Antigen tests the WHO recommendation is  $\underline{\Sigma} = 0.8$  and if the PCR test is not highly specific then Expression (4) might be violated. For example, the dilation test of Table 1 has  $\zeta = 0.49$  and if the PCR has sensitivity of  $\sigma = 0.6$  then not only Expression (4) is violated but the test is a dilation. More specifically, Expression (3) can be used to find the exact threshold sensitivity  $\sigma^*$  below which a given test turns into a dilation. For the dilation test this value is  $\sigma^* = 60.79\%$ .

### 2.3 INFORMATIVENESS OF ADDITIONAL TESTING

Whereas the new test produces ambiguous information, the established test is always informative. Therefore, practitioners might want to perform an additional established test depending on whether a person obtains a negative or positive result from the new test. For example, if an Antigen test is used to detect SARS-

---

<sup>33</sup>Usually, the minimum requirements include also a threshold for specificity, but this does not matter here.

<sup>34</sup>It is worth recalling that this is a sufficient condition when, additionally,  $\frac{\gamma\zeta}{P(y=1, z=1|t=1)} \leq \frac{\gamma(1-\zeta)}{P(y=1, z=0|t=1)}$  and in many applications this inequality becomes irrelevant for Expression (3) because the right-hand side is greater than one.

CoV-2 and the result is positive, a common practice is verifying the result by means of a PCR test. Since PCR tests are the reference test for evaluating the accuracy of Antigen tests, the current framework can be used to shed light on how informative this additional test is.

**Proposition 4** (Combined testing). *Under [Assumption 1](#), [Assumption 2](#), and [Assumption 4](#),*

$$P(x = 1|y = 1, z = 1, t = 1) = P(x = 1|y = 1, z = 0, t = 1) = 1,$$

$$P(x = 0|y = 0, z = 0, t = 1) \in \begin{cases} \left[1 - \frac{\bar{\chi}(1-\sigma)}{P(y=0, z=0|t=1)}, 1\right] & \text{in case (C)} \\ \left[1 - \frac{\bar{\chi}(1-\sigma)}{P(y=0, z=0|t=1)}, \frac{1-\bar{\chi}}{P(y=0, z=0|t=1)}\right] & \text{in case (I)} \\ [0, 1] & \text{in case (U)} \\ \left[0, \frac{1-\bar{\chi}}{P(y=0, z=0|t=1)}\right] & \text{in case (X),} \end{cases}$$

and

$$P(x = 0|y = 0, z = 1, t = 1) \in \begin{cases} [0, 1] & \text{in case (C)} \\ \left[0, \frac{1-\bar{\chi}}{P(y=0, z=1|t=1)}\right] & \text{in case (I)} \\ \left[1 - \frac{\bar{\chi}(1-\sigma)}{P(y=0, z=1|t=1)}, 1\right] & \text{in case (U)} \\ \left[1 - \frac{\bar{\chi}(1-\sigma)}{P(y=0, z=1|t=1)}, \frac{1-\bar{\chi}}{P(y=0, z=1|t=1)}\right] & \text{in case (X).} \end{cases}$$

*Proof.* If  $y = 1$  the the PPV has to be one independent of the new test's result because of [Assumption 2](#). For NPV, again start from [Table 20](#) and [Table 21](#) with  $\tau = 1$ . First, the case of both tests matching, i.e.  $y = 0 = z$ :

$$\begin{aligned} \underline{P}(x = 0, y = 0, z = 0|t = 1) &= \max \{0, \gamma - \bar{\chi} + P(y = 0, z = 0|t = 1)\} \\ &= \max \{0, P(y = 0, z = 0|t = 1) - \bar{\chi}(1 - \sigma)\} \end{aligned}$$

and

$$\overline{P}(x = 0, y = 0, z = 0|t = 1) = \min \{1 - \bar{\chi}, P(y = 0, z = 0|t = 1)\}.$$

Now, divide by  $P(y = 0, z = 0|t = 1)$  to obtain  $P(x = 0|y = 0, z = 0, t = 1)$ .

In case of differing test results, the relevant probability are:

$$\begin{aligned}\underline{P}(x = 0, y = 0, z = 1|t = 1) &= \min \left\{ 1 - \frac{\gamma}{\sigma}, P(y = 0, z = 1|t = 1) \right\} \\ &= \min \left\{ 1 - \frac{\gamma}{\sigma}, 1 - \gamma - P(y = 0, z = 0|t = 1) \right\} \\ &= 1 - \gamma - \max \{ \bar{\chi}(1 - \sigma), P(y = 0, z = 0|t = 1) \} \\ &\text{and}\end{aligned}$$

$$\begin{aligned}\bar{P}(x = 0, y = 0, z = 1|t = 1) &= \max \{ 1 - \bar{\chi} - P(y = 0, z = 0|t = 1), 0 \} \\ &= \max \{ P(y = 0, z = 1|t = 1) - \bar{\chi}(1 - \sigma), 0 \}.\end{aligned}$$

Note that  $\underline{P}(x = 0, y = 0, z = 1|t = 1) \geq \bar{P}(x = 0, y = 0, z = 1|t = 1)$  in this case. Division by  $P(y = 0, z = 1|t = 1)$  gives  $P(x = 0|y = 0, z = 1, t = 1)$ . ■

**Proposition 4** once more reveals that tests in the category (U) and (X) should be avoided. Even if the two test results match and are both negative, the possibility of zero (negative) predictive value cannot be ruled out. **Proposition 4** also makes clear the naming convention of the cases defined in **Table 3**. A confirmatory test (C) provides accurate information when both test produce a negative result, but is completely uninformative if and only if the new test has a positive result. An informative test (I), however, always provides some information in the sense of producing not completely trivial bounds. A contradictory test (X) provides information, but leans against the result of the established test. The uninformative test (U) provides no information at all even when both tests agree on a negative result. Of course, performing the additional test is always informative in the sense of not being a dilation. The established test does not produce false-negatives and therefore a positive result from the established test is always a perfect predictor of being infected regardless of the new test.

## 2.4 PREDICTIVE VALUES FOR THE OVERALL POPULATION

**Proposition 1** bounds the established test's NPV for the overall population, not only for the tested population. The new test, on the other hand, was analyzed for the testing pool only so far. The full characterization in **Appendix A** allows to extend the analysis of the new test to make an evaluation for the overall population. Since this involves more cumbersome notation, I only illustrate the resulting bounds for the NPV  $= P(x = 0|z = 0)$ . The analysis of PPV would proceed in a similar matter.

**Proposition 5** (Unconditional NPV). *Under [Assumption 1](#)–[Assumption 4](#), the new test's (unconditional) negative predictive value is sharply bounded by*

$$\begin{aligned}
 & \left[ \frac{1 - \bar{\chi} - \tau P(y = 0, z = 1|t = 1)}{1 - \tau\zeta}, \frac{1 - \underline{\chi}}{1 - \tau\zeta} \right] && \text{in case } (C^*) \\
 & \left[ \frac{1 - \bar{\chi} - \tau P(y = 0, z = 1|t = 1)}{1 - \tau\zeta}, 1 - \frac{P(y = 1, z = 0|t = 1)}{1 - \tau\zeta} \right] && \text{in case } (I^*) \\
 & \left[ 0, \frac{1 - \underline{\chi}}{1 - \tau\zeta} \right] && \text{in case } (U^*) \\
 & \left[ 0, 1 - \frac{P(y = 1, z = 0|t = 1)}{1 - \tau\zeta} \right] && \text{in case } (X^*),
 \end{aligned}$$

where

$$\begin{aligned}
 (C^*) \quad & \dots \quad P(y = 0, z = 0|t = 1) \geq \max \left\{ 1 - \gamma - \frac{1 - \bar{\chi}}{\tau}, 1 - \bar{\chi} \right\} \\
 (I^*) \quad & \dots \quad 1 - \bar{\chi} > P(y = 0, z = 0|t = 1) \geq 1 - \gamma - \frac{1 - \bar{\chi}}{\tau} \\
 (U^*) \quad & \dots \quad 1 - \gamma - \frac{1 - \bar{\chi}}{\tau} > P(y = 0, z = 0|t = 1) \geq 1 - \bar{\chi} \\
 (X^*) \quad & \dots \quad \min \left\{ 1 - \gamma - \frac{1 - \bar{\chi}}{\tau}, 1 - \bar{\chi} \right\} > P(y = 0, z = 0|t = 1).
 \end{aligned}$$

*Proof.* From [Table 20](#) and [Table 21](#), we obtain respectively:

$$\begin{aligned}
 \underline{P}(x = 0, z = 0) &= \max \{0, 1 - \bar{\chi} - \tau P(y = 0, z = 1|t = 1)\} \\
 \bar{P}(x = 0, z = 0) &= 1 - \tau \max \{\bar{\chi}, 1 - P(y = 0, z = 0|t = 1)\} \\
 &= 1 - \tau \max \{\bar{\chi}, \zeta + P(y = 1, z = 0|t = 1)\}.
 \end{aligned}$$

Now, the result follows from dividing by  $P(z = 0) = 1 - \tau\zeta$ . ■

If instead of predictive values the interest lies in the new test's sensitivity or specificity in the whole population another complication arises. Conditional on the testing pool, both of these measures can be derived as in [Subsection 2.2](#). For example, for sensitivity one could use the proof of [Proposition 2](#) and divide by  $P(x = 1|t = 1) = \bar{\chi}$  instead of  $P(z = 1|t = 1) = \zeta$ . The bounds for sensitivity are again determined by considering the extremes of [Table 20](#) and [Table 21](#). For the unconditional sensitivity, however, the the numerator and the denominator are both set identified because  $P(x = 1) \in [\underline{\chi}, \bar{\chi}]$ . Therefore, the lower bound might not be attained at either of the extreme distributions. This makes solving for

a closed-form expression for sensitivity intractable. Nonetheless, the bounds can easily be obtained computationally by considering a fixed  $\Gamma := P(y = 1) \in [\tau\gamma, \gamma]$  with corresponding  $p = \Gamma/\sigma$ . For this  $\Gamma$ , sharp bounds of sensitivity, say  $[L_\Gamma, H_\Gamma]$ , can be obtained by using [Table 18](#) and [Table 19](#). To find the overall bounds for sensitivity, two (non-linear) optimization problems across all values of  $\Gamma$  need to be performed to give  $[\min_\Gamma L_\Gamma, \max_\Gamma H_\Gamma]$ .

### 3 APPLICATIONS

In this section, the theoretic framework will be illustrated with several applications. First, I analyze the (hypothetical) dilation test presented in the introduction. Then, I examine two real SARS-CoV-2 detecting tests. Finally, I show that CT-scanning procedures to detect COVID-19 are prone to being dilations.

#### 3.1 A DILATION TEST

As argued before the hypothetical test data in [Table 1](#) corresponds to a dilation. Suppose the test data is derived for an Antigen test to detect SARS-CoV-2 and the reference test is a PCR test.<sup>35</sup> The test satisfies the WHO's (2020) minimum requirements with apparent sensitivity ( $\Sigma = 80.6\%$ ) and specificity (97.1%) above the specified thresholds of 80% and 97%, respectively.<sup>36</sup> For such a setting the current framework is applicable. Especially, [Assumption 2](#) seems to be warranted because a PCR test is highly specific. However, it is known that a PCR test might lack high sensitivity. [Alcoba-Florez et al. \(2020\)](#) report sensitivity for several PCR tests with point estimates ranging from  $\sigma = 60.2\%$  to  $\sigma = 97.9\%$ .<sup>37</sup> All of the 95% confidence intervals exclude perfect sensitivity,  $\sigma = 1$ .

Using the results from [Subsection 2.2](#), [Table 4](#) summarizes some key statistics for the dilation test. When the PCR sensitivity is close to one, the new (hypothetical) test produces relative accurate measurements with PPV close to one and NPV above 75%. However, if the PCR test lacks high sensitivity then we cannot be sure of the dilation test's quality. In the worst-case for PCR sensitivity ( $\sigma = 0.6$ ), the new test is indeed a dilation: Before a test result was obtained the prevalence (in

<sup>35</sup>Recall that for SARS-CoV-2 detection a PCR test is the established test used to evaluate other tests. ([Esbin et al., 2020](#))

<sup>36</sup>These numbers are calculated as if the reference test is perfect. This is similar to [Corollary 1](#) and [Corollary 3](#).

<sup>37</sup>[Alcoba-Florez et al.](#) differentiate values based on the targeted gene. The range reported here is across all genes and tests.

the testing pool) is 98.8%, after obtaining *either* dilation test's result the possible probability of being infected is at least the interval [97.7%, 100%]. In fact, potentially even more puzzling is that the lowest value after a negative test is strictly higher than after a positive result. Using [Expression \(3\)](#),  $\sigma^* = 60.8\%$  represents the cutoff PCR sensitivity below which a dilation occurs.

Table 4: Dilation Test Statistics

$\sigma$	0.6	0.85	0.98	1
$\bar{\chi} = P(x = 1 t = 1)$	98.8%	69.8%	60.5%	59.3%
PPV <sub>z</sub>	[97.6%, 100%]	[97.6%, 100%]	[97.6%, 100%]	97.6%
1 − NPV <sub>z</sub>	[97.7%, 100%]	[40.7%, 43.1%]	[22.6%, 24.9%]	22.6%
NPV <sub>z</sub>	[0%, 2.29%]	[56.9%, 59.3%]	[75.1%, 77.43%]	77.43%
Dilation Threshold	$\sigma^* = 60.8\%$			

### 3.2 STANDARD Q COVID-19 RAPID ANTIGEN TEST

Next, consider the Standard Q (*StQ*) COVID-19 Rapid Antigen Test of SD Biosensor/Roche for detection of SARS-CoV-2 as analyzed by [Kaiser et al. \(2020\)](#). They use results of PCR tests as comparison (see [Footnote 35](#)). The testing data is summarized in [Table 5](#). When  $\sigma = 1$ , then StQ's PPV and NPV are obtained

Table 5: StQ Test results from [Kaiser et al. \(2020, p. 3\)](#)

$z \backslash y$	$y = 0$	$y = 1$	Sum
$z = 0$	63.71%	3.97%	67.67%
$z = 1$	0.19%	32.14%	32.33%
Sum	63.89%	36.11%	

with [Corollary 1](#) and [Corollary 3](#) which yields 99.42% and 94.13%, respectively. These are the reported values of [Kaiser et al. \(2020\)](#). However, as explained above PCR are not perfectly sensitive.<sup>38</sup> Thus, to evaluate the StQ test the current framework is applicable.

Focusing first on the testing pool only, [Table 6](#) summarizes PPV and NPV for different values of PCR sensitivity ( $\sigma$ ) using [Proposition 2](#) and [Proposition 3](#). Even if the PCR test lacks high sensitivity, StQ has a close to perfect positive

<sup>38</sup>[Kaiser et al. \(2020\)](#) use PCR tests targeting *E* genes, which tend to have higher sensitivity in the analysis of [Alcoba-Florez et al. \(2020\)](#). The lowest reported sensitivity for a PCR test targeting *E* genes is 65.33%.

predicative value ( $\text{PPV}_z \approx 1$ ). However, the values for NPV drop significantly as  $\sigma$  decreases. In the worst case, a negative StQ result becomes close to a fair coin flip. However, the test is very informative overall as can be seen by the low dilation threshold  $\sigma^* = 36.3\%$ .

Table 6: Accuracy of StQ

$\sigma$	0.6	0.85	0.98
$\text{PPV}_z$	[99.4%, 100%]	[99.4%, 100%]	[99.4%, 100%]
$\text{NPV}_z$	[58.6%, 58.9%]	[84.7%, 85%]	[93.1%, 94.1%]
Dilation Threshold	$\sigma^* = 36.3\%$		

Kaiser et al. (2020, p. 1) state “study participants were representative of the usual population seeking testing in our center (main testing center in Geneva). The majority were presenting with symptoms compatible with a SARS-CoV2 infection and a minority were asymptomatic but with a known positive contact or were asymptomatic healthcare workers.” The current framework allows to use the obtained testing data to evaluate StQ’s quality for the overall population (of Geneva) as analyzed in Subsection 2.4. Furthermore, this explanation supports Assumption 3.

Table 7 shows bounds on prevalence using the baseline analysis in Section 2. For low values of  $\tau$ , i.e. the testing pool was highly non-representative of the overall population, the width of the intervals is rather wide. However, even the lowest number is close to 2% indicating a thorough spread of the virus in Geneva at the time of testing.<sup>39</sup> When testing becomes representative ( $\tau \rightarrow 1$ ) the prevalence converges to the prevalence in the testing pool.

Table 7: Bounds on prevalence  $p \in [\underline{\chi}, \bar{\chi}]$ 

$\tau \setminus \sigma$	0.6	0.85	0.98
$1/20$	[3.01%, 60.2%]	[2.12%, 42.5%]	[1.84%, 36.8%]
$1/10$	[6.02%, 60.2%]	[4.25%, 42.5%]	[3.68%, 36.8%]
$1/2$	[30.1%, 60.2%]	[21.2%, 42.5%]	[18.4%, 36.8%]
$19/20$	[57.2%, 60.2%]	[40.4%, 42.5%]	[35.7%, 36.8%]
1	60.2%	42.5%	36.8%

<sup>39</sup>Note that this is a one time analysis. It does not answer the question of how many people were cumulatively infected by SARS-CoV-2 up to the time of testing.



At this time, if a Genevese obtains a negative PCR result, what is the probability of her being infected? If testing is not competently representative, a unique number cannot be given. However, [Proposition 1](#) provides sharp bounds for this case and the results are shown in [Table 8](#). How do these PCR results compare to

Table 8: Bounds on  $NPV_y$ 

$\tau \setminus \sigma$	0.6	0.85	0.98
$1/20$	[62.3%, 98.8%]	[90.0%, 99.7%]	[98.9%, 100%]
$1/10$	[62.3%, 97.5%]	[90.0%, 99.3%]	[98.9%, 99.9%]
$1/2$	[62.3%, 85.3%]	[90.0%, 96.1%]	[98.9%, 99.6%]
$19/20$	[62.3%, 65.2%]	[90.0%, 90.8%]	[98.9%, 98.9%]
1	62.3%	90.0%	98.9%

results from StQ? [Table 9](#) provides the numbers using [Proposition 5](#). The lower bounds are significantly lower than for the PCR test. This makes the width of the interval also significantly wider. The widening is a reflection of the combination of the two missing data problems inherit in the testing procedure without a perfect reference test: (i) unknown overall prevalence (which also affects PCR's NPV) and (ii) missing correlation data (which does not affect the PCR's NPV).

Table 9: Bounds on STQ's  $NPV_z$  for overall population

$\tau \setminus \sigma$	0.6	0.85	0.98
$1/20$	[40.5%, 96.0%]	[58.5%, 96.0%]	[64.2%, 96.0%]
$1/10$	[41.1%, 95.9%]	[59.4%, 95.9%]	[65.3%, 95.9%]
$1/2$	[47.4%, 83.4%]	[68.5%, 94.0%]	[75.2%, 95.2%]
$19/20$	[57.2%, 61.8%]	[82.8%, 86.1%]	[90.9%, 93.8%]

### 3.3 BIAXNOW COVID-19 AG HOME TEST

The BiAxNOW (*BiN*) Covid-19 Ag Home Test of Abbott is one of the first rapid Antigen tests for use at home which is able to detect the SARS-CoV-2 virus that was emergency approved by the [FDA \(2020a\)](#). BiN's clinical performance for approval by the FDA was conducted with a PCR test as a reference. The data are shown in [Table 10](#) and [Table 11](#) shows the implied accuracy measures.

Relative to StQ, BiN has significantly lower PPV and also rules out perfect PPV for lower values of PCR specificity. On the other hand, NPV is uniformly

Table 10: BiN Test results from FDA (2020a, p. 20)

$z \backslash y$	$y = 0$	$y = 1$	Sum
$z = 0$	73.48%	3.91%	77.39%
$z = 1$	1.09%	21.52%	22.61%
Sum	74.57%	25.43%	

Table 11: Accuracy of BiN

$\sigma$	0.6	0.85	0.98
PPV <sub><math>z</math></sub>	[95.2%, 100%]	[95.2%, 100%]	[95.2%, 97.5%]
NPV <sub><math>z</math></sub>	[73.0%, 74.4%]	[89.1%, 90.6%]	[94.3%, 94.9%]
Dilation Threshold	$\sigma^* = 26.7\%$		

greater for BiN compared to StQ. Even for the worst-case PCR sensitivity, BiN’s possible NPV values are reasonably high. Furthermore, the dilation threshold is extremely low at  $\sigma^* = 26.7\%$ .

The FDA (2020a) also provides additional data about BiN results by including the cycle threshold obtained by the PCR test.<sup>40</sup> Table 12 shows this data. This

Table 12: BiN Test results from FDA (2020a, p. 21) with Cycle Threshold Count

$z \backslash y$	$Ct < 33$	$Ct \geq 33$	Sum
$z = 0$	16.2%	6.94%	23.12%
$z = 1$	9.83%	67.05%	76.89%
Sum	26.01%	73.99%	

is additional data a PCR test produces, which can be used to refine bounds on predictive values of the Antigen test. However, an extension of the current setting is needed, because such additional information is not accounted for in the current setting with binary tests. Subsection 4.2 discusses a possible extension of the current setting to allow for this additional data.

<sup>40</sup>Public Health England (2020) explains: “Cycle threshold (Ct) is a semi-quantitative value that can broadly categorise the concentration of viral genetic material in a patient sample following testing by RT PCR as low, medium or high — that is, it tells us approximately how much viral genetic material is in the sample. A low Ct indicates a high concentration of viral genetic material, which is typically associated with high risk of infectivity. A high Ct indicates a low concentration of viral genetic material which is typically associated with a lower risk of infectivity.”

## 3.4 CT SCAN TO DETECT COVID-19

Ai et al. (2020) and Gietema et al. (2020) propose using chest CT scans for early identifying COVID-19 in patients. In Gietema et al.’s study, all COVID-19 symptomatic patients of a single Dutch emergency department have a chest CT scan and a PCR test for detecting SARS-CoV-2. Their study design exactly fits the framework of the current paper: (i) non-representative sampling of the testing pool and (ii) missing correlation data due to use of an imperfect reference test (with perfect specificity). The testing data are reproduced in Table 13. Compared

Table 13: CT scan data from Gietema et al. (2020, Table 2)

$z \backslash y$	$y = 0$	$y = 1$	Sum
$z = 0$	38.86%	4.66%	43.52%
$z = 1$	18.13%	38.34%	56.48%
Sum	56.99%	43.01%	

to the previously studies Antigen tests, the data for CT scans seems less aligned with the PCR test results. This is an indication that such a CT test is less informative: the dilation threshold of  $\sigma^* = 63.35\%$  is higher than for the Antigen tests. Thus, this testing procedure is completely uninformative if  $\sigma = 60\%$ —the lowest sensitivity of a PCR test reported by Alcoba-Florez et al. (2020). In this case, Table 14 shows (sharp bounds on) population prevalence, PCR NPVs, and CT scan NPVs. The PCR’s (assumed) low sensitivity means that its NPV might

Table 14: Implications for overall population from Gietema et al. (2020).

$\tau \downarrow$	$p := P(x = 1)$	$\text{NPV}_y$	$\text{NPV}_z$
$1/20$	[3.58%, 71.7%]	[49.7%, 98.5%]	[23.4%, 65.1%]
$1/10$	[7.17%, 71.7%]	[49.7%, 97.0%]	[23.4%, 65.1%]
$1/2$	[35.8%, 71.7%]	[49.7%, 81.7%]	[23.4%, 65.1%]
$19/20$	[68.1%, 71.7%]	[49.7%, 54.0%]	[23.4%, 65.1%]

be quite low, but it is at least close to 50%, irrespective of  $\tau$ . On the other hand, the CT scan has both a sizable interval and a low lower bound of possible NPVs. Since  $\sigma = 0.6$  is below the dilation threshold, the CT scan is completely uninformative for the tested population (equivalently for  $\tau = 1$ ). Furthermore, this remains true for the overall population if  $\tau \geq 1/2$  as shown in Table 14. For exam-

ple, for a non-COVID-indicative CT scan the set of possible infection probabilities  $P(x = 1|z = 0)$  increases relative to the prior information  $p$ .<sup>41</sup>

Even more striking is the data of [Ai et al. \(2020\)](#) shown in [Table 15](#).<sup>42</sup> [Ai et al.](#) also use Chest CT scans to test for COVID-19. Their data is obtained in Wuhan, China and like the study of [Gietema et al. \(2020\)](#) a PCR test is used as a reference. The data reveals a very low apparent specificity but a high apparent sensitivity of  $\Sigma = 96.51\%$ . [Expression \(4\)](#) indicates that a high yield of the new test,  $\zeta$ , might be problematic. Here, this yield is very high with  $\zeta = 87.57\%$ . The problem becomes even more apparent by looking at the dilation threshold, which is high with a value of  $\sigma^* = 90.74$ . This implies that even if the PCR test is quite sensitive, the CT scan is completely uninformative for the tested people in Wuhan.

Table 15: CT scan data from [Ai et al. \(2020, Table 2\)](#)

$z \backslash y$	$y = 0$	$y = 1$	Sum
$z = 0$	10.36%	2.07%	12.43%
$z = 1$	30.37%	57.20%	87.57%
Sum	40.73%	59.27%	

## 4 DISCUSSION AND EXTENSION

### 4.1 IMPERFECT SPECIFICITY OF THE ESTABLISHED TEST AND THE USE OF ADDITIONAL DATA

[Assumption 2](#) might be too strong for some applications. Although, this assumption simplifies the algebraic expression sometimes significantly, it is not a crucial assumption conceptually. The crucial characterization of joint distributions in [Appendix A](#) can easily be extended to allow for false-positives of the established test.

Evaluations of a new test sometimes have more data available than just  $P(y, z|t = 1)$ . For example, blood samples from before the existence of a virus can serve as true-negative samples. On the other hand, specific blood samples could be analyzed with more sophisticated (and usually much more expensive) methods than just using an established test as reference. These methods would lead to samples

<sup>41</sup>Recall that  $P(x = 1|z = 0) = 1 - \text{NPV}_z$ .

<sup>42</sup>I thank Filip Obradovic for providing this reference.

with true positives (or at least with very high probability).<sup>43</sup> Either of these methods would be provide additional data and therefore would also reduce the missing data problem. In general, this supplementary knowledge leads to narrower bounds, but unless these extra methods are performed for the whole tested population, the missing correlation issues remains. Of course, these methods cannot be applied for the untested population. Therefore the the missing data on prevalence cannot be avoided with these extraneous data.

## 4.2 BEYOND BINARY TESTS

The current framework only allows for binary outcomes for both tests and also for the underlying health state. This seems to be the most common situation studied in the literature on diagnostic testing. (Zhou et al., 2014) Often tests provide ternary results (with the additional result of ‘inconclusive’ or ‘invalid’), or allow for even more detailed information, like the Cycle Threshold Count of a PCR test as mentioned in Subsection 3.3. In such situations, the theoretic analysis does not provide the appropriate machinery. However, the crucial application to characterize the set of all joint distribution is a result in copula theory (Joe, 1997, Theorem 3.10), which does not rely on any dimension being binary. Indeed, the result even works for continuous outcomes on each dimension.

Similarly, one could use other results in Joe (1997) to characterize the set of possible joint distributions if multiple tests are conducted simultaneously as studied in Zhou et al. (2014, Chapter 9). In this case, and like in the characterization of Appendix A, the testing data are higher-dimensional marginal distribution of an overall joint distribution with an additional dimension (the health state). When considering such an extension, a caution has to be taken because sometimes sharp bounds on the set of possible higher-dimensional distributions may not be known.

## 4.3 COVID-19 RELATED TESTING

Since testing has a potentially big impact on the economy, an accurate description of the available testing technology is crucial. From the microeconomic perspective, the testing technology affects how test should be optimally allocated (see Ely et al. (2021), Lipnowski and Ravid (2021)) and also how much people engage in social distancing as studied by Acemoglu et al. (2020). But also the macroeconomy

---

<sup>43</sup>For example, Olbrich et al. (2021) combine these methods to evaluate SARS-CoV-2 antibody tests.

is highly affected by testing strategies and an optimal choice might reduce the economic costs of pandemics considerably. ([Alvarez et al., 2021](#); [Eichenbaum et al., 2022](#)) Although, these studies establish the importance of testing and also address varying testing technologies, all of them assume that a test corresponds to an experiment à la [Blackwell \(1951\)](#) and therefore is always informative (sometimes the assumption is even that the test itself provides perfect information).

This paper demonstrates that the assumption of unambiguous information in test results is only applicable if a perfect reference is available when evaluating new tests. In particular, new Antigen test for detection of SARS-CoV-2 are evaluated relative to an imperfect PCR test and therefore—as shown in this paper—these Antigen test produce ambiguous information. An optimal testing procedure should take this ambiguity into account. Similarly, practitioner guides (like [Galeotti et al. \(2020\)](#); [Watson et al. \(2020\)](#)) might want to consider addressing uncertainty in test results in more detail.

## A CHARACTERIZATION OF THE SET OF JOINT DISTRIBUTIONS

$$P(x, y, z)$$

Consider a fixed  $\Gamma := P(y = 1) \in [\tau\gamma, \gamma]$  with corresponding prevalence  $p := P(x = 1) = \Gamma/\sigma$ .

**Lemma 1.** *Suppose [Assumption 1](#) and [Assumption 2](#) hold. For a given  $\Gamma$ ,  $P(x, y)$  is given by [Table 16](#).*

Table 16: Joint distribution of  $P(x, y)$

$P(x \setminus y)$	$y = 0$	$y = 1$
$x = 0$	$1 - \Gamma/\sigma$	$0$
$x = 1$	$\Gamma \frac{1-\sigma}{\sigma}$	$\Gamma$
	$1 - \Gamma$	$\Gamma$

*Proof.* 1.  $P(x = 1, y = 1) = P(y = 1) - \underbrace{P(x = 0, y = 1)}_{=0 \text{ by [Assumption 2](#)}} = \Gamma$

2.  $P(x = 1, y = 0) = P(x = 1) - P(x = 1, y = 1) = \Gamma(1 - \sigma)/\sigma$ .

3.  $P(x = 0, y = 0) = P(y = 0) - P(x = 1, y = 0) = 1 - \Gamma/\sigma$ . ■

Using the law of total probability and rearranging gives  $P(y = 1|t = 0) = \frac{\Gamma - \tau\gamma}{1 - \tau}$ . Furthermore, by the nature of testing  $P(y = 0, z = 1|t = 0) = P(y = 1, z = 1|t = 0) = 0$ . Therefore,  $P(y, z)$  is given by [Table 17](#).

Table 17: Joint distribution of  $P(y, z)$

$P(z \setminus y)$	$y = 0$	$y = 1$
$z = 0$	$1 - \Gamma - P(y = 0, z = 1 t = 1)\tau$	$\Gamma - P(y = 1, z = 1 t = 1)\tau$
$z = 1$	$P(y = 0, z = 1 t = 1)\tau$	$P(y = 1, z = 1 t = 1)\tau$
	$1 - \Gamma$	$\Gamma$

By [Joe \(1997, Theorem 3.10\)](#) the set of all distributions  $P(x, y, z)$  with marginals given by  $P(x, y)$  and  $P(y, z)$  is bounded by two extreme distributions:<sup>44</sup>  $\underline{F}_\Gamma \leq F \leq \overline{F}_\Gamma$ , where  $F$  is the CDF corresponding to  $P(x, y, z)$ ,  $\underline{F}_\Gamma$  is given by [Table 18](#), and  $\overline{F}_\Gamma$  is given by [Table 19](#).

Table 18: CDF  $\underline{F}_\Gamma$ 

		$x = 1$		$x = 0$	
		$z = 1$	$z = 0$	$z = 1$	$z = 0$
$y = 1$	1		$1 - \tau\zeta$	$1 - \frac{\Gamma}{\sigma}$	$\max\{0, 1 - \Gamma/\sigma - P(y = 0, z = 1 t = 1)\tau\}$
$y = 0$	$1 - \Gamma$	$1 - \Gamma - P(y = 0, z = 1 t = 1)\tau$		$1 - \frac{\Gamma}{\sigma}$	$\max\{0, 1 - \Gamma/\sigma - P(y = 0, z = 1 t = 1)\tau\}$

Table 19: CDF  $\overline{F}_\Gamma$ 

		$x = 1$		$x = 0$	
		$z = 1$	$z = 0$	$z = 1$	$z = 0$
$y = 1$	1		$1 - \tau\zeta$	$1 - \frac{\Gamma}{\sigma}$	$\min\{1 - \frac{\Gamma}{\sigma}, 1 - \Gamma - P(y = 0, z = 1 t = 1)\tau\}$
$y = 0$	$1 - \Gamma$	$1 - \Gamma - P(y = 0, z = 1 t = 1)\tau$		$1 - \frac{\Gamma}{\sigma}$	$\min\{1 - \frac{\Gamma}{\sigma}, 1 - \Gamma - P(y = 0, z = 1 t = 1)\tau\}$

Since  $\underline{F}_\Gamma$  and  $\overline{F}_\Gamma$  are both nonincreasing in  $\Gamma$ , sharp bounds for the CDF  $F$  across all  $\Gamma := P(y = 1) \in [\tau\gamma, \gamma]$  are  $\underline{F} := \underline{F}_\gamma \leq F \leq \overline{F}_{\tau\gamma} =: \overline{F}$ . For the lower, we have  $1 - \Gamma/\sigma - P(y = 0, z = 1|t = 1) = (1 - \gamma)(1 - \tau) + \tau$ .

For the upper bound, note that  $1 - \tau\gamma - P(y = 0, z = 1|t = 1)\tau = 1 - \tau(1 - P(y = 0, z = 0|t = 1))$ . The corresponding probability mass functions are given by Table 20 and Table 21.

Table 20: Lower bound PMF with  $\mathcal{P}_{01} := P(y = 0, z = 1|t = 1)$ 

		$x = 1$		$x = 0$	
		$z = 1$	$z = 0$	$z = 1$	$z = 0$
$y = 1$	$P(y = 1, z = 1 t = 1)\tau$		$\gamma - P(y = 1, z = 1 t = 1)\tau$	0	0
$y = 0$	$\max\{0, \frac{\gamma}{\sigma} + \mathcal{P}_{01}\tau - 1\}$		$\min\{\gamma\frac{1-\sigma}{\sigma}, 1 - \gamma - \mathcal{P}_{01}\tau\}$	$\min\{1 - \frac{\gamma}{\sigma}, \mathcal{P}_{01}\tau\}$	$\max\{0, 1 - \frac{\gamma}{\sigma} - \mathcal{P}_{01}\tau\}$

Table 21: Upper bound PMF with  $\mathcal{P}_{00} := P(y = 0, z = 0|t = 1)$ 

		$x = 1$		$x = 0$	
		$z = 1$	$z = 0$	$z = 1$	$z = 0$
$y = 1$	$P(y = 1, z = 1 t = 1)\tau$		$P(y = 1, z = 0 t = 1)\tau$	0	0
$y = 0$	$\tau[\min\{\frac{\gamma}{\sigma}, 1 - \mathcal{P}_{00}\} - \gamma]$		$\tau\max\{\mathcal{P}_{00} + \frac{\gamma}{\sigma} - 1, 0\}$	$\tau\max\{1 - \frac{\gamma}{\sigma} - \mathcal{P}_{00}, 0\}$	$1 - \tau\max\{\frac{\gamma}{\sigma}, 1 - \mathcal{P}_{00}\}$

## REFERENCES

ACEMOGLU, D., A. MAKHDOUNI, A. MALEKIAN, AND A. OZDAGLAR (2020): “Testing, Voluntary Social Distancing and the Spread of an Infection,” Tech. Rep. w27483, National Bureau of Economic Research, <https://www.nber.org/papers/w27483>.

<sup>44</sup>The set of all these distributions, often called Fréchet class, is a convex set. Hence, here it suffices to consider the extreme points only.



- AI, T., Z. YANG, H. HOU, C. ZHAN, C. CHEN, W. LV, Q. TAO, Z. SUN, AND L. XIA (2020): “Correlation of Chest CT and RT-PCR Testing for Coronavirus Disease 2019 (COVID-19) in China: A Report of 1014 Cases,” *Radiology*, 296, E32–E40, <https://pubs.rsna.org/doi/10.1148/radiol.2020200642>.
- ALCOBA-FLOREZ, J., H. GIL-CAMPESINO, D. G.-M. DE ARTOLA, R. GONZÁLEZ-MONTELONGO, A. VALENZUELA-FERNÁNDEZ, L. CIUFREDA, AND C. FLORES (2020): “Sensitivity of Different RT-qPCR Solutions for SARS-CoV-2 Detection,” *International Journal of Infectious Diseases*, 99, 190–192, <http://www.sciencedirect.com/science/article/pii/S1201971220306032>.
- ALVAREZ, F., D. ARGENTE, AND F. LIPPI (2021): “A Simple Planning Problem for COVID-19 Lock-down, Testing, and Tracing,” *American Economic Review: Insights*, 3, 367–82.
- BLACKWELL, D. (1951): “Comparison of Experiments,” in *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, The Regents of the University of California, <https://projecteuclid.org/euclid.bsmmsp/1200500222>.
- (1953): “Equivalent Comparisons of Experiments,” *Annals of Mathematical Statistics*, 24, 265–272, <https://projecteuclid.org/euclid.aoms/1177729032>.
- CASSANITI, I., F. NOVAZZI, F. GIARDINA, F. SALINARO, M. SACHS, S. PERLINI, R. BRUNO, F. MOJOLI, AND F. BALDANTI (2020): “Performance of VivaDiag COVID-19 IgM/IgG Rapid Test Is Inadequate for Diagnosis of COVID-19 in Acute Patients Referring to Emergency Room Department,” *Journal of Medical Virology*, 92, 1724–1727, <https://onlinelibrary.wiley.com/doi/abs/10.1002/jmv.25800>.
- DE OLIVEIRA, H. (2018): “Blackwell’s Informativeness Theorem Using Diagrams,” *Games and Economic Behavior*, 109, 126–131, <http://www.sciencedirect.com/science/article/pii/S0899825617302270>.
- EICHENBAUM, M. S., S. REBELO, AND M. TRABANDT (2022): “The macroeconomics of testing and quarantining,” *Journal of Economic Dynamics and Control*, 138, 104337.

- ELY, J., A. GALEOTTI, O. JANN, AND J. STEINER (2021): “Optimal test allocation,” *Journal of Economic Theory*, 193, 105236.
- ESBIN, M. N., O. N. WHITNEY, S. CHONG, A. MAURER, X. DARZACQ, AND R. TJIAN (2020): “Overcoming the Bottleneck to Widespread Testing: A Rapid Review of Nucleic Acid Testing Approaches for COVID-19 Detection,” *RNA*, 26, 771–783, <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7297120/>.
- FDA (2020a): “BinaxNOW COVID-19 Ag Card Home Test: Healthcare Provider Instructions for Use,” *FDA*, <https://www.fda.gov/media/144574/download>.
- (2020b): “EUA Authorized Serology Test Performance,” *FDA*, <https://www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/eua-authorized-serology-test-performance>.
- GALEOTTI, A., J. STEINER, AND P. SURICO (2020): “Merit of Test: Perspective of Information Economics,” *Health Policy and Technology*, 9, 575–577, <http://www.sciencedirect.com/science/article/pii/S2211883720300848>.
- GIETEMA, H. A., N. ZELIS, J. M. NOBEL, L. J. G. LAMBRIKS, L. B. VAN ALPHEN, A. M. L. O. LASHOF, J. E. WILDBERGER, I. C. NELISSEN, AND P. M. STASSEN (2020): “CT in Relation to RT-PCR in Diagnosing COVID-19 in The Netherlands: A Prospective Study,” *PLOS ONE*, 15, e0235844, <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0235844>.
- GUL, F. AND W. PESENDORFER (2018): “Evaluating Ambiguous Random Variables and Updating by Proxy,” <https://www.princeton.edu/~fgul/updating.pdf>.
- JOE, H. (1997): *Multivariate Models and Multivariate Dependence Concepts*, CRC Press.
- KAISER, L., I. ECKERLE, M. SCHIBLER, AND A. BERGER (2020): “Validation Report: SARS-CoV-2 Antigen Rapid Diagnostic Test,” Tech. rep., [https://www.hug.ch/sites/interhug/files/structures/laboratoire\\_de\\_virologie/documents/Centre\\_maladies\\_virales\\_infectieuses/ofsp\\_rdt\\_report\\_gcevd\\_27.10.2020.pdf](https://www.hug.ch/sites/interhug/files/structures/laboratoire_de_virologie/documents/Centre_maladies_virales_infectieuses/ofsp_rdt_report_gcevd_27.10.2020.pdf).

- LIPNOWSKI, E. AND D. RAVID (2021): “Pooled testing for quarantine decisions,” *Journal of Economic Theory*, 198, 105372.
- MACHINA, M. J. AND M. SINISCALCHI (2014): “Chapter 13 - Ambiguity and Ambiguity Aversion,” in *Handbook of the Economics of Risk and Uncertainty*, ed. by M. Machina and K. Viscusi, North-Holland, vol. 1 of *Handbook of the Economics of Risk and Uncertainty*, 729–807, <http://www.sciencedirect.com/science/article/pii/B9780444536853000131>.
- MANSKI, C. F. (1989): “Anatomy of the Selection Problem,” *The Journal of Human Resources*, 24, 343–360, <https://www.jstor.org/stable/145818>.
- (2018): “Credible Ecological Inference for Medical Decisions with Personalized Risk Assessment,” *Quantitative Economics*, 9, 541–569, <https://onlinelibrary.wiley.com/doi/abs/10.3982/QE778>.
- (2020): “Bounding the accuracy of diagnostic tests, with application to COVID-19 antibody tests,” *Epidemiology*, 32, 162–167.
- MANSKI, C. F. AND F. MOLINARI (2021): “Estimating the COVID-19 Infection Rate: Anatomy of an Inference Problem,” *Journal of Econometrics*, 220, 181–192, <http://www.sciencedirect.com/science/article/pii/S0304407620301676>.
- OLBRICH, L., N. CASTELLETTI, Y. SCHÄLTE, M. GARÍ, P. PÜTZ, A. BAKULI, M. PRITSCH, I. KROIDL, E. SAATHOFF, J. M. GUGGENBUEHL NOLLER, V. FINGERLE, R. LE GLEU, L. GILBERG, I. BRAND, P. FALK, A. MARKGRAF, F. DEÁK, F. RIESS, M. DIEFENBACH, T. ESER, F. WEINAUER, S. MARTIN, E.-M. QUENZEL, M. BECKER, J. DURNER, P. GIRL, K. MÜLLER, K. RADON, C. FUCHS, R. WÖLFEL, J. HASENAUER, M. HOELSCHER, AND A. WIESER (2021): “Head-to-head evaluation of seven different seroassays including direct viral neutralisation in a representative cohort for SARS-CoV-2,” *Journal of General Virology*, 102.
- PUBLIC HEALTH ENGLAND (2020): “Understanding Cycle Threshold (Ct) in SARS-CoV-2 RT-PCR,” Tech. rep., Department of Health and Social Care, [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/926410/Understanding\\_Cycle\\_Threshold\\_\\_Ct\\_\\_in\\_SARS-CoV-2\\_RT-PCR\\_.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/926410/Understanding_Cycle_Threshold__Ct__in_SARS-CoV-2_RT-PCR_.pdf).

REITSMA, J. B., A. W. S. RUTJES, K. S. KHAN, A. COOMARASAMY, AND P. M. BOSSUYT (2009): “A Review of Solutions for Diagnostic Accuracy Studies with an Imperfect or Missing Reference Standard,” *Journal of Clinical Epidemiology*, 62, 797–806, <http://www.sciencedirect.com/science/article/pii/S0895435609000638>.

SACKS, D. W., N. MENACHEMI, P. EMBI, AND C. WING (forthcoming): “What can we learn about SARS-CoV-2 prevalence from testing and hospital data?” *Review of Economics and Statistics*, <https://dansacks.files.wordpress.com/2022/01/sacks-menachemi-embi-wing-2022-covid-prevalence-testing-hospitalization.pdf>.

SEIDENFELD, T. AND L. WASSERMAN (1993): “Dilation for Sets of Probabilities,” *Annals of Statistics*, 21, 1139–1154, <https://projecteuclid.org/euclid.aos/1176349254>.

SHISHKIN, D. AND P. ORTOLEVA (2021): “Ambiguous Information and Dilation: An Experiment,” [http://denisshishkin.com/papers/ambiguous\\_information.pdf](http://denisshishkin.com/papers/ambiguous_information.pdf).

STOYE, J. (2021): “Bounding infection prevalence by bounding selectivity and accuracy of tests: with application to early COVID-19,” *The Econometrics Journal*, 25, 1–14.

WATSON, J., P. F. WHITING, AND J. E. BRUSH (2020): “Interpreting a Covid-19 Test Result,” *BMJ*, 369, m1808, <https://www.bmj.com/content/369/bmj.m1808>.

WHO (2020): “Antigen-Detection in the Diagnosis of SARS-CoV-2 Infection Using Rapid Immunoassays,” <https://www.who.int/publications-detail-redirect/antigen-detection-in-the-diagnosis-of-sars-cov-2infection-using-rapid-immunoassays>

ZHOU, X.-H., N. A. OBUCHOWSKI, AND D. K. MCCLISH (2014): *Statistical Methods in Diagnostic Medicine*, John Wiley & Sons.