

1 **SARS-CoV-2 diagnostic testing rates determine the sensitivity of genomic**
2 **surveillance programs**

3
4 Alvin X. Han^{1,*}, Amy Toporowski², Jilian A. Sacks³, Mark D. Perkins³, Sylvie Briand³, Maria
5 van Kerkhove³, Emma Hannay², Sergio Carmona², Bill Rodriguez², Edyth Parker⁴, Brooke E.
6 Nichols^{1,2,5,†}, Colin A. Russell^{1,5,†,*}

7 ¹Department of Medical Microbiology & Infection Prevention, Amsterdam University Medical
8 Center, University of Amsterdam, Amsterdam, The Netherlands

9 ²Foundation for Innovative New Diagnostics (FIND), Geneva, Switzerland

10 ³Department of Epidemic and Pandemic Preparedness and Prevention, Emergency Preparedness
11 Programme, World Health Organization, Geneva, Switzerland

12 ⁴Department of Immunology and Microbiology, The Scripps Research Institute, La Jolla, CA,
13 USA

14 ⁵Department of Global Health, School of Public Health, Boston University, Boston, MA, USA

15
16 †Contributed equally

17 *Corresponding authors. Email: x.han@amsterdamumc.nl and c.a.russell@amsterdamumc.nl
18

19 **Abstract (143 words)**

20 The first step in SARS-CoV-2 genomic surveillance is testing to identify infected people.

21 However, global testing rates are falling as we emerge from the acute health emergency and

22 remain low in many low- and middle-income countries (LMICs) (mean = 27 tests/100,000

23 people/day). We simulated COVID-19 epidemics in a prototypical LMIC to investigate how

24 testing rates, sampling strategies, and sequencing proportions jointly impact surveillance

25 outcomes and showed that low testing rates and spatiotemporal biases delay time-to-detection of

26 new variants by weeks-to-months and can lead to unreliable estimates of variant prevalence even

27 when the proportion of samples sequenced is increased. Accordingly, investments in wider

28 access to diagnostics to support testing rates of ~100 tests/100,000 people/day could enable more

29 timely detection of new variants and reliable estimates of variant prevalence. The performance of

30 global SARS-CoV-2 genomic surveillance programs is fundamentally limited by access to

31 diagnostic testing.

32 **Introduction (3489 words excluding Methods)**

33 Since the start of the COVID-19 pandemic in 2019, unprecedented expansion of genomic
34 surveillance efforts has led to the generation of more than 10 million SARS-CoV-2 sequences
35 deposited in the publicly accessible GISAID database (<https://www.gisaid.org/>) as of May 2022.
36 These efforts have been integral to understanding the COVID-19 pandemic¹, including the
37 identification of the Alpha variant in the United Kingdom during the fall 2020², the Delta variant
38 in India in late 2020³, and the Omicron variant in Southern Africa in November 2021⁴. Despite
39 the value of these efforts for monitoring the evolution of SARS-CoV-2, the intensity of genomic
40 surveillance is highly heterogenous across countries. High-income countries (HICs) on average
41 produced 16 times more SARS-CoV-2 sequences per reported case than low- and middle-income
42 countries (LMICs) as a result of longstanding socioeconomic inequalities and consequent
43 underfunding of laboratory and surveillance infrastructures⁵. To strengthen global pandemic
44 preparedness, initiatives such as the Access to COVID-19 Tools Accelerator Global Risk
45 Monitoring Framework, the Pan American Health Organization COVID-19 Genomic
46 Surveillance Regional Network, the Africa Pathogen Genomics Initiative, as well as the Global
47 Influenza Surveillance and Response System, among others, have supported LMICs in
48 developing pathogen genomic surveillance programs.

49
50 As resources are finite, it is critical that sequencing sample sizes, and the diagnostic testing
51 needed to obtain samples for sequencing, are carefully set for genomic surveillance programs to
52 detect and monitor variants as efficiently as possible. Current recommended sample sizes are

53 based on sampling theory⁵⁻⁸ and assume that the volume of diagnostic testing is large enough
54 such that the diversity of sampled viruses is representative of the diversity of viruses circulating
55 in the population. However, LMICs test at a mean rate of 27 tests per 100,000 persons per day
56 (tests/100k/day) as opposed to >800 tests/100k/day across HICs based on observational data
57 collected between January 2020 and March 2022⁹, with even higher testing rates in some HICs
58 (Fig. 1). Low testing rates lead to spotty information and smaller virus specimen pools available
59 for sequencing, resulting in sampling biases. These factors can render efforts to monitor the
60 emergence of new variants or prevalence of existing variants highly unreliable.

61
62 Here, we studied how different testing rates can impact genomic surveillance outcomes.
63 Specifically, we developed and used the *Propelling Action for Testing And Treating* (PATAT)
64 model, an individual-based modelling framework, to simulate concurrently-circulating wild-type
65 SARS-CoV-2 (pre-Alpha viruses)/Alpha-like epidemics as well as Delta-/Omicron (BA.1)-like
66 epidemics in Zambia as a representative LMIC archetype where recent demographic census data
67 required by the model was available (Online Methods). We assumed that Alpha and Omicron
68 (BA.1) were more transmissible than the respective extant virus to achieve growth rates of
69 ~0.15/day and ~0.35/day respectively^{2,10} and simulated SARS-CoV-2 infection waves in a
70 population of 1,000,000 individuals over a 90-day period that begins with an initial 1%
71 prevalence of the extant SARS-CoV-2 variant and the mutant variant being introduced at 0.01%.
72 We assumed that clinic-based professional-use Antigen Rapid Diagnostic Tests (Ag-RDTs) form
73 the basis of testing given persistent reports that polymerase chain reaction (PCR) tests are poorly

74 accessible for detection of symptomatic cases outside of tertiary medical facilities in many
75 LMICs¹¹.

76

77 We then simulated different genomic surveillance sampling strategies to elucidate how testing,
78 sequencing volumes and the degree of sampling bias arising from sources of specimens jointly
79 impact the timeliness of variant detection and the accuracy of variant monitoring (Online
80 Methods). These strategies include: (i) sending all samples from community clinics and tertiary
81 hospitals to a centralized facility for possible sequencing (i.e. population-wide strategy); (ii)
82 sampling and sequencing a portion of positive specimens collected at one tertiary sentinel facility
83 for the population of 1,000,000 simulated people (including mild individuals seeking
84 symptomatic testing and severe patients who sought tertiary care at the facility); or sampling and
85 sequencing a portion of positive specimens collected at (iii) 10%, (iv) 25%, (v) 50%, and (vi)
86 100% of all tertiary sentinel facilities.

87

88 **Results**

89 Performance of current guidance

90 We first assessed various suggested sample sizes of positive specimens to sequence to detect
91 SARS-CoV-2 variants at low prevalence for simulated wild-type/Alpha and Delta/Omicron
92 epidemics in Zambia with a mean testing rate of 27 tests/100k/day (based on the observed mean
93 rate of testing in LMICs) (Fig. 2). We used recommended sample sizes from three prominent

94 guidances: (1) The World Health Organization and European Centre for Disease Prevention and
95 Control computed sample size using the binomial method^{7,8}; (2) By subsampling genomic
96 surveillance data generated in Denmark in 2020-2021 when the country was testing at >2,000
97 tests/100k/day on average, Brito et al. suggested that sequencing 0.5% of all detected cases with
98 a turnaround time of 21 days would result in a 20% of variant detection before reaching 100
99 cases⁵; (3) Wohl et al. formulated a novel framework computing sequencing sample size by
100 modeling the biological and logistical processes that impact sampled variant proportions⁶.
101 Critically, all three methods did not consider how low testing rates and spatial nonuniformity in
102 sampling coverage impact sampled variant proportions, and in turn, speed of variant detection.
103 The assumptions, mathematical background, and lack of accounting for spatiotemporal bias in
104 sample size estimation of each guidance are detailed in Table 1 and Supplementary Notes.

105
106 As such, even when assuming negligible turnaround time (i.e. time from specimen collection to
107 acquisition of sequencing data), the recommended approaches were insufficient to detect the
108 variant on their respective target detection day when testing rates were low, due to poor
109 representativeness, regardless of the genomic surveillance sampling strategy. The first strategy of
110 sampling specimens collected from the whole population that were sent to one sequencing
111 facility (i.e. population-wide strategy) led to the best performance (closest to target detection
112 day) for all recommendations, as it involves random uniform sampling of all available samples, a
113 fundamental assumption made by all current guidance. However, if the specimen pools available
114 for sequencing are restricted to those collected from a subset of sentinel tertiary facilities only,

115 the non-uniformity in sampling coverage results in spatiotemporal bias within the sequenced
116 samples, and leads to delayed detection of variants-of-concern (VOCs), which gets progressively
117 worse as the proportion of tertiary facilities performing sequencing decreases to one facility.

118

119 Variant detection

120 To elucidate how SARS-CoV-2 testing rates and the proportion of positive specimens sequenced
121 impact the speed of variant detection, we simulated wild-type SARS-CoV-2/Alpha and
122 Delta/Omicron epidemics at different Ag-RDT availability ranging from 27 tests/100k/day to
123 1,000 tests/100k/day (Fig. 3). We assumed that specimens to be sequenced are sampled on their
124 collection day, and varied the proportion of positive specimens to sample for sequencing each
125 day between 1% and 100%. We analyzed the impact of testing rates and sequencing proportions
126 on the expected day when the first specimen sampled for sequencing containing the variant was
127 collected as a measure of variant detection speed. In Fig. 3, we did not consider the time between
128 sample collection and sequencing nor the turnaround time to obtaining sequencing results as they
129 would only delay the actual day of variant detection by the assumed turnaround time.

130

131 For all testing rates, the relationship between the expected day when the first sample containing
132 the variant was collected and the proportion of positive specimens sequenced per day can be
133 described by a convex operating curve, reflecting rapidly diminishing returns in the speed of
134 variant detection as more specimens are sampled for sequencing. Across all genomic

135 surveillance sampling strategies, relatively larger marginal improvements to the speed of variant
136 detection are generally made when the sequencing proportion is increased up to ~10% of all
137 samples collected. Further sequencing only minimally shortens the expected time to variant
138 detection, as the operating curve asymptotically approaches the earliest possible day of detection.
139 Importantly, increasing SARS-CoV-2 testing allows smaller sequencing proportions to attain
140 similar detection day targets, and higher testing rates lower the earliest possible detection day.
141 For both the Alpha and Omicron variants, increasing testing rates from 27 tests/100k/day to 100
142 tests/100k/day brings forward the expected day of sampling the first variant sequence by at least
143 one week (Fig. 3).

144

145 For the same level of testing and sequencing proportion, the population-wide strategy led to the
146 earliest initial detection of a variant sequence. If sequencing were restricted to samples collected
147 at a subset of tertiary sentinel facilities only, increasing the number of facilities sending samples
148 for sequencing reduced the spatiotemporal bias in the specimen pool, thereby shaping the
149 operating curves closer to the ones observed for the population-wide strategy. Interestingly,
150 results similar to the population-wide strategy could be attained if all tertiary facilities acted as
151 sentinel sites and sent the samples they collected for sequencing to increase the
152 representativeness of sampling.

153

154 Observed variant proportion

155 Test availability and sampling coverage also affect the accuracy of the observed variant
156 proportion (Figs. 4 and Extended Data S1). At a testing rate of 27 tests/100k/day, the observed
157 variant proportion maximally differs from the true circulating proportion by >30% for both the
158 Alpha and Omicron variants and for more than 15% of the time, the proportional difference
159 between the observed and true variation was greater than 20%. Both the maximum absolute
160 difference and percentage of timepoints where the difference is >20% can be lowered to <20%
161 and <5% respectively if testing rate is increased to 100 or more tests/100k/day.

162
163 Critically, when the representativeness of the specimen pool is spatiotemporally biased by
164 sequencing samples collected at tertiary sentinel facilities only, increasing the proportion of
165 specimens to be sequenced only marginally lowers the maximum absolute difference or lessens
166 the number of times where observed variant proportion deviates less than 20% from true
167 circulating proportions (Fig. 4, near vertical isoclines at low daily rates of testing). Increasing
168 testing rates at sentinel surveillance sites provides more accurate detection in changes to
169 circulating prevalence than sequencing more samples in the context of low testing rates.

170

171 Sensitivity analyses

172 We repeated our analyses using virus properties (i.e., incubation period, maximum viral load,
173 protection against infection by the mutant virus after extant virus infection) of the Omicron
174 variant but varied different relative transmissibility to the Delta variant (1.0 to 4.0) as well as the

175 initial proportion of individuals who had been infected by the Delta variant (10% and 40%). The
176 variant growth rates simulated for these hypothetical Delta/Omicron epidemics ranged from
177 0.17/day to 0.42/day.

178

179 Under these varied conditions, the expected day when the specimen of the first variant sequence
180 is collected still follows a convex-shaped operating curve against the daily proportion of positive
181 specimens to sequence. For all curves, the larger marginal improvements in shortening variant
182 detection are still in sequencing proportions of up to ~10% (Extended Data Fig. S2). In terms of
183 the accuracy of observed variant to true circulating proportions, the maximum absolute
184 difference and percentage of timepoints where difference is >20% are both substantially lowered
185 if testing rate is increased to at least 100 tests/100k/day (Extended Data Fig. S3-4).

186

187 We also varied the prevalence of extant Delta infections when the Omicron variant was
188 introduced (Extended Data Fig. S5). We found that lower test availability causes a delay in
189 sampling the first variant specimen if the variant is introduced when pre-existing extant variant
190 circulation is high. At 27 tests/100k/day, regardless of specimen proportions sequenced,
191 detection could be delayed by ~1 week if Omicron was introduced when Delta was circulating at
192 10% prevalence as opposed to 1%. This is because a greater share of tests would be used to
193 diagnose the more prevalent extant virus infections which in turn decreases the likelihood of
194 detecting the newly introduced variant at low proportions.

195

196 **Discussion**

197 Our findings show that the emphasis on the proportion of samples referred for genomic
198 surveillance is misplaced if testing capacity is insufficient and sample sources are highly
199 spatiotemporally biased. As such, at the current mean rate of testing in LMICs (27
200 tests/100k/day), current guidance⁵⁻⁸ on sequencing sample size estimation could likely lead to
201 later-than-predicted detection of novel variants at best or, at worst, leave new variants undetected
202 until they have infected a majority of a population.

203

204 Based on our work, we identified three major areas of improvement that could be prioritized to
205 enhance the robustness of genomic surveillance programs (Fig. 5). First, the most substantial
206 improvements are likely to come from increasing the mean testing rate in LMICs from 27
207 tests/100k/day (Fig. 5A) to at least 100 tests/100k/day (Fig. 5B). Even if one were to conduct
208 sentinel surveillance only at one tertiary facility, this increase in testing rate for the catchment
209 area of the facility would speed up variant detection by 1-2 weeks.

210

211 Second, the representativeness of a specimen pool for sequencing can be further improved by
212 expanding sampling coverage. In our model, variant detection was further sped up by 1-3 weeks
213 by increasing the percentage of tertiary sentinel facilities sending the samples they had collected
214 for sequencing to 25% of facilities (Fig. 5C). Additionally, in terms of prevalence monitoring, if

215 25% of tertiary facilities sequenced 5% of all positive specimens they had collected to detect and
216 monitor an Alpha-like variant, the maximum absolute difference to true circulating proportion is
217 expected to decrease from >50% (assuming a single sentinel facility) to no more than 20%.

218

219 Third, reducing turnaround time from samples referred to sequencing output results in a 1:1
220 decrease in time to new variant detection regardless of the proportion of sequenced samples, test
221 availability or sampling coverage (Fig. 5). These gains require scale up in sample transport
222 networks, access to sequencing machinery, trained personnel, and/or increases in numbers of
223 sequenced samples to make the most efficient use of each sequencing run¹¹. Furthermore, LMICs
224 also often face high costs and extended delivery delays of laboratory reagents and consumables
225 that were sometimes further exacerbated by recurring travel bans during the acute phase of the
226 pandemic^{5,12,13}.

227

228 After reducing spatiotemporal bias in the specimen pool through increased testing and sampling
229 coverage, sequencing up to 5-10% of the positive specimens collected could return the greatest
230 information gains while minimizing resource wastage. For an Alpha-like variant, 100
231 tests/100k/day with sampling from 25% of tertiary sentinel facilities for sequencing amounts to
232 an estimated 5-10 sequences per week averaged over a 90-day period per 1,000,000 people. If
233 turnaround time is kept within one week, the variant would likely be detected within one month
234 at ~4% circulating proportion (Fig. 5D). Similarly, at the same testing rate, sampling coverage

235 and turnaround time (i.e. average 5-11 sequences per week per 1,000,000 people), an Omicron-
236 like variant would be detected before the first month since its introduction but at ~23%
237 circulating proportion owing to its faster transmission (Extended Data Fig. S6).

238

239 Our findings serve to inform expectations of genomic surveillance initiative and should be
240 interpreted according to the public health objectives of each program. If the objective is to serve
241 as an early warning system for the *de novo* emergence of new variants before they are likely to
242 have spread widely, then all factors above can be considered essential and could require
243 substantially more than 100 tests/100k/day. Critically, determining that a new variant is a threat
244 requires not only detection of the variant itself but also the capacity to reliably monitor changes
245 in its prevalence and potential clinical impact on short timescales. The results presented here also
246 inform the design of programs for the sensitive and reliable detection of changes in variant
247 prevalence. Otherwise, if the objective is to detect for introduction of novel variants from
248 overseas, some of the factors above may be relaxed depending on the public health objectives.
249 For instance, if the aim is to attempt containment, all factors should still be considered to
250 promptly detect and monitor the spread of the variant. However, if the aim to ensure sufficient
251 time for control strategies to be enacted, less samples could be sequenced or turnaround time
252 could be longer, for example, so long as the mitigation strategies remain useful when
253 implemented.

254

255 Despite performing our simulations using demographic parameters from Zambia, the emergence
256 and detection of each VOC to date represents interesting case studies for the work described here
257 (Supplementary Notes). For example, at the time of first detection of the Omicron variant, in
258 South Africa in November 2021, the daily SARS-CoV-2 testing rate was 51 tests/100,000
259 people/day⁹, which was among the highest testing rates in Africa. However, the Omicron variant
260 was only detected 6-8 weeks after its likely emergence⁴. At that point, Omicron had already
261 infected a substantial portion of the population in Gauteng, South Africa (i.e., the estimated
262 circulating variant proportion was >80% by mid-November)⁴. Not only had the variant already
263 spread across the rest of South Africa and to neighboring Botswana⁴, Omicron samples were also
264 collected in multiple other countries, including Hong Kong¹⁴, Denmark¹⁵, and the Netherlands¹⁶
265 before the initial reports on the identification of the Omicron variant. This situation is consistent
266 with our modelling findings, where novel variant detection is possible with <100 tests/100k/day
267 but only after the new variant has spread widely across the population.

268
269 In another example, Germany randomly sequenced ~60-70 sequences per week (i.e. <1% of
270 cases sequenced per day) in December 2020¹⁷. During this time, testing rates in Germany
271 averaged at ~300 tests/100k/day⁹. Germany was able to detect the Alpha variant one week before
272 WHO declared the lineage a variant-of-concern in mid-December 2020¹⁷. The Alpha variant
273 likely emerged in the UK in mid-September 2020¹⁸ and rapidly proliferated across the country
274 before it was reported in December 2020¹⁹. Our analyses showed that the expected time before
275 the first Alpha variant specimen was sampled for sequencing since its introduction is >4 weeks

276 (i.e. around November 2020) at Germany's testing and sequencing rate. This falls in line with the
277 likely period of Alpha's introduction into Germany, similar to the period estimated for its
278 European neighbors such as the Netherlands²⁰.

279

280 While we find that routine representative sampling is vital for monitoring SARS-CoV-2
281 evolution, additional surveillance systems, including targeted surveillance of particular
282 populations and settings (such as immunocompromised individuals or unusual events), and
283 wastewater sampling, could enable increased variant detection sensitivity²¹. In particular, recent
284 advances in wastewater sequencing and deconvolution methods to resolve multiple viral lineages
285 in mixed wastewater samples enabled detection of emerging variants before they were captured
286 by clinical genomic surveillance²²⁻²⁴. However, sequence quality is often poor in wastewater
287 samples and in turn, these methods depend on a priori knowledge of the lineage-defining
288 mutations of VOCs and variants-of-interest, which are currently still identified based on
289 significant upsurges in clinically diagnosed cases. Furthermore, centralized wastewater
290 management systems, which these methods rely on for accurate determination of relative lineage
291 prevalence, are currently non-existent in many LMICs. Substantial investments, coordination and
292 time are needed to enable local sanitation infrastructures suitable for wastewater surveillance²⁵.
293 Detection of genetic markers such as S-gene target failure in PCR assays may also provide faster
294 notification of viral lineages with these specific mutations. However, whole genome sequencing
295 is still needed for unambiguous genotyping of SARS-CoV-2 samples. Ultimately, clinical

296 diagnostic testing and surveillance will remain the core mode of SARS-CoV-2 surveillance in
297 most countries.

298

299 During the initial phase of the pandemic in 2020, due to limited testing and sequencing
300 capacities, many LMICs were initially focused on genomic surveillance efforts at points of entry
301 at country borders to deter introductions^{26–28}. Over time, especially after the emergence of VOCs,
302 SARS-CoV-2 genomic surveillance gradually expanded to include community surveillance as
303 many LMICs enhanced their sequencing capacities^{4,28–30}. This was done either by establishing
304 regional sequencing networks to maximize available resources, investing in local sequencing
305 capacities or partnering with global collaborators^{30–32}. Sequencing turnaround time has also
306 improved from an average of ~170 days in 2020 to ~30 days in 2021 across the African
307 continent, albeit with substantial variation among countries³⁰. While sequencing capabilities have
308 expanded in LMICs, obtaining spatiotemporally representative samples remains a key
309 challenge³⁰. Our work shows that the sensitivity of genomic surveillance programs is highly
310 dependent on diagnostic testing rate and that a mean testing rate of 100 tests/100k/day at sentinel
311 sites that are geographically spread out across the community is a good basis for monitoring
312 virus variants. While a reflexive PCR test after a positive Ag-RDT diagnosis is currently
313 performed to obtain samples suitable for sequencing (and is possible in many tertiary facilities in
314 LMICs), this presents additional cost and logistical barriers. Recent studies showed that SARS-
315 CoV-2 sequencing can be performed using materials obtained from Ag-RDTs performed at
316 point-of-care^{33–35}. Importantly, whole genomes can be recovered up to eight days after testing,

317 providing opportunities for sequencing to be performed on samples performed through self-
318 testing as well³³.

319

320 Expanding genomic sequencing capabilities, especially in LMICs, is a global priority³⁶ and
321 current investments in sequencing must continue^{30,31}. Simultaneously, sustained investments in
322 public health systems are required to expand access to, and availability of, diagnostic testing to
323 underpin SARS-CoV-2 surveillance programs. Here, we primarily focused on LMICs but our
324 findings on the impact of testing rates and representativeness on genomic surveillance programs
325 are equally important for HICs as parts of their testing and surveillance infrastructures are
326 dismantled following the acute health emergency of the COVID-19 pandemic. Ultimately,
327 detecting the next SARS-CoV-2 variant or pathogen that causes the next pandemic requires
328 fundamental clinical diagnostic capacity to monitor existing and emerging pathogens.

329

330 **Online Methods**

331 Simulating SARS-CoV-2 epidemics with the Propelling Action for Testing And Treating

332 (PATAT) model

333 We used PATAT, a stochastic individual-based model to simulate SARS-CoV-2 epidemics in a
334 community with demographic profiles, contact mixing patterns, and level of public health
335 resources mirroring those typically observed in LMICs. Here, the model was based on Zambia.
336 PATAT creates an age-structured population, linking individuals within contact networks of
337 multi-generational households, schools, workplaces, and churches (i.e., regular mass gatherings)
338 (Extended Data Table S1). The simulated number of healthcare facilities (i.e., community clinics
339 and tertiary hospitals) where individuals with mild symptoms seek symptomatic testing and have
340 their virus specimens collected was based on an empirical clinic-to-population ratio (i.e. one
341 healthcare facility for every 7,000 individuals on average)^{37,38}. Although PATAT does not
342 explicitly simulate the spatial location of individuals, contact networks and healthcare facilities
343 are ordered to approximate localized community structures (i.e. the closer the number order of a
344 facility, the closer they are in the same neighborhood) that is most illustrative of urban centers.
345 Households are proximally ordered and distributed around these facilities based on an empirical
346 distance-structured distribution that correlates with probabilities of symptomatic individuals
347 seeking testing at clinics (Extended Data Table S1).

348

349 We then simulated SARS-CoV-2 infection waves in a population of 1,000,000 individuals over a
350 90-day period that begins with an initial 1% prevalence of an extant SARS-CoV-2 variant and

351 the introduction of a mutant variant at 0.01%. We assumed that clinic-based professional-use
352 Antigen Rapid Diagnostic Tests (Ag-RDTs) are the predominant SARS-CoV-2 diagnostic used
353 for SARS-CoV-2 testing.³⁹ As Ag-RDT sensitivity depends on within-host viral loads⁴⁰, PATAT
354 generates viral load trajectories, measured in cycle threshold (Ct) values, for infected individuals
355 by randomly sampling from known viral load distributions of different SARS-CoV-2
356 variants^{41,42}. We performed simulations for two variant replacement scenarios – Alpha variant
357 introduction while the wild-type virus was circulating (wild-type/Alpha) and Omicron (BA.1)
358 variant introduction while Delta was circulating (Delta/Omicron), applying known distributions
359 of their peak viral load, incubation, and virus clearance periods^{41,43} (Extended Data Table S1).
360 Before simulating the two-variant epidemic, we first calibrated the transmission probability
361 parameter for the extant variant such that it would spread in a completely susceptible population
362 at $R_0 = 2.5-3.0$. We then assumed Alpha and Omicron (BA.1) were more transmissible than the
363 respective extant virus to achieve growth rates of $\sim 0.15/\text{day}$ and $\sim 0.35/\text{day}$ respectively^{2,10}.

364
365 For both sets of simulations, we assumed that 10% of the population had infection-acquired
366 immunity against the extant strain initially with some level of protection against infection by the
367 mutant virus (wild-type SARS-CoV-2: 80% protection against Alpha⁴⁴; Delta: 20% protection
368 against Omicron¹⁰). We also investigated the scenario where 40% of the population had
369 infection-acquired immunity as part of sensitivity analyses (see below). We did not investigate
370 scenarios involving vaccine-acquired immunity due to low vaccine uptake in most LMICs⁴⁵.

371

372 PATAT uses the SEIRD (Susceptible-Exposed-Infected-Recovered/Death) epidemic model for
373 disease progression and stratifies infected individuals based on their symptom presentation
374 (asymptomatic, mild, or severe). After an assumed random delay post-symptom onset (mean = 1
375 day; s.d. = 0.5 day), symptomatic individuals who seek testing would do so at their nearest
376 healthcare facility, where test-positive samples may be reflexively collected for sequencing. We
377 assumed that symptomatic individuals sought testing based on a probability distribution of health
378 services-seeking behaviour that inversely correlates with the distance between the individual's
379 household and the nearest healthcare facility (Extended Data Table S1)⁴⁶.

380
381 We varied levels of Ag-RDT stocks per day (i.e., 27, 100, and 200-1,000 (in increments of 200)
382 tests/100k/day), running 10 independent epidemic simulations for each testing rate. Given the
383 start of a week on Monday, we assumed that a week's worth of tests are delivered to healthcare
384 facilities every Monday and unused Ag-RDTs in the previous week are carried forward into the
385 next week. If test stocks for a particular week were exhausted before the end of the week, testing
386 for the rest of that week ceased. Due to overlapping symptoms between COVID-19 and other
387 respiratory diseases, a proportion of available Ag-RDTs would be used by individuals who are
388 not infected with SARS-CoV-2. Based on test positivity rates reported by various countries in
389 the second half of 2021⁴⁷, we assumed 10% test positivity rate at the start and end of the
390 simulated epidemic, and 20% test positivity at its peak, linearly interpolating the rates between
391 these timepoints. We also assumed that false positive specimens could be sampled based on
392 reported Ag-RDT specificity of 98.9%⁴⁰.

393

394 We assumed that any specimens collected for genomic surveillance after positive detection
395 through Ag-RDT would be reflexively confirmed with PCR. We also assumed that all
396 symptomatic individuals who have severe symptoms require hospitalization, and are tested
397 separately from mild symptomatic persons who sought testing. Given that likely only ~10-20%
398 of COVID-19 deaths in Zambia were tested for the disease in life^{48,49}, we assumed that only 20%
399 of individuals with severe disease would be tested by Ag-RDT or PCR upon presenting severe
400 symptoms and have specimens collected for sequencing.

401

402 Full technical details of PATAT is described in the Supplementary Notes. The full model source
403 code is available at <https://github.com/AMC-LAEB/PATAT-sim>.

404

405 Genomic surveillance strategies

406 Twenty percent of healthcare facilities were assumed to be tertiary facilities based on empirical
407 data collected from Zambia^{37,38}. We assumed that tertiary facilities provide testing for mild
408 symptomatic individuals as well as hospitalized patients with severe symptoms. Given that
409 healthcare facilities were proximally ordered, we randomly selected tertiary facilities in each
410 independent surveillance simulation (see below) but ensured that the selected facilities were not
411 consecutively ordered. In sum, all tertiary facilities accounted for a median 18.4% (interquartile
412 range = 17.7-19.1%) of total testing volume across all simulations. We assumed that a proportion
413 of tertiary facilities serve as sentinel surveillance sites that reflexively collect SARS-CoV-2

414 positive samples for sequencing. We then simulated six strategies with varying degrees of
415 sampling coverage where positive specimens collected from testing sites would be consolidated
416 and sampled for sequencing: (i) all samples from community clinics and tertiary hospitals are
417 sent to a centralized facility and further sampled for sequencing (i.e. *population-wide* strategy);
418 (ii) only *one* tertiary sentinel facility for the population of 1,000,000 simulated people would
419 sequence a portion of positive specimens it has collected, both from mild individuals seeking
420 symptomatic testing and severe patients who sought tertiary care at the facility; or only (iii) 10%,
421 (iv) 25%, (v) 50%, and (vi) 100% of all tertiary sentinel facilities would sample and sequence a
422 proportion of the specimens they have collected.

423

424 For all strategies, we assumed that a proportion (1%-100%; in 2% increments between 1% and
425 5%, in 5% increments between 5% and 100%) of positive specimens are collected daily for
426 sequencing. We also assumed that positive specimens sampled within each week for sequencing
427 are consolidated into a batch before they are referred for sequencing. Turnaround time refers to
428 the time between collection of each weekly consolidated batch of positive specimens to the
429 acquisition of its corresponding sequencing data. Since the within-host viral loads of infected
430 individuals were simulated, we assumed that only high-quality samples where Ct values < 30
431 could be sequenced and that sequencing success rate is 80% as assumed in other studies⁶.

432

433 For each strategy and sequencing proportion, we performed 100 independent surveillance
434 simulations for each epidemic simulation with a given test stock availability, thus totaling to

435 1,000 random simulations for each set of variables (i.e., testing rate, sequencing proportion, and
436 strategy).

437
438 **Acknowledgments:** The authors are pleased to acknowledge that all computational work
439 reported in this paper was performed on the Shared Computing Cluster which is administered by
440 Boston University's Research Computing Services (www.bu.edu/tech/support/research/). This
441 work was supported by the Rockefeller Foundation, and the Governments of Germany, Canada,
442 UK, Australia, Norway, Saudi Arabia, Kuwait, Netherlands and Portugal. A.X.H. and C.A.R.
443 were supported by ERC NaviFlu (No. 818353). C.A.R. was also supported by NIH R01
444 (5R01AI132362-04) and an NWO Vici Award (09150182010027).

445 **Author contributions:** Conceptualization: A.X.H., B.E.N., C.A.R., J.A.S., M.D.P., S.B.,
446 M.V.K., A.T., E.H., S.C., B.R.; Methodology: A.X.H., B.E.N., C.A.R.; Software: A.X.H.;
447 Validation: A.X.H., E.P., B.E.N., C.A.R.; Formal analysis: A.X.H., B.E.N., C.A.R.;
448 Investigation: A.X.H., B.E.N., C.A.R. J.A.S., M.P., S.B., M.V.K.; Resources: B.E.N., C.A.R.,
449 A.T., E.H., S.C., B.R.; Data curation: A.X.H., B.E.N., C.A.R.; Visualization: A.X.H., E.P.,
450 B.E.N., C.A.R.; Supervision: B.E.N., C.A.R.; Project administration: B.E.N., C.A.R.; Funding
451 acquisition: B.E.N., C.A.R., A.T., E.H., S.C., B.R.; Writing – original draft: A.X.H., B.E.N.,
452 C.A.R.; Writing – review & editing: All authors.

453 **Declaration of interests:** A.T., E.H., S.C., B.R. and B.E.N. declare that they are employed by
454 FIND, the global alliance for diagnostics. J.A.S., M.D.P., S.B., M.V.K: The findings and

455 conclusions in this manuscript are those of the authors and do not represent the official position
456 of the World Health Organization.

457 **Data availability:** All data relevant to the study are included in the Article, the Supplementary
458 Information and the GitHub repository (<https://github.com/AMC-LAEB/PATAT-sim>). The
459 PATAT model source code is also available at <https://github.com/AMC-LAEB/PATAT-sim>.

460

461 **References**

- 462 1. Robishaw, J. D. *et al.* Genomic surveillance to combat COVID-19: challenges and
463 opportunities. *Lancet Microbe* **2**, e481–e484 (2021).
- 464 2. Davies, N. G. *et al.* Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7
465 in England. *Science (1979)* **372**, eabg3055 (2021).
- 466 3. Cherian, S. *et al.* SARS-CoV-2 Spike Mutations, L452R, T478K, E484Q and P681R, in
467 the Second Wave of COVID-19 in Maharashtra, India. *Microorganisms 2021, Vol. 9,*
468 *Page 1542* **9**, 1542 (2021).
- 469 4. Viana, R. *et al.* Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in
470 southern Africa. *Nature 2022* 1–10 (2022) doi:10.1038/s41586-022-04411-y.
- 471 5. Brito, A. F. *et al.* Global disparities in SARS-CoV-2 genomic surveillance. *medRxiv*
472 2021.08.21.21262393 (2021) doi:10.1101/2021.08.21.21262393.

- 473 6. Wohl, S., Lee, E. C., DiPrete, B. L. & Lessler, J. Sample Size Calculations for Variant
474 Surveillance in the Presence of Biological and Systematic Biases. *medRxiv*
475 2021.12.30.21268453 (2022) doi:10.1101/2021.12.30.21268453.
- 476 7. European Centre for Disease Prevention and Control. Sequencing of SARS-CoV-2 - first
477 update. <https://www.ecdc.europa.eu/en/publications-data/sequencing-sars-cov-2> (2022).
- 478 8. World Health Organization. Guidance for surveillance of SARS-CoV-2 variants: interim
479 guidance, 9 August 2021. <https://apps.who.int/iris/handle/10665/343775> (2021).
- 480 9. FIND. Test tracker - FIND. <https://www.finddx.org/covid-19/test-tracker/> (2022).
- 481 10. Imperial College London. Report 49 - Growth, population distribution and immune escape
482 of Omicron in England | Faculty of Medicine | Imperial College London.
483 [https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-49-](https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-49-Omicron/)
484 [Omicron/](https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-49-Omicron/) (2021).
- 485 11. Inzaule, S. C., Tessema, S. K., Kebede, Y., Ogwel Ouma, A. E. & Nkengasong, J. N.
486 Genomic-informed pathogen surveillance in Africa: opportunities and challenges. *Lancet*
487 *Infect Dis* **21**, e281–e289 (2021).
- 488 12. Mallapaty, S. Omicron-variant border bans ignore the evidence, say scientists. *Nature* **600**,
489 199 (2021).
- 490 13. Merhi, G. *et al.* SARS-CoV-2 genomic epidemiology: data and sequencing infrastructure.
491 *Future Microbiol* **17**, 1001–1007 (2022).

- 492 14. Gu, H. *et al.* Probable Transmission of SARS-CoV-2 Omicron Variant in Quarantine
493 Hotel, Hong Kong, China, November 2021. *Emerg Infect Dis* **28**, 460 (2022).
- 494 15. Espenhain, L. *et al.* Epidemiological characterisation of the first 785 SARS-CoV-2
495 Omicron variant cases in Denmark, December 2021. *Eurosurveillance* **26**, 2101146
496 (2021).
- 497 16. Rijksinstituut voor Volksgezondheid en Milieu. Omicron variant found in two previous
498 test samples | RIVM. [https://www.rivm.nl/en/news/omicron-variant-found-in-two-](https://www.rivm.nl/en/news/omicron-variant-found-in-two-previous-test-samples)
499 [previous-test-samples](https://www.rivm.nl/en/news/omicron-variant-found-in-two-previous-test-samples) (2021).
- 500 17. Oh, D. Y. *et al.* Advancing Precision Vaccinology by Molecular and Genomic
501 Surveillance of Severe Acute Respiratory Syndrome Coronavirus 2 in Germany, 2021.
502 *Clinical Infectious Diseases* **75**, S110–S120 (2022).
- 503 18. Hill, V. *et al.* The Origins and Molecular Evolution of SARS-CoV-2 Lineage B.1.1.7 in
504 the UK. *Virus Evol* (2022) doi:10.1093/VE/VEAC080.
- 505 19. Volz, E. *et al.* Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England.
506 *Nature* (2021) doi:10.1038/s41586-021-03470-x.
- 507 20. Han, A. X. *et al.* Regional importation and asymmetric within-country spread of SARS-
508 CoV-2 variants of concern in the Netherlands. *medRxiv* 2022.03.21.22272611 (2022)
509 doi:10.1101/2022.03.21.22272611.
- 510 21. Knyazev, S. *et al.* Unlocking capacities of genomics for the COVID-19 response and
511 future pandemics. *Nature Methods* 2022 19:4 **19**, 374–380 (2022).

- 512 22. Karthikeyan, S. *et al.* Wastewater sequencing reveals early cryptic SARS-CoV-2 variant
513 transmission. *Nature* 2022 1–4 (2022) doi:10.1038/s41586-022-05049-6.
- 514 23. Baaijens, J. A. *et al.* Variant abundance estimation for SARS-CoV-2 in wastewater using
515 RNA-Seq quantification. *medRxiv* 2021.08.31.21262938 (2021)
516 doi:10.1101/2021.08.31.21262938.
- 517 24. Jahn, K. *et al.* Early detection and surveillance of SARS-CoV-2 genomic variants in
518 wastewater using COJAC. *Nature Microbiology* 2022 7:8 7, 1151–1160 (2022).
- 519 25. Shrestha, S. *et al.* Wastewater-Based Epidemiology for Cost-Effective Mass Surveillance
520 of COVID-19 in Low- and Middle-Income Countries: Challenges and Opportunities.
521 *Water* 2021, Vol. 13, Page 2897 13, 2897 (2021).
- 522 26. Butera, Y. *et al.* Genomic sequencing of SARS-CoV-2 in Rwanda reveals the importance
523 of incoming travelers on lineage diversity. *Nature Communications* 2021 12:1 12, 1–11
524 (2021).
- 525 27. Salyer, S. J. *et al.* The first and second waves of the COVID-19 pandemic in Africa: a
526 cross-sectional study. *The Lancet* 397, 1265–1275 (2021).
- 527 28. Morang’a, C. M. *et al.* Genetic diversity of SARS-CoV-2 infections in Ghana from 2020-
528 2021. *Nature Communications* 2022 13:1 13, 1–11 (2022).
- 529 29. Cowley, L. A. *et al.* Genomics, social media and mobile phone data enable mapping of
530 SARS-CoV-2 lineages to inform health policy in Bangladesh. *Nature Microbiology* 2021
531 6:10 6, 1271–1278 (2021).

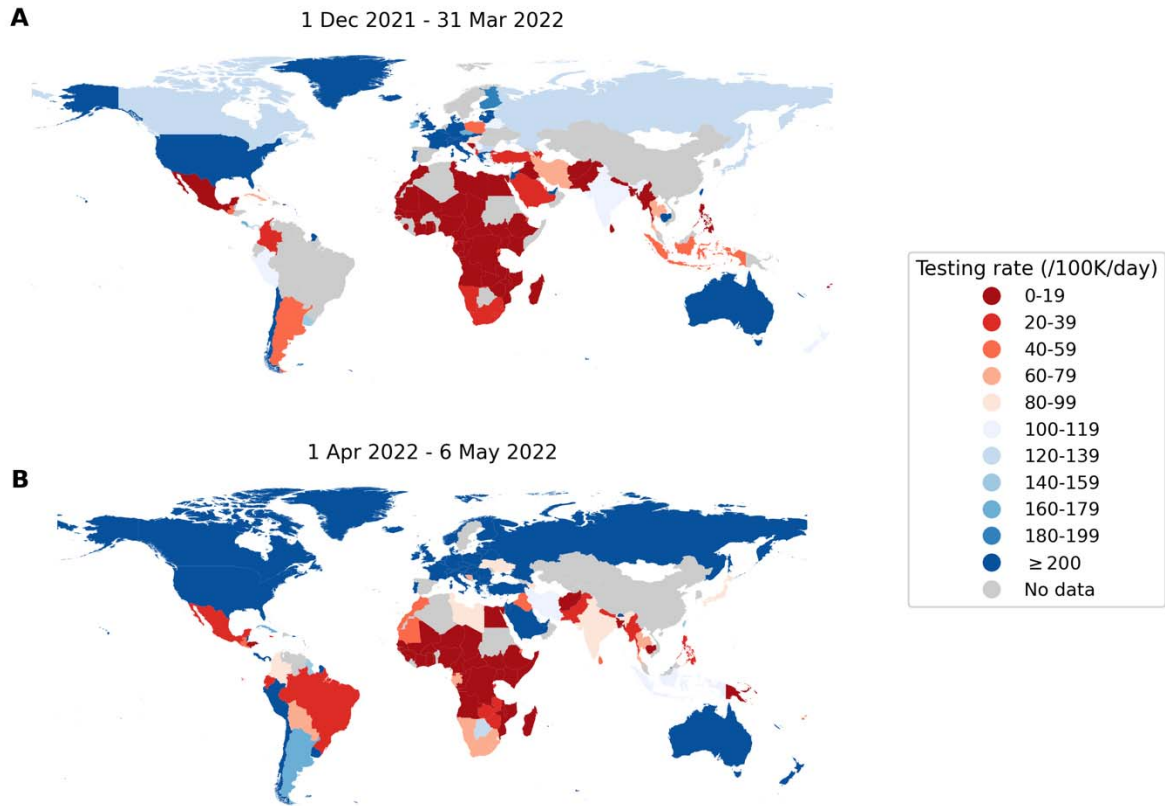
- 532 30. Tegally, H. *et al.* The evolving SARS-CoV-2 epidemic in Africa: Insights from rapidly
533 expanding genomic surveillance. *medRxiv* (2022)
534 doi:<https://doi.org/10.1101/2022.04.17.22273906>.
- 535 31. Leite, J. A. *et al.* Implementation of a COVID-19 Genomic Surveillance Regional
536 Network for Latin America and Caribbean region. *PLoS One* **17**, e0252526 (2022).
- 537 32. Dhar, M. S. *et al.* Genomic characterization and epidemiology of an emerging SARS-
538 CoV-2 variant in Delhi, India. *Science* (1979) **374**, 995–999 (2021).
- 539 33. Martin, G. E. *et al.* Maintaining genomic surveillance using whole-genome sequencing of
540 SARS-CoV-2 from rapid antigen test devices. *Lancet Infect Dis* **0**, (2022).
- 541 34. Nazario-Toole, A. *et al.* Sequencing SARS-CoV-2 from antigen tests. *PLoS One* **17**,
542 e0263794 (2022).
- 543 35. Macori, G. *et al.* Inactivation and Recovery of High Quality RNA From Positive SARS-
544 CoV-2 Rapid Antigen Tests Suitable for Whole Virus Genome Sequencing. *Front Public*
545 *Health* **10**, 1135 (2022).
- 546 36. Adepoju, P. Challenges of SARS-CoV-2 genomic surveillance in Africa. *Lancet Microbe*
547 **2**, e139 (2021).
- 548 37. Girdwood, S. J. *et al.* Optimizing viral load testing access for the last mile: Geospatial cost
549 model for point of care instrument placement. *PLoS One* **14**, e0221586 (2019).

- 550 38. Nichols, B. E. *et al.* Monitoring viral load for the last mile: what will it cost? *J Int AIDS*
551 *Soc* **22**, e25337 (2019).
- 552 39. Adepoju, P. Closing Africa’s wide COVID-19 testing and vaccination gaps. *Lancet*
553 *Microbe* **2**, e573 (2021).
- 554 40. Brümmer, L. E. *et al.* Accuracy of novel antigen rapid diagnostics for SARS-CoV-2: A
555 living systematic review and meta-analysis. *PLoS Med* **18**, e1003735- (2021).
- 556 41. Linton, N. M. *et al.* Incubation Period and Other Epidemiological Characteristics of 2019
557 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly
558 Available Case Data. *Journal of Clinical Medicine* 2020, Vol. 9, Page 538 **9**, 538 (2020).
- 559 42. Kissler, S. M. *et al.* Viral dynamics of acute SARS-CoV-2 infection and applications to
560 diagnostic and public health strategies. *PLoS Biol* **19**, e3001333- (2021).
- 561 43. Hay, J. A. *et al.* Viral dynamics and duration of PCR positivity of the SARS-CoV-2
562 Omicron variant. *medRxiv* 2022.01.13.22269257 (2022)
563 doi:10.1101/2022.01.13.22269257.
- 564 44. Pouwels, K. B. *et al.* Impact of Delta on viral burden and vaccine effectiveness against
565 new SARS-CoV-2 infections in the UK. *medRxiv* 2021.08.18.21262237 (2021)
566 doi:10.1101/2021.08.18.21262237.
- 567 45. Mathieu, E. *et al.* A global database of COVID-19 vaccinations. *Nature Human Behaviour*
568 *2021 5:7* **5**, 947–953 (2021).

- 569 46. Dovel, K. *et al.* Frequency of visits to health facilities and HIV services offered to men,
570 Malawi. *Bull World Health Organ* **99**, 618–626 (2021).
- 571 47. Hasell, J. *et al.* A cross-country database of COVID-19 testing. *Scientific Data* *2020 7:1*
572 **7**, 1–7 (2020).
- 573 48. Mwananyanda, L. *et al.* Covid-19 deaths in Africa: prospective systematic postmortem
574 surveillance study. *BMJ* **372**, (2021).
- 575 49. Gill, C. J. *et al.* Sustained high prevalence of COVID-19 deaths from a systematic post-
576 mortem study in Lusaka, Zambia: one year later. *medRxiv* 2022.03.08.22272087 (2022)
577 doi:10.1101/2022.03.08.22272087.
- 578
- 579

580

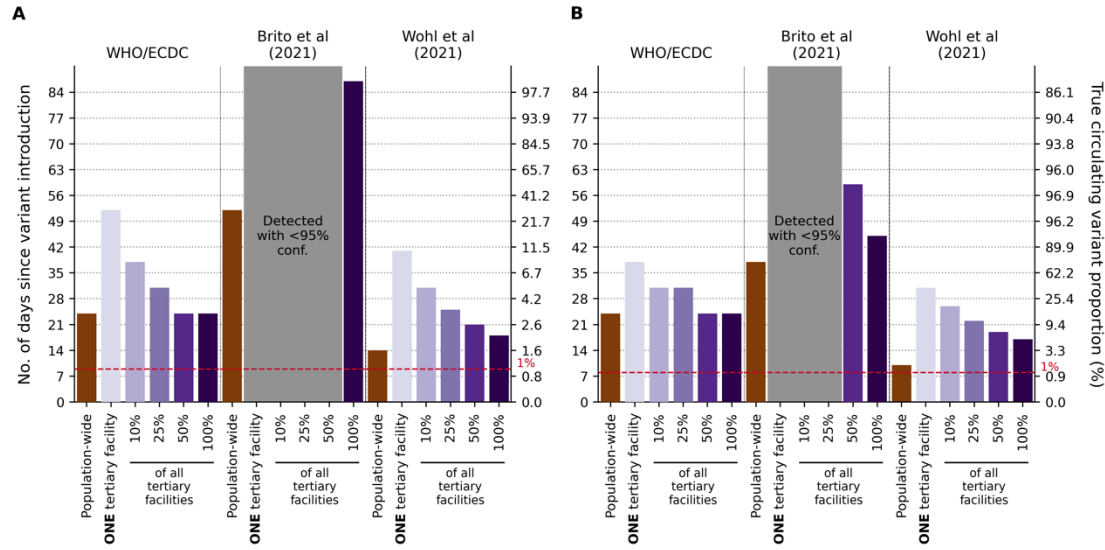
Figures



581

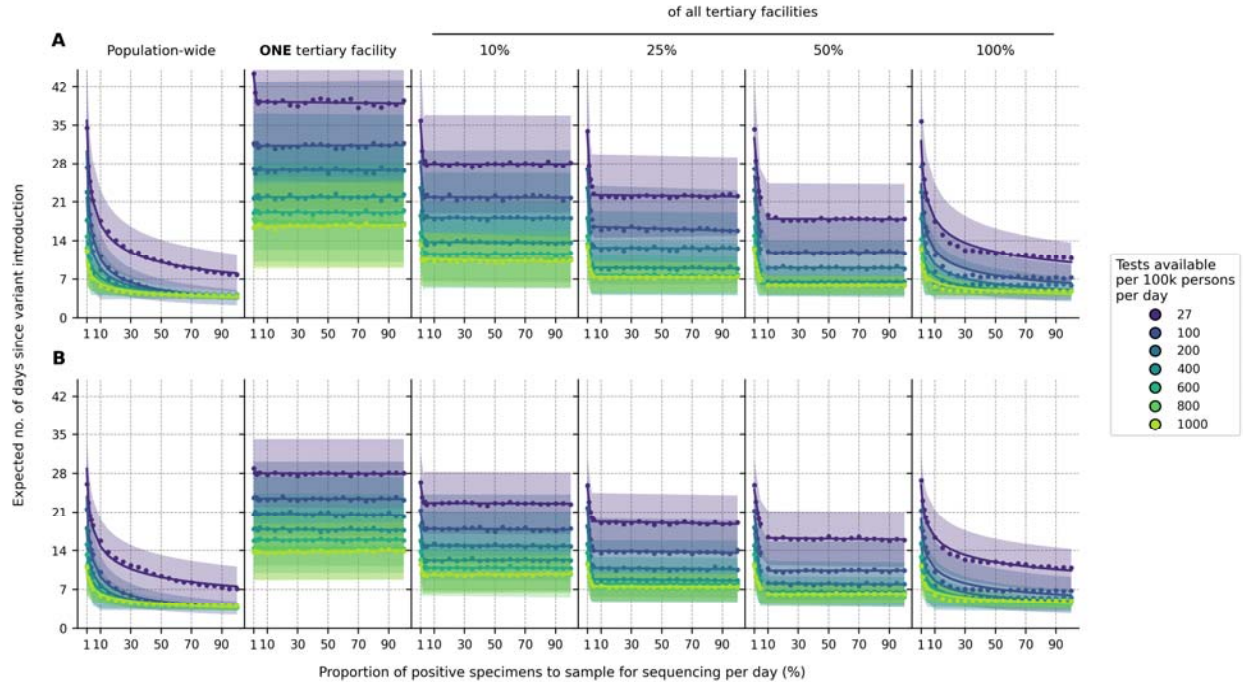
582 **Fig. 1. Global disparities in SARS-CoV-2 testing rates.** Each country is colored by the average
583 total number of SARS-CoV-2 tests performed per 100,000 persons per day (/100K/day) (A)
584 between 1 December 2021 and 31 March 2022 when the Omicron variant-of-concern spread
585 around the world; (B) between 1 April 2022 and 6 May 2022 when most countries were past
586 peak Omicron wave of infections⁹.

587



588

589 **Fig. 2. Performance of current guidance on number of positive specimens to sequence for**
590 **variant detection with testing rate at 27 tests per 100,000 persons per day.** First day of
591 detection since variant introduction at 95% confidence and the corresponding circulating variant
592 proportion using guidance from the World Health Organization (WHO)/European Centre for
593 Disease Prevention and Control (ECDC)^{7,8}, Brito et al.⁵, and Wohl et al.⁶ (Table 1) under
594 different genomic surveillance strategies with varying sampling coverage (i.e. all collected
595 specimens from all healthcare facilities are sent to one facility to be sampled for sequencing
596 (*population-wide* strategy); only *one*, 10%, 25%, 50%, or 100% of tertiary sentinel facilities
597 would sample the specimens they collected for sequencing). Turnaround time (i.e. time from
598 specimen collection to acquisition of sequencing data) was assumed to be negligible. 1,000
599 random independent simulations were performed for each guidance/surveillance strategy. We
600 simulated epidemics for (A) Wild-type SARS-CoV-2/Alpha. (B) Delta/Omicron. Grey regions
601 denote that we could not reliably detect the variant virus with 95% confidence using the
602 guidance in question under the assumed genomic surveillance strategy.
603

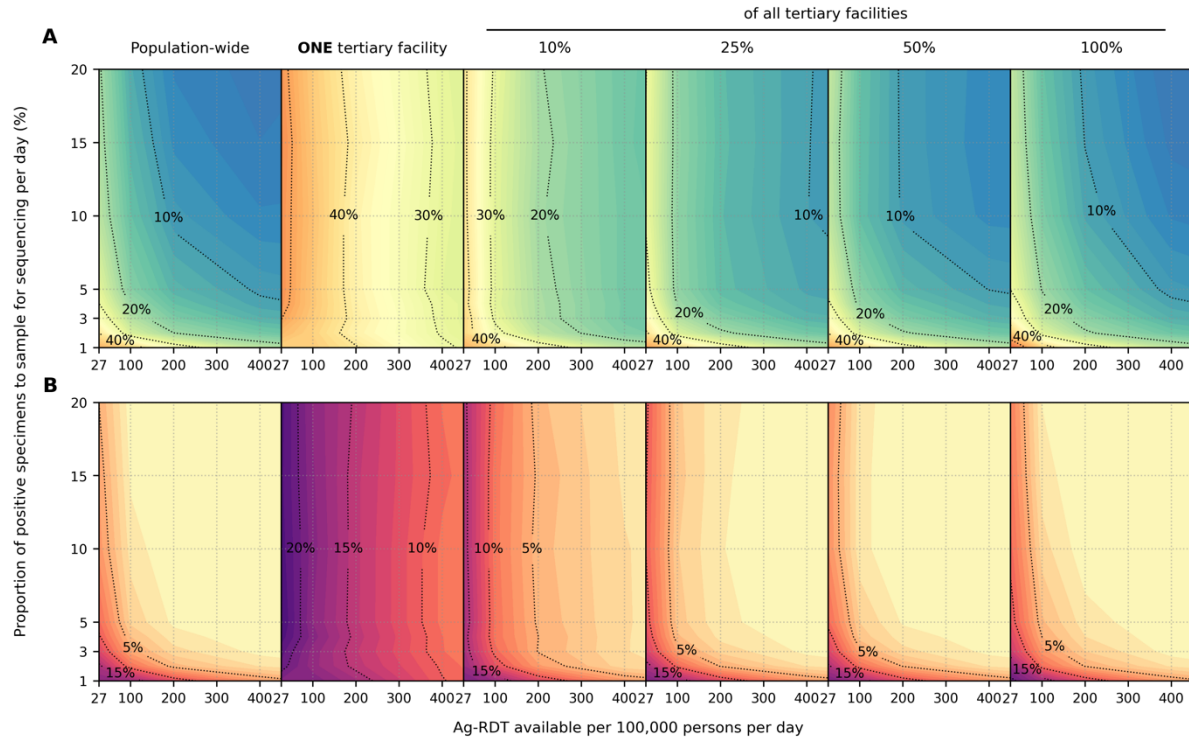


604

605 **Fig. 3. Impact of SARS-CoV-2 testing rates and proportion of positive specimens to**
606 **sequence on variant detection.** For each mean daily test availability (line and shading color),
607 the expected day when the first variant specimen to be sequenced is sampled since its
608 introduction is plotted against the proportion of positive specimens to be sampled for sequencing
609 daily. Different genomic surveillance strategies with varying sampling coverage (i.e. all
610 specimens collected from all healthcare facilities sent to one facility to be sampled for
611 sequencing (population-wide strategy); only one, 10%, 25%, 50%, or 100% of tertiary sentinel
612 facilities would sample the specimens they collected for sequencing) were simulated. (A) Wild-
613 type SARS-CoV-2/Alpha. (B) Delta/Omicron. The expected day when the first variant specimen
614 is sampled was computed from 1,000 random independent simulations for each surveillance
615 strategy. The shaded region depicts the standard deviation across simulations.

616

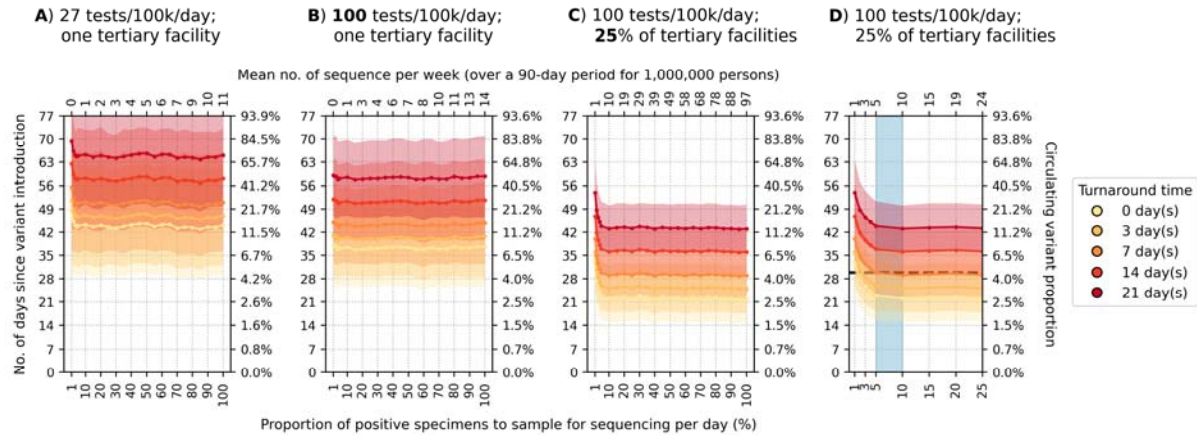
617



618

619 **Fig. 4. Impact of SARS-CoV-2 testing rates on the capacity to monitor changes in variant**
620 **prevalence based on diagnostic test availability and proportion of test-positive samples**
621 **sequenced.** Different genomic surveillance strategies (i.e. all specimens collected from all
622 healthcare facilities sent to one facility to be sampled for sequencing (*population-wide* strategy);
623 only *one*, 10%, 25%, 50%, or 100% of tertiary sentinel facilities would sample the specimens
624 they collected for sequencing) were simulated. (A) Maximum absolute difference between
625 observed and circulating variant proportions. (B) Proportion of timepoints when sequencing was
626 performed that the absolute difference between observed and circulating variant proportions is
627 greater than 20%. All results were computed from 1,000 random independent simulations for
628 each surveillance strategy.

It is made available under a [CC-BY-NC-ND 4.0 International license](https://creativecommons.org/licenses/by-nc-nd/4.0/) .



629

630 **Fig. 5. Recommended approach to enhance genomic surveillance robustness.** In each plot,
631 the operating curves of the expected day when the first Alpha variant sequence is generated are
632 plotted for different proportion of specimens to sample for sequencing per day and turnaround
633 times. We assumed that the Alpha variant was circulating at 1% initially with wild-type SARS-
634 CoV-2 in the background. We also assumed that positive specimens sampled within each week
635 for sequencing are consolidated into a batch before they are referred for sequencing. Turnaround
636 time refers to the time between collection of each weekly consolidated batch of positive
637 specimens to the acquisition of its corresponding sequencing data. The vertical axes denote the
638 number of days passed since the introduction of the Alpha variant (left) and its corresponding
639 circulating proportion (right). The horizontal axes denote the proportion of positive specimens to
640 sample for sequencing per day (bottom) and the corresponding mean number of sequences to be
641 generated per week per 1,000,000 people over a 90-day epidemic period. (A) Specimen pools for
642 sequencing from *one* tertiary sentinel facility with testing rate at 27 tests per 100,000 persons per
643 day (tests/100k/day). (B) Specimen pools for sequencing from *one* tertiary sentinel facility with
644 testing rate at 100 tests/100k/day. (C) Specimen pools for sequencing from 25% of all tertiary
645 sentinel facilities with testing rate at 100 tests/100k/day. (D) Zoomed-in plot of (C) to highlight
646 sequencing proportions varying between 1-25%. Sequencing 5-10% of positive specimens (blue
647 shaded region) would ensure that we would expectedly detect Alpha in 30 days if turnaround
648 time is kept within one week. All results were computed from 1,000 random independent
649 simulations for each surveillance strategy. The shaded region depicts the standard deviation
650 across simulations.
651

652 **Tables**

653 **Table 1. Current guidance by various stakeholder and academic groups on the number of**
 654 **specimens to sequence for detection of novel variants at low prevalence.**

	Recommendation on number/proportion of positive cases to sequence		Critical considerations
World Health Organization (WHO)/European Centre for Disease Prevention and Control (ECDC) ^{7,8}	<i>No. of positive cases</i>	<i>No. of sequences to detect at 1% with 95% confidence</i>	<ul style="list-style-type: none"> • Agnostic to variant properties • Assumes specimen pool to be sampled for sequencing is representative of circulating diversity but acknowledges that unless testing coverage is evenly distributed this will be a biased sample • Notes that in countries with limited sequencing capacity monitoring relative prevalence of variants should be prioritized
	<1000 cases	141	
	1001 – 2,500	196	
	2,500 – 5,000	243	
	5,001 – 10,000	270	
>10,000	285		
Brito et al., 2021 ⁵	At least 0.5% of all cases with a turnaround time of 21 days to detect novel lineage before it reaches 100 cases at 20% probability		<ul style="list-style-type: none"> • Based on sequencing data from Denmark which is testing at an average of >2,000 tests per 100,000 persons per day⁹
Wohl et al., 2022 ⁶	1-29 sequences per day to detect an Alpha-like variant based on 0.03% initial introduction for a population of 10,000 (assuming growth rate of 0.1/day) at 1% with 95% confidence. We used the spreadsheet (https://github.com/HopkinsIDD/VOCsamplesize) provided and input appropriate parameters to obtain the recommendation relevant to the simulated epidemics.		<ul style="list-style-type: none"> • Assumes that the <i>observed</i> variant proportion in the positive specimens collected is representative of the <i>circulating</i> variant proportions among the infected population. This requires a large number of specimens that are randomly collected for assumption to hold true at low circulating variant proportions • A correction factor is included to correct for biases in the <i>observed</i> variant proportion but only pertaining to those arising from the relative differences in diagnostic sensitivity, sample qualities and conditional asymptomatic and symptomatic testing probabilities between the two circulating variants.

655