

1 *Interpreting Wastewater SARS-CoV-2 Results using Bayesian*

2 *Analysis*

3

4 Kyle Curtis*, Raul A. Gonzalez

5

6 Hampton Roads Sanitation District, 1434 Air Rail Avenue, Virginia Beach, VA 23455

7

8 Keywords

9 SARS-CoV-2

10 COVID-19

11 Bayesian Analysis

12 Wastewater Surveillance

13 RT-PCR

14

15

16

17

18 **Abstract**

19 Wastewater surveillance of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has proven a
20 practical complement to clinical data for assessing community-scale infection trends. Clinical assays,
21 such as the CDC-promulgated N1, N2, and N3 have been used to detect and quantify viral RNA in
22 wastewater but, to date, have not included estimates of reliability of true positive or true negative.
23 Bayes' Theorem was applied to estimate Type I and Type II error rates for detections of the virus in
24 wastewater. Conditional probabilities of true positive or true negative were investigated when one
25 assay was used, or multiple assays were run concurrently. Cumulative probability analysis was used to
26 assess the likelihood of true SARS-CoV-2 detection using multiple samples. Results demonstrate highly
27 reliable positive (>0.86 for priors >0.25) and negative (>0.80 for priors = 0.50) results using a single
28 assay. Using N1 and N2 concurrently caused greater reliability (>0.99 for priors <0.05) when results
29 concurred but generated potentially counterintuitive interpretations when results were discordant.
30 Regional wastewater surveillance data was investigated as a means of setting prior probabilities.
31 Probability of true detection with a single marker was investigated using cumulative probability across
32 all combinations of positive and negative results for a set of three samples. Findings using a low (0.11)
33 and uniformed (0.50) initial prior resulted in high probabilities of detection (>0.95) even when a set of
34 samples included one or two negative results, demonstrating the influence of high sensitivity and
35 specificity values. Analyses presented here provide a practical framework for understanding analytical
36 results generated by wastewater surveillance programs.

37

38 **Introduction**

39 The on-going coronavirus disease-2019 (COVID-19) pandemic has reached most regions of the
40 world, with global cases and deaths currently greater than 185 million and 4 million, respectively (Dong

41 et al., 2020). COVID-19 wastewater surveillance has emerged as a complementary epidemiological tool
42 used to identify community level infection trends, as an early warning system, and even to screen for
43 infected individuals at the building level. The dominant analytical techniques to quantify severe acute
44 respiratory syndrome 2 (SARS-CoV-2), the virus that causes COVID-19, are PCR-based (e.g., RT-qPCR, RT-
45 dPCR) and use clinical assays. Globally, the research community (Bivins et al., 2020; Naughton et al.
46 2021) has made considerable progress in effectively applying this tool to quantify the virus in
47 wastewater and observe community-scale trends. The next phase of implementation should focus on
48 appropriate interpretation of results with respect to confidence or probability.

49 Bayes' Theorem can be used to calculate results in terms of likelihood, or in the language of
50 Bayesian statistics, posterior probability. Because of type I and II errors in PCR data, results are more
51 accurately reported in terms of probability, which includes a consideration of prior likelihood of an event
52 as well as false positive and negative rates. This approach computes the conditional probability of
53 reliable result (true negative or true positive), then continually updates prior assumptions as new
54 information is acquired. A Bayesian framework for interpreting analytical results from environmental
55 samples has been applied in the field of microbial source tracking (MST) (Kildare et al. 2007; Johnston et
56 al. 2013, Curtis and Gonzalez 2019, Chen et al. 2021). Here PCR markers for human-associated fecal
57 microbes are used to identify waterways potentially impacted by wastewater. Employing a Bayesian
58 approach characterizes results in terms of contamination probability rather than simple
59 presence/absence or concentration of the marker. This method provides a quantitative way to
60 incorporate new information (e.g. recent sample results) into the interpretation of subsequent results.

61 Applying a probabilistic framework for understanding the reliability of results requires only the
62 sensitivity and specificity of the analytical test and the ability to set a prior. Here we examine this
63 approach using the Centers for Disease Control (CDC) SARS-CoV-2 panel (N1, N2, N3) under a variety of

64 wastewater surveillance scenarios. The likelihood of true positive and true negative is examined using
65 individual assays across a range of prior probabilities and also using multiple assays concurrently. A
66 proposed technique for setting realistic priors based on wastewater sampling results is described using
67 data from a wastewater surveillance network with nine treatment facilities in southeastern Virginia.
68 Finally, cumulative probability is used to investigate the interpretation of multiple sample results under
69 two scenarios.

70

71 **Methods**

72 *CDC Assay Sensitivity and Specificity*

73 The CDC SARS-CoV-2 real-time reverse transcription PCR panel (N1, N2, N3) was used to assess
74 the likelihood that wastewater samples contained SARS-CoV-2 RNA based on analytical results, and
75 assay sensitivity and specificity. These assays were developed for clinical identification of SARS-CoV-2
76 infection but have been widely applied to environmental samples, particularly wastewater, for virus
77 detection and enumeration (e.g. Stadler et al. 2020, Agrawal et al. 2021, McClary-Gutierrez 2021). The
78 specificity of each assay was tested in-vitro by Lu et al. (2020) finding N1, N2 and N3 assays with
79 specificity values of 1.0, 1.0, and 0.955 respectively. Specificity can be described as the tendency of an
80 assay to not detect the target of interest when it is absent and is calculated by dividing the number of
81 true negative values by the number of test samples which did not contain the target. Thus, a specificity
82 of 0.955 suggests a false-negative rate of 4.5%.

83 The sensitivity of an assay describes its ability to detect the target when it is present.
84 Wastewater sensitivity values were calculated using regional clinical data from the Virginia Department
85 of Health (<https://www.vdh.virginia.gov/coronavirus/>) and SARS-CoV-2 wastewater results (Gonzalez et

86 al. 2020). Clinical PCR positive case counts by zip code for all zip codes within the 9 wastewater
87 treatment facility sewersheds in Hampton Roads, Virginia were gathered from May 15, 2020 through
88 November 17, 2020. True positives were defined as when a weekly wastewater sample from a facility
89 was positive and at least one zip code in the facility's catchment had a clinically confirmed infected
90 individual that week. For each of the 3 assays, the number of weekly wastewater samples from the 9
91 facilities were pooled. The number of samples with a true positive were divided by the total number of
92 samples to determine the sensitivity of each assay in wastewater. The sensitivity for N1, N2, and N3
93 assays are 75% (200/266), 83% (220/266), and 80% (212/266) respectively. Sensitivity and specificity as
94 used in subsequent equations are described below.

95

96 Sensitivity (Se) and Specificity (Sp);

97
$$Se = P(A^+ | CoV2^+)$$

98
$$Sp = P(A^- | CoV2^-)$$

99 where;

100 Se = assay sensitivity

101 Sp = assay specificity

102 A⁺ = a positive detection with a given assay

103 A⁻ = a negative analytical result with a given assay

104

105

106 *Bayes' Theorem*

107 Bayes' Theorem (eq 1) was applied to CDC assay sensitivity and the wastewater-derived specificity to
108 investigate the likelihood that SARS-CoV-2 is present in wastewater samples given positive and negative
109 analytical results.

110

111 Equation 1.

$$112 \quad P(\text{CoV2}^+ | A^+) = \frac{P(A^+ | \text{CoV2}^+) * P(\text{CoV2}^+)}{[P(A^+ | \text{CoV2}^+) * P(\text{CoV2}^+)] + [P(A^+ | \text{CoV2}^-) * P(\text{CoV2}^-)]}$$

113

114 where;

115 CoV2^+ = SARS-CoV-2 is present

116 CoV2^- = SARS-CoV-2 is absent

117

118 The posterior probability that SARS-CoV-2 is present in a sample given detection of each assay (N1, N2,
119 N3) was tested across all possible priors (0.0-1.0) using Equation 1. A similar conditional probability was
120 derived to test the likelihood that SARS-CoV-2 is absent from a sample given a negative assay result and
121 was also tested across all priors.

122 As noted, the specificity of the N1 and N2 assays determined by Lu et al. were 1.0. A specificity value of
123 1.0 guarantees the absence of false positive results regardless of assay sensitivity and prior probability.

124 As a result, it is difficult to observe the effect of differing approaches for setting priors and the

125 implications of using one or multiple assays. To investigate these factors, we have used a specificity
126 value of 0.999 for all calculations.

127

128

129 *Surveillance Using Multiple Assays Concurrently*

130 A multiple-assay approach to environmental sampling has been supported in the field of
131 microbial source tracking (MST) (Harwood et al. 2014; Ahmed et al. 2016; Jardé et al. 2018;) and is
132 currently employed by many researchers conducting wastewater surveillance of SARS-CoV-2 (Stadler et
133 al. 2020; Peñarrubia et al. 2020; Agrawal et al. 2021). The likelihood that SARS-CoV-2 is present in a
134 sample was investigated using multiple assays concurrently. Equation 2 describes the probability that
135 SARS-CoV-2 is present given positive detection of the virus with two assays. N1 and N2 assay were
136 selected for this analysis due to CDC's recommendation to remove N3 from the clinical panel (Lu et al.
137 2020). An analogous equation was used to test the likelihood that SARS-CoV-2 is present when failing to
138 detect the virus using two assays. All combinations of N1 and N2 detection and non-detection were
139 tested across the range of possible priors (0.0-1.0).

140

141 Equation 2.

$$142 \quad P(\text{CoV2}^+ | A_1^+ \cap A_2^+) = \frac{Se_1 * Se_2 * P(\text{CoV2}^+)}{Se_1 * Se_2 * P(\text{CoV2}^+) + (1 - Sp_1) * (1 - Sp_2) * P(\text{CoV2}^-)}$$

143

144 where;

145 Se_i = sensitivity of assay i

146 Sp_i = specificity of assay i

147 A_i^+ = detection of assay i

148 A_i^- = failing to detect assay i

149 *Prior Probabilities Based on Wastewater Detection*

150 Wastewater surveillance data using the N2 assay and the N1 and N2 assays concurrently were
151 used to set priors based on percent positive wastewater samples. In this case the prior probability of
152 SARS-CoV-2 in wastewater was set as the percent of samples which were positive for one or both assay
153 in the week prior. Wastewater data for these analyses come from the wastewater surveillance program
154 being conducted by the Hampton Roads Sanitation District (HRSD) and comprise a raw influent sample
155 from each of 9 regional treatment facilities. For a detailed description of sampling and analytical
156 methodologies see Gonzalez et al. 2020. Using wastewater-derived priors, posterior probability of SARS-
157 CoV-2 presence was calculated for wastewater samples using N2 alone and N1 and N2 concurrently. For
158 samples that were negative for N2 or both N1 and N2 the posterior probability that SARS-CoV-2 is
159 absent in a given wastewater sample was reported.

160

161 *Cumulative Probability from Multiple Samples*

162 The cumulative probability of SARS-CoV-2 based on results from three separate samples was
163 investigated for two scenarios representing a comparatively low and high initial prior. The low initial
164 prior (0.11) was based on the lowest possible non-zero value for percent positive treatment facilities (1
165 of 9 facilities with a detection) in the Hampton Roads wastewater surveillance network. The high initial

166 prior (0.50) was used to examine a situation where there is no reliable way to estimate prior likelihood
167 and thus priors are uninformed. Cumulative probability of the presence of SARS-CoV-2 was calculated
168 using Equation 4 for each sample and the resulting posterior probability was used as the prior for
169 subsequent samples.

170 Equation 4.

171
$$P(\text{CoV2}^+)_c = 1 - \prod (1 - P(\text{CoV2}^+|A^\pm))$$

172

173 Where;

174 $P(\text{CoV2}^+)_c$ = cumulative probability of SARS-CoV-2

175 $P(\text{CoV2}^+|A^\pm)$ = probability of SARS-CoV-2 given a positive or negative assay

176

177

178 **Results and Discussion**

179 *Confidence Using a Single Assay*

180 Bayes' Theorem was applied in conjunction with sensitivity and specificity values for the three
181 CDC assays to calculate posterior probabilities of true positive and true negative across all possible
182 priors (Figure 1). The high specificity values of each assay resulted in expectedly high posterior
183 probabilities of SARS-CoV-2 given a positive analytical result (Figure 1a). The resulting curves suggest
184 that when N1 or N2 are used, a positive analytical result is highly reliable (>0.97), even at priors below
185 0.05. While not as overwhelming as N1 or N2, the N3 assay also generated highly reliable positive

186 results with posterior probabilities of SARS-CoV-2 >0.86 for priors >0.25 . The marginally increased
187 likelihood of false positive results for N3 was expected given the slightly lower specificity (0.955). As
188 described by Lu et al., the N3 assay was intentionally designed to also detect SARS-CoV, resulting in a
189 lower specificity value when cross-reactivity was tested against closely related respiratory viruses. The
190 likelihood of obtaining a true negative result was high with all assays when SARS-CoV-2 was absent
191 (Figure 1b). Values denoting likelihood of true negative result are more similar among assays than those
192 for true positives, owing to the similarity in calculated sensitivity values. Findings suggest that when
193 using an uninformed prior (0.50) a single assay generates acceptable confidence (0.80) in negative
194 results. Given no additional data to inform prior selection all assays performed well and result in a
195 reasonably high likelihood of true negative results. If there is reason to reduce the estimate of prior
196 probability the reliability of a true negative increases dramatically and is greater than 0.92 for all priors
197 <0.25 (Figure 1).

198

199 *Surveillance Using Multiple Assays Concurrently*

200 Many COVID-19 wastewater surveillance publications have concurrently used multiple SARS-
201 CoV-2 assays. When different combinations of N1 and N2 assay results were tested across all priors, the
202 resulting posterior probabilities for SARS-CoV-2 presence track in a generally predictable manner (Figure
203 2). When both assays detect the virus there is an exceedingly high probability (>0.99) that this detection
204 is reliable, even at very low priors (0.005). Similarly, when either assay is used individually and detects
205 the virus, these results are also highly reliable, as described in the previous section. Discordant results
206 (N1+/N2- or N2+/N1-) using two assays also yields a high probability that SARS-CoV-2 is present (>0.87),
207 even when priors are low (0.05). This is a potentially counterintuitive result which requires an
208 understanding of test characteristics (sensitivity and specificity) as well as the effect of prior probability

209 to appropriately interpret. Failing to consider these results in a Bayesian framework, one might tacitly
210 assign a likelihood of 0.5 to the probability that SARS-CoV-2 is present, given that two highly accurate
211 tests yielded conflicting results. Such an interpretation not only underestimates the likelihood that
212 SARS-CoV-2 is present, it also provides no more information about SARS-CoV-2 presence in wastewater
213 than what was known prior to sampling. In other words, this interpretation suggests that the virus is just
214 as likely present as not. Alternatively, one might assume a low prior probability that SARS-CoV-2 is
215 present, given low community incidence at the time of sampling, and therefore conclude that conflicting
216 results actually suggest the virus is absent by considering the one positive result a false positive. This
217 would also be an incorrect interpretation, given what is known about the sensitivity and specificity of
218 these assays. The implication of this interpretation is assigning a 0.0 probability that SARS-CoV-2 is
219 present when in fact there is a >0.93 likelihood that it is present for all priors above 0.10.

220

221 *Wastewater Detections to Inform Priors*

222 One strategy for selecting prior probability of SARS-CoV-2 in wastewater samples is to use
223 recent wastewater surveillance results. Surveillance programs often sample at routine intervals, and
224 some include multiple facilities. Therefore, a simple metric such as percent positive samples over recent
225 weeks or across multiple facilities can be used to inform prior probability. Relatively large priors (0.39-
226 1.0) resulted when percent positive facilities were calculated using the Hampton Roads surveillance data
227 due to the ubiquity of the virus in this region for the study period. With most priors being >0.50 positive
228 results are highly reliable for all samples, regardless if one or both assays were used (Fig 3). These
229 findings suggest that a single assay likely provides sufficient confidence in positive results when
230 detections are routinely seen at monitored facilities, causing priors to be set relatively high.

231 Conversely, the likelihood of true negative is somewhat driven down (Figure 3) when
232 wastewater-derived priors result in high values. This effect is understandable considering that
233 numerous recent detections should suggest a greater likelihood of detection for the current set of
234 samples. The consequence of high priors in constraining reliability of negative results is clearly shown in
235 the 7/21/20 data from Figure 3. This figure shows the likelihood of true positive for samples given
236 detection of one or both assay and likelihood of true negative when one or both assays fail to detect
237 viral RNA. Here, given a prior probability of 83% that SARS-CoV-2 is present, a single negative N2 result
238 indicates a 54% likelihood of true negative. Finding negative results for both assays increases the
239 likelihood of true negative, but only to 74%. This is a significant departure for the reliability associated
240 with concordant negative results when priors are lower (<0.15), which approaches 100%. In a scenario
241 of high priors based on numerous recent detections a negative result should be assigned less reliability
242 than one when few recent samples are positive, given that wastewater detection of viral RNA is an
243 indication of community incidence for a catchment. However, such an interpretation may not be
244 intuitive if one simply considers the sensitivity of the assay but fails to account for the prior likelihood of
245 detection.

246 Finding conflicting results when running multiple assays in the context of higher, wastewater-
247 derived priors also produces potentially difficult to interpret results. For the range of wastewater-
248 derived priors tested (0.39-1.0), discordant results regardless of which assay was detected, provide
249 strong evidence that SARS-CoV-2 is present (Figure 3). As described above, failing to consider prior
250 probability of detection could lead a researcher to interpret conflicting results as a 0.5 likelihood that
251 SAR-CoV-2 RNA is present in the sample. Interpretation using a Bayesian framework would suggest that,
252 even at the lowest prior for which there were conflicting results (0.56), the likelihood that the virus is
253 present is >0.99 .

254

255 *Multiple Samples*

256 Collecting multiple samples is an effective alternative for increasing reliability in results while
257 avoiding the need to run multiple assays. This approach generates high confidence via the cumulative
258 probability of true positive/negative results across samples, rather than the intersection of
259 positive/negative results for two or more assays. A multiple sample strategy was applied using a prior of
260 0.11 to examine the effect of a wastewater-derived prior for a situation where one of the nine facilities
261 in the HRSD surveillance network had a positive detection in the previous week. At this prior, any
262 combinations of samples that include a positive result yield >0.95 likelihood that viral RNA is present. A
263 negative result in sample one reduces the likelihood that SARS-CoV-2 is present from 0.11 to 0.02. In
264 this scenario the second sample is highly informative. Finding a positive result in sample two, even after
265 a negative for sample one, yields a 0.95 likelihood that the virus is present. High cumulative probability
266 that the virus is present given conflicting results from two samples using a single assay is consistent with
267 the analogous finding of high likelihood given the intersection of two conflicting, different assays run on
268 a single sample. With no detection in sample one, a second negative result has little impact on the
269 cumulative probability, which is 0.025, giving good evidence that the virus is absent. This pattern
270 continues for each sample three endpoint, wherein any positive result drives the cumulative probability
271 that SARS-CoV-2 is present in a set of three samples above 0.94 while consecutive negative results yield
272 cumulative probabilities (0.020 – 0.028) which remain one order of magnitude lower than the prior
273 (0.11).

274 Using an uniformed prior (0.50) results in a similarly high posterior probability (>0.99) that SARS-
275 CoV-2 is present for all combinations of sampling results other than three consecutive negatives. A key
276 difference with this prior is that even samples with only negative results still yield a 0.15-0.20 likelihood

277 that SARS-CoV-2 is present. As described previously, this strategy for setting priors is reasonable when
278 there is no reliable way to estimate likelihood of detecting viral RNA before samples have been
279 analyzed. Using a prior of 0.50 suggests a 50% likelihood that a sample contains the SARS-CoV-2 virus at
280 a concentration greater than the analytical limit of detection. The interaction of a prior that suggests an
281 equal chance of detecting RNA as not (0.50) with the high sensitivity and specificity of the N2 assay
282 generates some cumulative probabilities that may seem counterintuitive. The main examples include
283 sets of three samples for which two results are negative and one is positive but the cumulative
284 probability that SARS-CoV-2 is present is still >0.99 . Intuitively, two negative results, regardless of where
285 they occur in the sequence of samples, seem to suggest the absence of SARS-CoV-2. Following a
286 Bayesian approach in which priors are updated each time new data is acquired helps clarify these
287 results. As described, the posterior probability from each sample is used as the prior for subsequent
288 samples. Sets of samples where the positive result occurred first in sequence results in high posterior
289 probabilities which are carried through as priors for the remaining samples. For these cases, even
290 negative downstream results yield high posterior probabilities, driven by priors of >0.99 used in samples
291 two and three. In cases where sample one is negative, the prior of 0.50 and high sensitivity and
292 specificity prevent the posterior probability from being driven below 0.14. As a result, a positive
293 detection in sample one or two results in a cumulative probability of >0.99 .

294

295

296 *Limitations and Assumptions*

297 Applying a Bayesian framework for interpreting analytical findings is fundamentally driven by the
298 sensitivity and specificity values assigned to the assays in question, along with the assumed prior. As a
299 result, the accuracy of these values is highly consequential for the analysis. For this preliminary exercise,

300 in vivo specificity values similar to those reported by Lu et al. (2020) were used in conjunction with a
301 novel in situ approach to determining sensitivity values from a regional SARS-CoV-2 wastewater
302 monitoring program. As noted, specificity values of 0.999 were used in place of the 1.0 specificities
303 reported by Lu et al. (2020) for the N1 and N2 assays. The goal of this approach was to assume a false
304 positive rate of 1/1000 in order to conservatively examine the reliability of analytical results when
305 neither sensitivity nor specificity are perfect. Where possible, future studies should conduct in-house
306 specificity testing using their own analytical methodology performed on the matrix being studied to
307 derive lab-specific specificity values. A similar recommendation is suggested for deriving lab-specific
308 sensitivity values and is likely more impacted by wastewater matrix effects such as inhibition.

309 After establishing assay sensitivity and specificity, prior probability is the final value required to conduct
310 the simple conditional probability analyses described above. Prior probability is another critical piece of
311 this framework as it also exerts a significant influence on the outcome and the process for setting a
312 value may not be intuitive. Establishing the prior probability of an event or result can be challenging
313 when little is known about the system being studied. Regarding wastewater-based epidemiology one
314 suggested approach for estimating prior probability of SARS-CoV-2 detection is to use recent results
315 from the same location (e.g., the proportion of positive results from the last set or sets of samples) or
316 results from other areas within a common region. Priors may also comprise a distribution, allowing for
317 some incorporation of uncertainty. For instance, noting that clinical cases in the area of study have
318 been increasing rapidly in the preceding weeks does not suggest a specific value for use as a prior
319 probability but does provide information which should not be ignored when assessing reliability of
320 wastewater detections. Integrating this clinical information into a Bayesian analysis of wastewater
321 results might include setting priors using a uniform distribution of 0.51-0.99, implying that little is known
322 about the true prior probability of SARS-CoV-2 detection, but regional case data suggests that a

323 detection is likelier than not. Performed this way the posterior probability is also a distribution which
324 allows for uncertainty around the estimate to be easily included.

325

326

327 AUTHOR INFORMATION

328 Corresponding Author

329 *Corresponding author contact information

330 Kyle Curtis

331 Hampton Roads Sanitation District

332 1434 Air Rail Avenue

333 Virginia Beach, VA 23455

334 Email: kcurtis@hrsdc.com

335

336 Author Contributions

337 The manuscript was written through contributions of all authors. All authors have given approval to the
338 final version of the manuscript.

339

340 Funding Sources

341 All funding for this work was provided by Hampton Roads Sanitation District.

342

343

344

345 **References**

346 1. Dong, E.; Du, H.; Gardner, L. An Interactive Web-Based Dashboard to Track COVID-19 in Real

347 Time. *Lancet Infect. Dis.* 2020, 3099 (20), 19– 20, DOI: 10.1016/S1473-3099(20)30120-1

348 2. Aaron Bivins, Devin North, Arslan Ahmad, Warish Ahmed, Eric Alm, Frederic Bivins, A., North, D.,

349 Ahmad, A., Ahmed, W., Alm, E., Been, F., Bhattacharya, P., Bijlsma, L., Boehm, A.B., Brown, J. and

350 Buttiglieri, G., 2020. Wastewater-based epidemiology: global collaborative to maximize

351 contributions in the fight against COVID-19. Been, Prosun Bhattacharya, Lubertus Bijlsma, Alexandria

352 B Boehm, Joe Brown, Gianluigi Buttiglieri, Vincenza Calabro, Annalaura Carducci, Sara Castiglioni,

353 Zeynep Cetecioglu Gurol, Sudip Chakraborty, Federico Costa, Stefano Curcio, Francis L de los Reyes

354 III, Jeseth Delgado Vela, Kata Farkas, Xavier Fernandez-Casi, Charles Gerba, Daniel Gerrity, Rosina

355 Girones, Raul Gonzalez, Eiji Haramoto, Angela Harris, Patricia A Holden, Md Tahmidul Islam, Davey L

356 Jones, Barbara Kasprzyk-Hordern, Masaaki Kitajima, Nadine Kotlarz, Manish Kumar, Keisuke Kuroda,

357 Giuseppina La Rosa, Francesca Malpei, Mariana Mautus, Sandra L McLellan, Gertjan Medema, John

358 Scott Meschke, Jochen Mueller, Ryan J Newton, David Nilsson, Rachel T Noble, Alexander Van Nuijs,

359 Jordan Peccia, T Alex Perkins, Amy J Pickering, Joan Rose, Gloria Sanchez, Adam Smith, Lauren

360 Stadler, Christine Stauber, Kevin Thomas, Tom Van Der Voorn, Krista Wigginton, Kevin Zhu, Kyle

361 Bibby. 2020. Wastewater-based epidemiology: global collaborative to maximize contributions in the

362 fight against COVID-19. *Environmental Science and Technology*, 54(13), pp.7754–7757

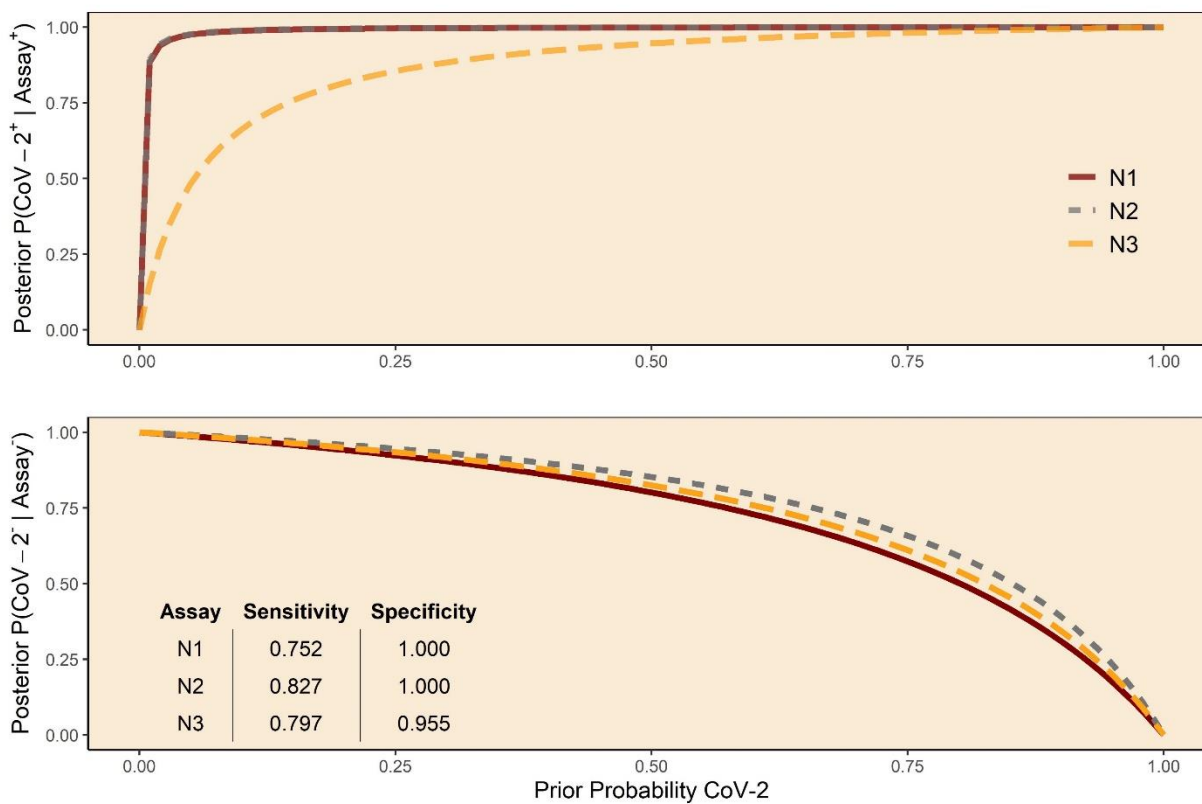
- 363 3. Naughton, C.C., Roman, F.A., Alvarado, A.G.F., Tariqi, A.Q., Deeming, M.A., Bibby, K., Bivins, A., Rose,
364 J.B., Medema, G., Ahmed, W. and Katsivelis, P., 2021. Show us the data: Global COVID-19
365 wastewater monitoring efforts, equity, and gaps. medRxiv.
- 366 4. Kildare, B.J., Leutenegger, C.M., McSwain, B.S., Bambic, D.G., Rajal, V.B. and Wuertz, S., 2007. 16S
367 rRNA-based assays for quantitative detection of universal, human-, cow-, and dog-specific fecal
368 Bacteroidales: a Bayesian approach. *Water research*, 41(16), pp.3701-3715.
- 369 5. Johnston, C., Byappanahalli, M.N., Gibson, J.M., Ufnar, J.A., Whitman, R.L. and Stewart, J.R., 2013.
370 Probabilistic analysis showing that a combination of *Bacteroides* and *Methanobrevibacter* source
371 tracking markers is effective for identifying waters contaminated by human fecal
372 pollution. *Environmental science & technology*, 47(23), pp.13621-13628.
- 373 6. Curtis, K. and Gonzalez, R.A., 2019. Integrating Bayesian Analysis and Cumulative Probability
374 Generates High Confidence Using a Single Microbial Source Tracking Marker. *Environmental science
375 & technology*, 53(23), pp.13929-13937.
- 376 7. Chen, H., Liu, C., Li, Y. and Teng, Y., 2021. Integrating metagenomic and Bayesian analyses to
377 evaluate the performance and confidence of CrAssphage as an Indicator for tracking human sewage
378 contamination in China. *Environmental Science & Technology*, 55(8), pp.4992-5000.
- 379 8. Stadler, L.B., Ensor, K., Clark, J.R., Kalvapalle, P., LaTurner, Z.W., Mojica, L., Terwilliger, A.L., Zhuo, Y.,
380 Ali, P., Avadhanula, V. and Bertolusso, R., 2020. Wastewater Analysis of SARS-CoV-2 as a Predictive
381 Metric of Positivity Rate for a Major Metropolis. MedRxiv.
- 382 9. Agrawal, S., Orschler, L. and Lackner, S., 2021. Long-term monitoring of SARS-CoV-2 RNA in
383 wastewater of the Frankfurt metropolitan area in Southern Germany. *Scientific reports*, 11(1), pp.1-
384 7.
- 385 10. McClary-Gutierrez, J., Mattioli, M., Marcenac, P., Silverman, A., Boehm, A., Bibby, K., Balliet, M.,
386 Gerrity, D., Griffith, J., Holden, P. and Katehis, D., 2021. Sars-Cov-2 Wastewater Surveillance for

- 387 Public Health Action: Connecting Perspectives From Wastewater Researchers and Public Health
388 Officials During a Global Pandemic.
- 389 11. Lu, X., Wang, L., Sakthivel, S.K., Whitaker, B., Murray, J., Kamili, S., Lynch, B., Malapati, L., Burke,
390 S.A., Harcourt, J. and Tamin, A., 2020. US CDC real-time reverse transcription PCR panel for
391 detection of severe acute respiratory syndrome coronavirus 2. *Emerging infectious diseases*, 26(8),
392 p.1654.
- 393 12. Gonzalez, R., Curtis, K., Bivins, A., Bibby, K., Weir, M.H., Yetka, K., Thompson, H., Keeling, D.,
394 Mitchell, J. and Gonzalez, D., 2020. COVID-19 surveillance in Southeastern Virginia using
395 wastewater-based epidemiology. *Water research*, 186, p.116296.
- 396 13. Harwood, V.J., Staley, C., Badgley, B.D., Borges, K. and Korajkic, A., 2014. Microbial source tracking
397 markers for detection of fecal contamination in environmental waters: relationships between
398 pathogens and human health outcomes. *FEMS microbiology reviews*, 38(1), pp.1-40.
- 399 14. Ahmed, W., Sidhu, J.P., Smith, K., Beale, D.J., Gyawali, P. and Toze, S., 2016. Distributions of fecal
400 markers in wastewater from different climatic zones for human fecal pollution tracking in Australian
401 surface waters. *Applied and environmental microbiology*, 82(4), pp.1316-1323.
- 402 15. Jardé, E., Jeanneau, L., Harrault, L., Quenot, E., Solecki, O., Petitjean, P., Lozach, S., Chevé, J. and
403 Gourmelon, M., 2018. Application of a microbial source tracking based on bacterial and chemical
404 markers in headwater and coastal catchments. *Science of The Total Environment*, 610, pp.55-63.
- 405 16. Peñarrubia, L., Ruiz, M., Porco, R., Rao, S.N., Juanola-Falgarona, M., Manissero, D., López-Fontanals,
406 M. and Pareja, J., 2020. Multiple assays in a real-time RT-PCR SARS-CoV-2 panel can mitigate the risk
407 of loss of sensitivity by new genomic variants during the COVID-19 outbreak. *International Journal of*
408 *Infectious Diseases*, 97, pp.225-229.
- 409

410

411

412 Figure 1.



413

414 Posterior probability of true positive (top) and true negative (bottom) result for the N1, N2, and N3

415 assays across all possible prior probabilities.

416

417

418

419

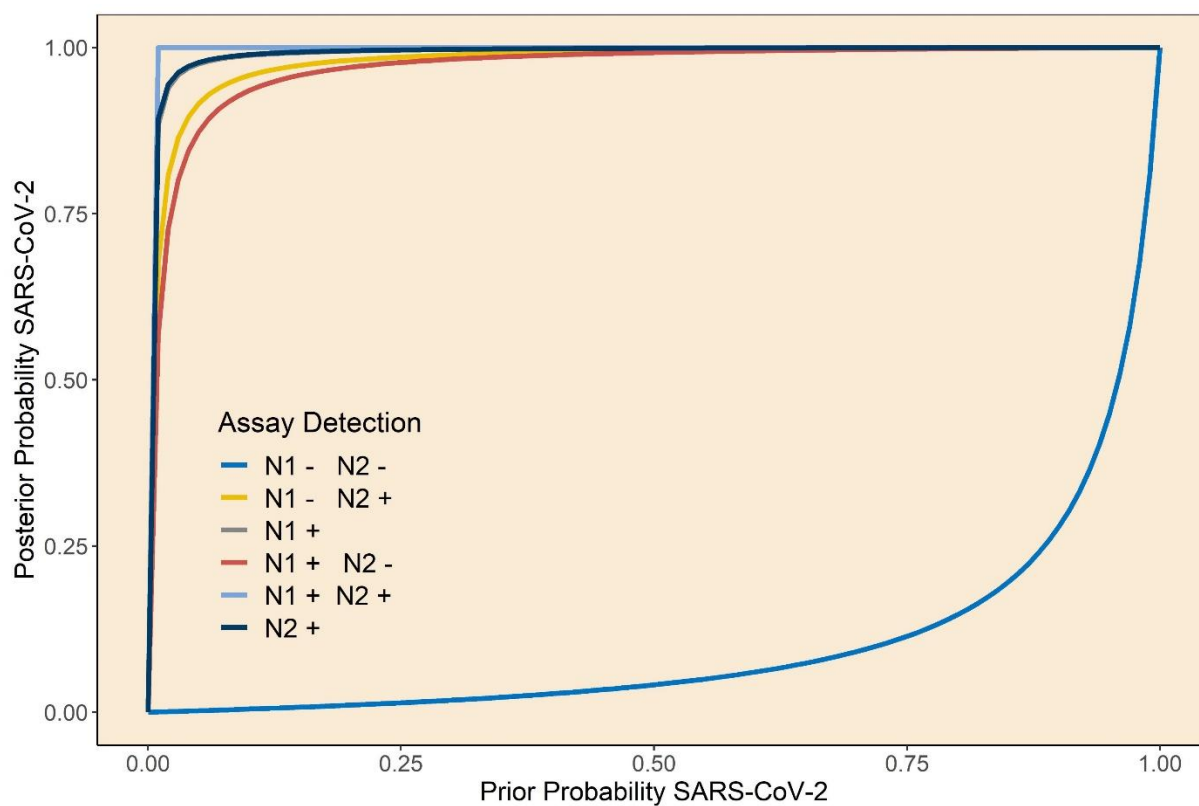
420

421

422

423

424 Figure 2.



425

426 Posterior probability that SARS-CoV-2 nucleic acid is present given combinations of results using one or
427 both of the N1 and N2 assays. Posteriors plotted across all possible prior probabilities that SARS-CoV-2
428 is present.

429

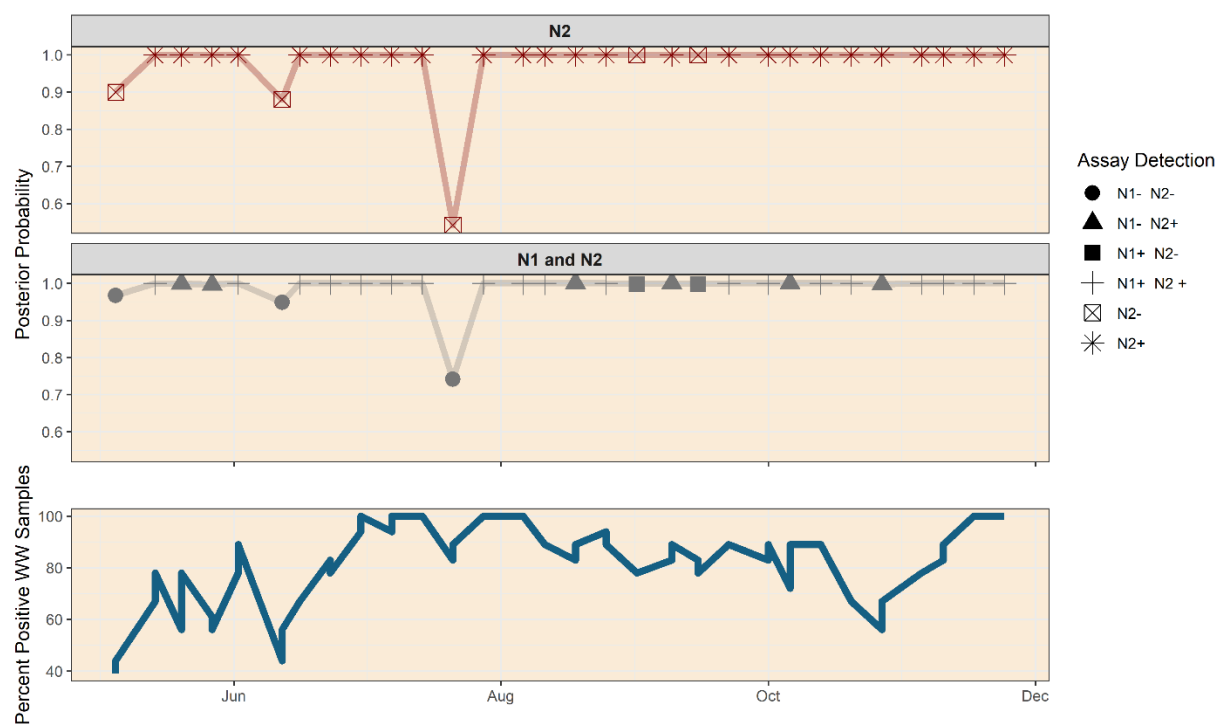
430

431

432

433 Figure 3.

434



435

436 Posterior probability of true positive or true negative using the N2 assay (top). The middle panel shows

437 the likelihood of true positive given detection with one or both of N1 and N2 assays and true negative

438 when both assays used concurrently fail to detect the virus. The bottom panel indicates the number of

439 treatment facilities (n=9) in which the virus was detected. These values are used as priors for calculating

440 posterior probabilities plotted in the top and middle panels.

441

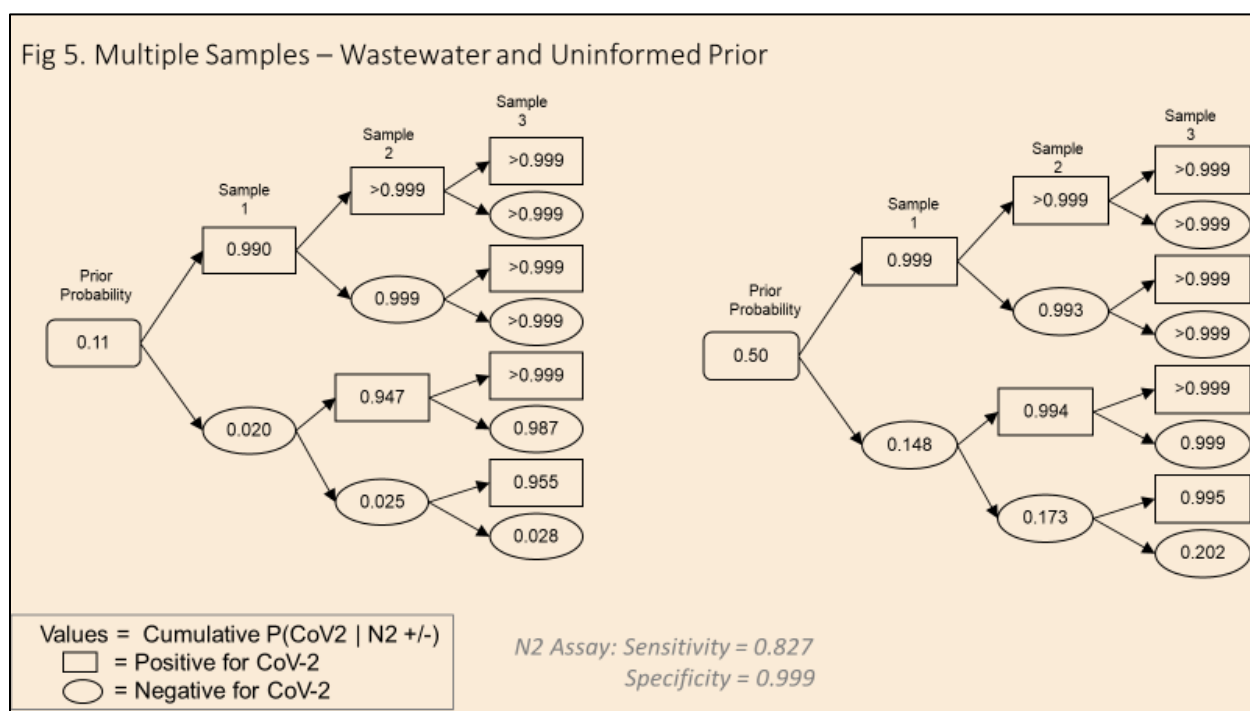
442

443

444

445 Figure 4.

446



447

448

449 Cumulative probability that SARS-CoV-2 nucleic acid is present given two initial priors (0.11, 0.50).

450 Cumulative posteriors are plotted across three sampling events with all combinations of

451 positive/negative results displayed.