

Using viral load and epidemic dynamics to optimize pooled testing in resource constrained settings

Authors: Brian Cleary^{1,*†}, James A. Hay^{2,3,*†}, Brendan Blumenstiel¹, Maegan Harden¹, Michelle Cipicchio¹, Jon Bezney¹, Brooke Simonton¹, David Hong⁴, Madikay Senghore², Abdul K. Sesay⁵, Stacey Gabriel¹, Aviv Regev^{6,7,8,9,*}, Michael J. Mina^{1,2,3,10,*}

Affiliations:

¹ Broad Institute of MIT and Harvard, Cambridge, MA 02142 USA

² Centre for Communicable Disease Dynamics, Department of Epidemiology, Harvard School of Public Health, Boston, MA, USA

³ Department of Immunology and Infectious Diseases, Harvard School of Public Health

⁴ Wharton Statistics, University of Pennsylvania, Philadelphia, PA, USA

⁵ Medical Research Council Unit The Gambia at London School of Hygiene and Tropical Medicine, PO Box 273, Banjul, The Gambia

⁶ Klarman Cell Observatory, Broad Institute of MIT and Harvard

⁷ Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142 USA

⁸ Howard Hughes Medical Institute, Chevy Chase, MD, USA

⁹ Current address: Genentech, 1 DNA Way, South San Francisco, CA, USA

¹⁰ Department of Pathology, Brigham and Women's Hospital, Harvard Medical School

†These authors contributed equally to this work

*Correspondence should be addressed to: bcleary@broadinstitute.org (B.C.), jhay@hsph.harvard.edu (J.H.), aregev@broadinstitute.org (A.R.), mmina@hsph.harvard.edu (M.M.)

23 **Abstract**

24 Extensive virological testing is central to SARS-CoV-2 containment, but many settings face
25 severe limitations on testing. Group testing offers a way to increase throughput by testing
26 pools of combined samples; however, most proposed designs have not yet addressed key
27 concerns over sensitivity loss and implementation feasibility. Here, we combine a
28 mathematical model of epidemic spread and empirically derived viral kinetics for SARS-CoV-
29 2 infections to identify pooling designs that are robust to changes in prevalence, and to ratify
30 losses in sensitivity against the time course of individual infections. Using this framework, we
31 show that prevalence can be accurately estimated across four orders of magnitude using only
32 a few dozen pooled tests without the need for individual identification. We then exhaustively
33 evaluate the ability of different pooling designs to maximize the number of detected infections
34 under various resource constraints, finding that simple pooling can identify up to 20 times as
35 many positives compared to individual testing with a given budget. We illustrate how pooling
36 affects sensitivity and overall detection capacity during an epidemic and on each day post
37 infection, finding that sensitivity loss is mainly attributed to individuals sampled at the end of
38 infection. Crucially, we confirm that our theoretical results can be accurately translated into
39 practice using pooled human nasopharyngeal specimens. Our results show that accounting
40 for variation in sampled viral loads provides a nuanced picture of how pooling affects sensitivity
41 to detect epidemiologically relevant infections. Using simple, practical group testing designs
42 can vastly increase surveillance capabilities in resource-limited settings.

43 Introduction

44 The ongoing pandemic of SARS-CoV-2, a novel coronavirus, has caused over 24 million
45 reported cases of coronavirus disease 2019 (COVID-19) and 800,000 reported deaths
46 between December 2019 and August 2020. (1) Although wide-spread virological testing is
47 essential to inform disease status and where outbreak mitigation measures should be targeted
48 or lifted, sufficient testing of populations with meaningful coverage has proven difficult. (2–7)
49 Disruptions in the global supply chains for testing reagents and supplies, as well as on-the-
50 ground limitations in testing throughput and financial support, restrict the usefulness of testing—
51 both for identifying infected individuals and to measure community prevalence and epidemic
52 trajectory. While these issues have been at the fore in even the highest-income countries, the
53 situation is even more dire in low income regions of the world. Cost barriers alone mean it is
54 often simply not practical to prioritize community testing in any useful way, with the limited
55 testing that exists necessarily reserved for the healthcare setting. These limitations urge new,
56 more efficient, approaches to testing to be developed and adopted both for individual
57 diagnostics and to enable public health epidemic control and containment efforts.

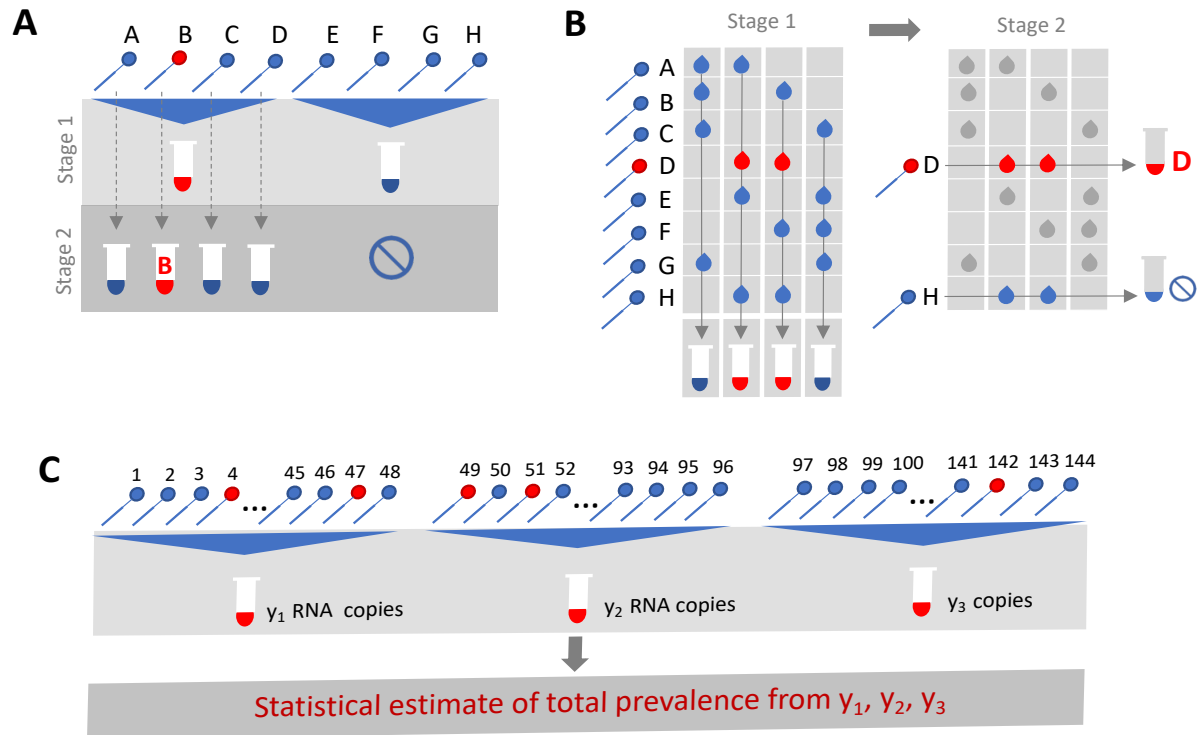
58 Group or pooled testing offers a way to increase efficiency by combining samples into a
59 group or pool and testing a small number of pools rather than all samples individually. (8–10)
60 For classifying individual samples, including for diagnostic testing, the principle is simple: if a
61 pool tests negative, then all of the constituent samples are assumed negative. If a pool tests
62 positive, then the constituent samples are putatively positive and must be tested again
63 individually (**Fig. 1A**). Further efficiency gains are possible through combinatorial pooling,
64 where, instead of testing every sample in every positive pool, each sample can instead be
65 represented across multiple pools and potential positives are identified based on the pattern
66 of pooled results (**Fig. 1B**). (9,10)

67 Simple pooling designs can also be used to assess prevalence without individual specimen
68 identification (**Fig. 1C**). It has already been shown that the frequency of positive pools can
69 allow estimation of the overall prevalence. (11) Crucially however, we show here that
70 prevalence estimates can be greatly honed by considering quantitative viral loads measured
71 in each positive pool, rather than simply using binary results (positive / negative). In short, the
72 viral (RNA) load measurement from a pool is proportional to the sum of the (diluted) viral loads
73 from each positive sample in the pool. Thus, here we show how evaluating the viral loads
74 greatly improves potential efficiency gains in prevalence estimates by providing crucial
75 information on the estimated number of positive samples in the pool – when the expected
76 distribution of viral loads across specimens is known, which is easily measured empirically in
77 a given lab. (12,13) Although this approach requires more complex statistical methods, the
78 efficiency gains for public health surveillance can be large, and simplifying templates can be
79 produced to improve ease of use and access to these types of analyses. The outcome is a
80 highly efficient method for estimating population prevalence and enabling robust public health
81 surveillance where it was previously out of reach.

82 Whilst the literature on theoretically optimized pooling designs for COVID-19 testing has
83 grown rapidly, formal incorporation of biological variation (*i.e.*, viral loads) and incorporation
84 of general position along the epidemic curve, has received little attention. (14–17) Test
85 sensitivity for example is not a fixed value, but depends on viral load, which can vary by
86 many orders of magnitudes across individuals and over the course of an infection. (18–20)
87 This large variation within a single infection affects sensitivity to detect infections at different
88 points in the disease course, which has implications for appropriate intervention and the
89 interpretation of a viral load measurement from a sample pool.

90 Here, we comprehensively evaluate designs for pooled testing of SARS-CoV-2 whilst
91 accounting for epidemic dynamics and variation in viral loads arising from viral kinetics and
92 extraneous features such as sampling variation. We demonstrate efficient, logistically feasible

93 pooling designs for individual identification (*i.e.*, diagnostics) and prevalence estimation (*i.e.*,
94 population surveillance). To do this, we use realistic simulated viral load data at the individual
95 level over time, representing the entire time course of an epidemic to generate synthetic data
96 that reflects the true distribution of viral loads in the population at any given time of the
97 epidemic. We then used these data to derive optimal pooling strategies for different use cases
98 and resource constraints *in-silico*. Finally, we demonstrate the approach using discarded de-
99 identified human nasopharyngeal swabs initially collected for diagnostic and surveillance
100 purposes.



101

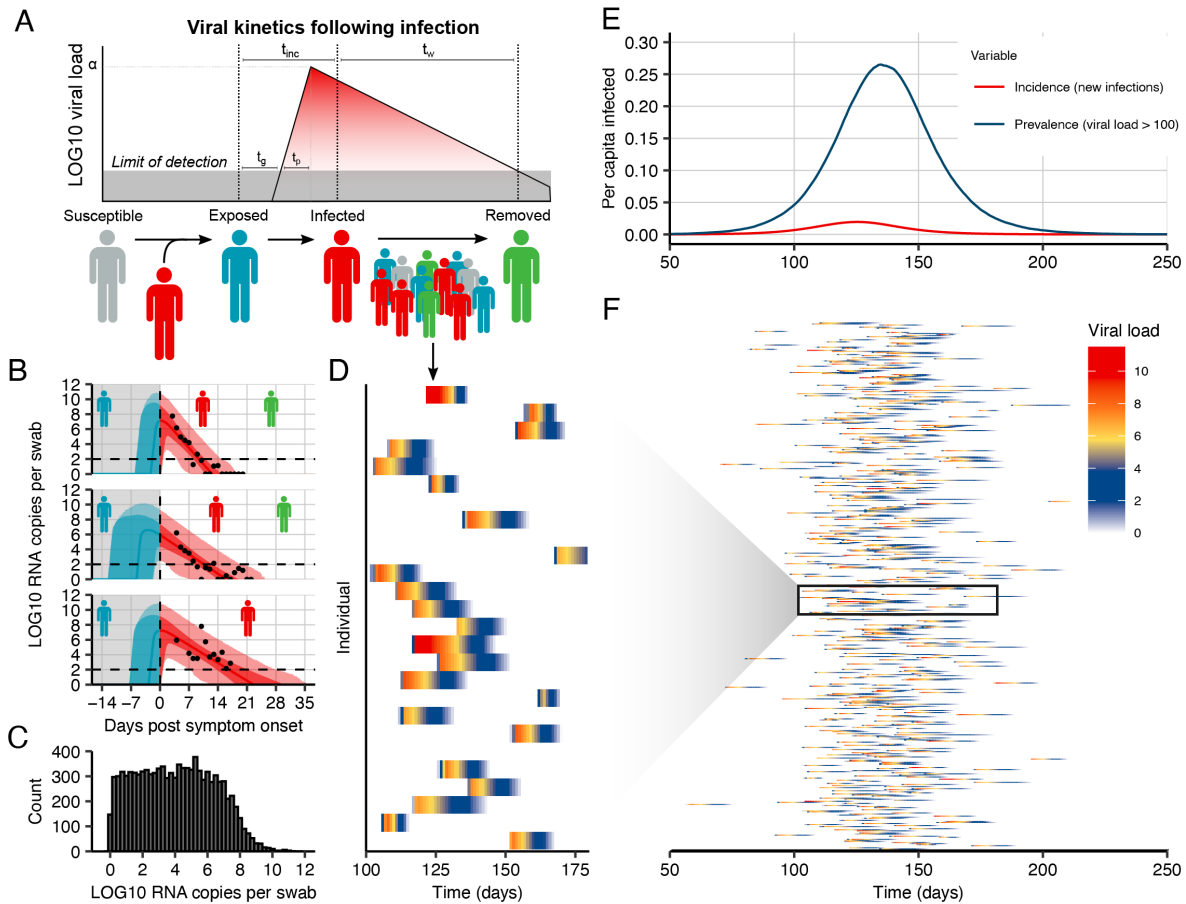
102 **Fig. 1.** Group testing designs for sample identification or prevalence estimation. In group
 103 testing, multiple samples are pooled and tests are run on one or more pools. The results of
 104 these tests can be used for identification of positive samples (**A, B**) or to estimate prevalence
 105 (**C**). (**A**) In the simplest design for sample identification, samples are partitioned into non-
 106 overlapping pools. In stage 1 of testing, a negative result (Pool 2) indicates each sample in
 107 that pool was negative, while a positive result (Pool 1) indicates at least one sample in the
 108 pool was positive. These putatively positive samples are subsequently individually tested in
 109 stage 2 to identify positive results. (**B**) In a combinatorial design, samples are included in
 110 multiple pools as shown in stage 1. All samples that were included in negative pools are
 111 identified as negative, and the remaining putatively positive samples that were not included in
 112 any negative test are tested individually in stage 2. (**C**) In prevalence estimation, samples are
 113 partitioned into pools. The pool measurement will depend on the number and viral load of
 114 positive samples, and the dilution factor. The (quantitative) results from each pool can be used

115 to estimate the fraction of samples that would have tested positive, had they been tested
116 individually.

117 **Results**

118 **Modelling a synthetic population to assess pooling designs**

119 To identify optimal pooling strategies for distinct scenarios, we required realistic estimates of
120 viral loads across epidemic trajectories. We developed a population-level mathematical model
121 of SARS-CoV-2 transmission that incorporates empirically measured within-host virus
122 kinetics, and used these simulations to generate population-level viral load distributions
123 representing real data sampled from population surveillance, either using nasopharyngeal
124 swab or sputum samples (**Fig. 2**). Full details are provided in **Materials and Methods** and
125 **Supplementary Material 1, sections 1-4**. These simulations generated a synthetic, realistic
126 epidemic with a peak daily per incidence of 19.5 per 1000 people, and peak daily prevalence
127 of RNA positivity (viral load greater than 100 virus RNA copies per ml) of 265 per 1000 (**Fig.**
128 **2E**). We used these simulation data to evaluate optimal group testing strategies at different
129 points along the epidemic curve for diagnostic as well as public health surveillance, where the
130 true viral loads in the population is known fully.



131

132

Fig. 2: Viral kinetics model fits, simulated infection dynamics and population-wide viral

133

load kinetics. (A) Schematic of the viral kinetics and infection model. Individuals begin

134

susceptible with no viral load, acquire the virus from another infectious individual (exposed),

135

experience an increase in viral load and possibly develop symptoms (infected), and finally

136

either recover following viral waning or die (removed). This process is simulated for many

137

individuals. **(B)** Model fits to time-varying viral loads in swab samples. The black dots show

138

observed *log*₁₀ RNA copies per swab; solid lines show posterior median estimates; dark

139

shaded regions show 95% credible intervals (CI) on model-predicted latent viral loads; light

140

shaded regions show 95% CI on simulated viral loads with added observation noise. The blue

141

region shows viral loads before symptom onset and red region shows time after symptom

142

onset. The horizontal dashed line shows the limit of detection. **(C)** Distribution of positive viral

143

loads from 10,000 individuals sampled at day 140. **(D)** 25 simulated viral loads over time. The

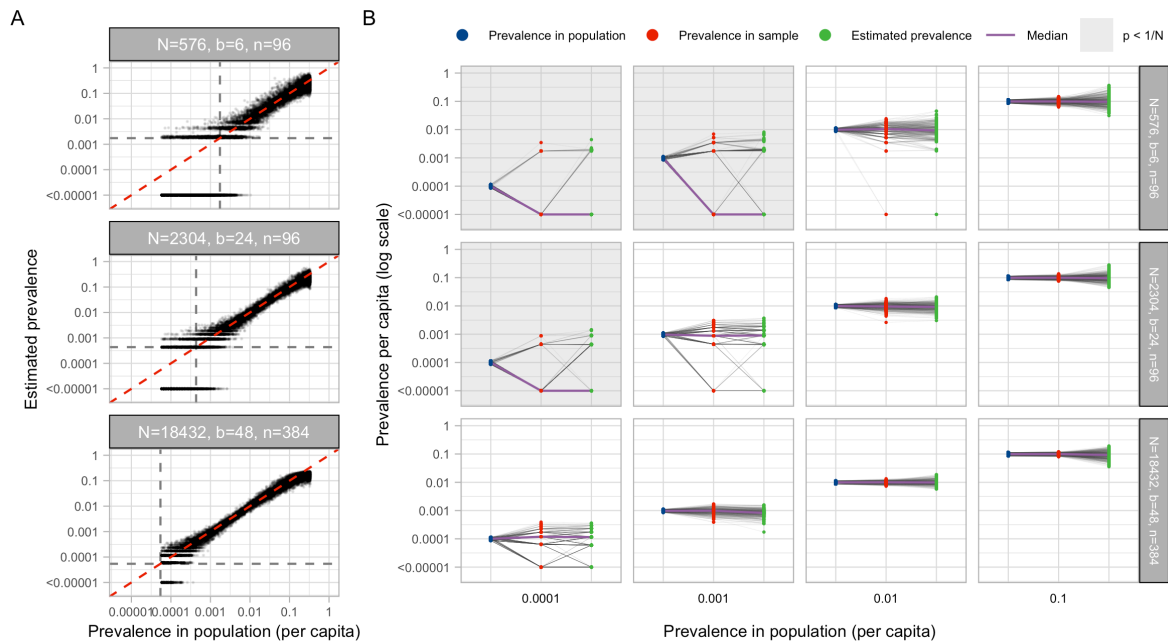
144 heatmap shows the viral load in each individual over time. **(E)** Simulated infection incidence
145 and prevalence of virologically positive individuals from the SEIR model. Incidence was
146 defined as the number of new infections per day divided by the population size. Prevalence
147 was defined as the number of individuals with viral load > 100 (\log_{10} viral load > 2) in the
148 population divided by the population size on a given day. **(F)** As in **(D)**, but for 500 individuals.
149 The distribution of viral loads reflects the increase and subsequent decline of prevalence. We
150 simulated from inferred distributions for the viral load parameters, thereby propagating
151 substantial individual-level variability through the simulations.

152 Improved testing efficiency for estimating prevalence

153 We developed a statistical method to estimate prevalence of SARS-CoV-2 based on cycle
154 threshold (Ct) values measured from pooled samples (**Materials and Methods**), potentially
155 using far fewer tests than would be required to assess prevalence based on number of
156 positive samples identified. We used our synthetic viral load data to assess inferential
157 accuracy under a range of sample availabilities and pooling designs. Across the spectrum of
158 simulated pools and tests we found that simple pooling allows accurate estimates of
159 prevalence across at least four orders of magnitude, ranging from 0.02% to 20%, with up to
160 400-times efficiency gains (*i.e.*, 400 times fewer tests) than would be needed without pooling
161 (**Fig. 3**). For example, in a population prevalence study that collects ~2,000 samples, we
162 accurately estimated infection prevalences as low as 0.05% by using only 24 total qPCR
163 tests (*i.e.*, 24 pools of 96 samples each; **Fig. 3A**; **Fig. S1**). Importantly, because the
164 distribution of Ct values may differ depending on the sample type (sputum vs. swab), the
165 instrument, and the phase of the epidemic (growth vs. decline, **Fig. S2**), in practice, the
166 method should be calibrated to viral load data (*i.e.*, Ct values) specific to the laboratory and
167 instrument (which can differ from one laboratory to the next) and the population under
168 investigation.

169 Estimation error arises in two stages: sample collection effects, and as part of the inference
170 method (**Fig. 3B**). Error from sampling collection became less important with increasing
171 numbers of positive samples, which occurred with increasing population prevalence or by
172 increasing the total number of tested samples (**Fig. 3B**; **Fig. S2**). At very low prevalence,
173 small sample sizes (N) risk missing positives altogether or becoming biased by false
174 positives. We found that accuracy in prevalence estimation is greatest when population
175 prevalence is greater than $1/N$ and that when this condition was met, partitioning samples
176 into more pools always improves accuracy (**Fig. S2**). In summary, very accurate estimates of

177 prevalence can be attained using only a small fraction of the tests that would be needed in
178 the absence of pooling.



179

180 **Fig. 3:** Estimating prevalence from a small number of pooled tests. In prevalence estimation,
 181 a total of N individuals are sampled and partitioned into b pools (with $n=N/b$ samples per pool).
 182 The true prevalence in the entire population (x -axis in A) varies over time with epidemic
 183 spread. Population prevalences shown here are during the epidemic growth phase. **(A)**
 184 Estimated prevalence against true population prevalence using 100 independent trials
 185 sampling N individuals at each day of the epidemic. Each facet shows a different pooling
 186 design (more pooling designs shown in **Fig. S1**). Dashed grey lines show one divided by the
 187 sample size, N . **(B)** For a given true prevalence (x -axis, blue points), estimation error is
 188 introduced both through binomial sampling of positive samples (red points) and inference on
 189 the sampled viral loads (green points). Sampling variation is a bigger contributor at low
 190 prevalence and low sample sizes. When prevalence is less than one divided by N (grey
 191 boxes), inference is less accurate due to the high probability of sampling only negative
 192 individuals or inclusion of false positives.

193 Pooled testing for individual identification

194 We next analyzed effectiveness of group testing for identifying individual sample results at
195 different points along the epidemic curve with the aim of identifying simple, efficient pooling
196 strategies that are robust to a range of prevalences (**Fig. 1A&B**). Using the simulated viral
197 load data described in **Materials and Methods**, we evaluated a large array of pooling designs
198 *in silico* (**Table S1**). Based on our models of viral kinetics and given a PCR limit of detection
199 of 100 viral copies per ml, we first estimated a baseline sensitivity of conventional (non-pooled)
200 PCR testing of 85% during the epidemic growth phase (*i.e.*, 15% of the time we sample an
201 infected individual with a viral load greater than 1 but below the LOD of 100 viral copies per
202 mL, **Fig. 4A**), which largely agrees with reported estimates. (21,22) This reflects sampling
203 during the latent period of the virus (after infection but prior to significant viral growth) or in the
204 relatively long duration of low viral titers during viral clearance.

205 Sensitivity of pooled tests, relative to individual testing, is affected by the dilution factor of
206 pooling and by the population prevalence – with lower prevalence resulting in generally
207 lower sensitivity as positives are diluted into many negatives (**Fig. 4A**). The decrease in
208 sensitivity is roughly linear with the log of the dilution factor employed, which largely depends
209 on the number and size of the pools and, for combinatorial pooling, the number of pools that
210 each sample is placed into (**Fig. S3A-C**).

211 There is a less intuitive relationship between sensitivity and prevalence as it changes over the
212 course of the epidemic. Early in an epidemic there is an initial dip in sensitivity for both
213 individual and pooled testing (**Fig. 4A**). Early during exponential growth of an outbreak, a
214 random sample of infected individuals will be sampled closer to their peak viral load, while
215 later on there is an increasing mixture of newly infected with individuals with lower viral loads
216 at the tail end of their infection. We found that this means at peak prevalence, sensitivity of
217 pooled testing increases as samples with lower viral loads, which would otherwise be missed

218 due to dilution, are more likely to be ‘rescued’ by coexisting in the same pool with high viral
219 load samples and thus get individually retested (at their undiluted concentration) during the
220 validation stage. During epidemic decline, fewer new infections arise over time and therefore
221 a randomly selected infected individual is more likely to be sampled during the recovery phase
222 of their infection, when viral loads are lower (**Fig. S4D**). Overall sensitivity is therefore lower
223 during epidemic decline, as more infected individuals have viral loads below the limit of
224 detection; during epidemic growth (up to day 108), overall sensitivity of RT-PCR for individual
225 testing is 85%, whereas during epidemic decline (from day 168 onward) it is 60% (**Fig. S5A**).
226 Sensitivity of RT-PCR for individual testing was ~75% across the whole epidemic. We note
227 that in practice, sensitivity is likely higher than estimated here, because individuals are not
228 sampled entirely at random. Together, these results describe how sensitivity is affected by the
229 combination of epidemic dynamics, viral kinetics, and pooling design when individuals are
230 sampled randomly from the population.

231 We find that on average the majority of false negatives arise from individuals sampled seven
232 days or more after their peak viral loads, or around seven days after what is normally
233 considered symptom onset (~75% in swab samples during epidemic growth; ~96% in swab
234 samples during epidemic decline; ~68% in sputum during epidemic growth). Importantly, only
235 ~3% of false negative swab samples arose from individuals tested during the first week
236 following peak viral load during epidemic growth, and only ~1% during epidemic decline
237 – (peak titers usually coincide with symptom onset) – and thus most false negatives are from
238 individuals with the least risk of onward transmission (**Fig. S3D&E**).

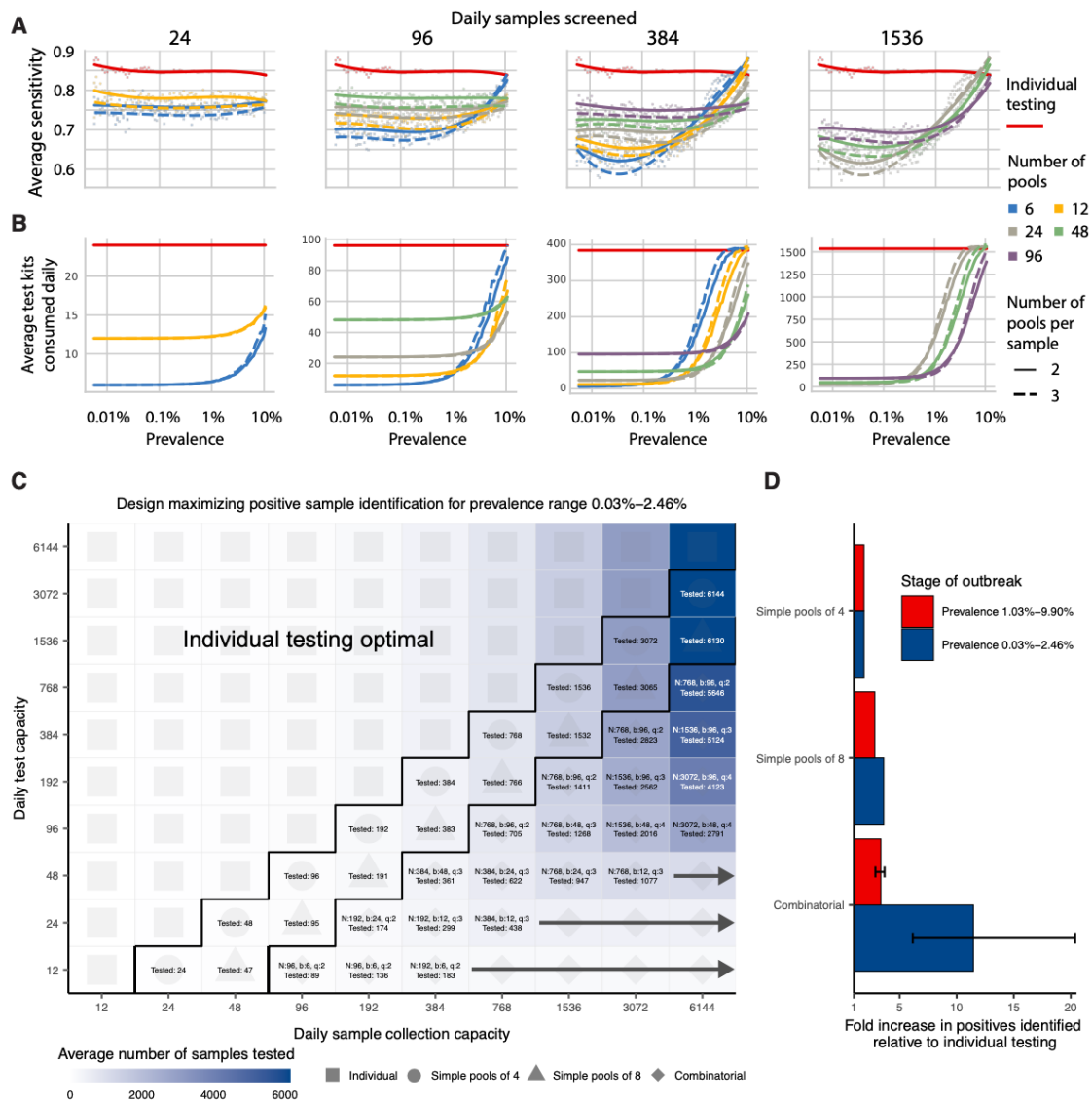
239 As mentioned above, the lower sensitivity of pooled testing is counterbalanced by gains in
240 efficiency. When prevalence is low, efficiency is roughly the number of samples divided by
241 the number of pools, since there are rarely putative positives to test individually. However,
242 the number of validation tests required will increase as prevalence increases, and designs
243 that are initially more efficient will lose efficiency (**Fig. 4B**). In general, we find that at very

244 low population prevalence the use of fewer pools each with larger numbers of specimens
245 offers relative efficiency gains compared to larger numbers of pools, as the majority of pools
246 will test negative. However, as prevalence increases, testing a greater number of smaller
247 pools pays off as more validations will be performed on fewer samples overall (**Fig. 4B**). For
248 combinatorial designs with a given number of total samples and pools, splitting each sample
249 across fewer pools results in a modest efficiency gains (dashed versus solid lines in **Fig.**
250 **4B**).

251 To address realistic resource constraints, we integrated our analyses of sensitivity and
252 efficiency with limits on daily sample collection and testing capacity to maximize the number
253 of positive individuals identified (see **Materials and Methods**). We analyzed the total number
254 of samples screened and the fold increase in the number of positive samples identified relative
255 to individual testing for a wide array of pooling designs evaluated over a period of 50 days
256 during epidemic spread (days 40-90 where point prevalence reaches ~2.5%; **Fig. 4C&D**).
257 Because prevalence changes over time, the number of validation tests may vary each day
258 despite constant pooling strategies. Thus, tests saved on days requiring fewer validation tests
259 can be stored for days where more validation tests are required.

260 Across all resource constraints considered, we found that effectiveness ranged from one
261 (when testing every sample individually is optimal) to 20 (*i.e.*, identifying 20x more positive
262 samples on a daily basis compared with individual testing within the same budget; **Fig. 4D**).
263 As expected, when capacity to collect samples exceeds capacity to test, group testing
264 becomes increasingly effective. Simple pooling designs are most effective when samples are
265 in slight excess of testing capacity (2-8x), whereas we find that increasingly complex
266 combinatorial designs become the most effective when the number of samples to be tested
267 greatly exceeds testing capacity. Additionally, when prevalence is higher (*i.e.*, sample
268 prevalence from 1.03% to 9.90%), the optimal pooling designs shift towards combinatorial
269 pooling, and the overall effectiveness decreases – but still remains up to 4x more effective

270 than individual testing (**Fig. S6**). Our results were qualitatively unchanged when evaluating
271 the effectiveness of pooling sputum samples, and the optimal pooling designs under each
272 set of sample constraints were either the same or very similar (**Fig. S7**). Furthermore, we
273 evaluated the same strategies during a 50-day window of epidemic decline (days 190-250)
274 and found that similar pooling strategies were optimally effective, despite lower overall
275 sensitivity as described above (**Fig. S5**).



276

277 **Fig. 4:** Group testing for sample identification. We evaluated a variety of group testing designs
 278 for sample identification (**Table S1**) on the basis of sensitivity (**A**), efficiency (**B**), total number
 279 of positive samples identified (**C**) and the fold increase in positive samples identified relative
 280 to individual testing (**D**). (**A** and **B**) The average sensitivity (**A**, y-axis, individual points and
 281 spline) and average number of tests needed to identify individual positive samples (**B**, y-axis)
 282 using different pooling designs (individual lines) were measured over days 20-110 in our
 283 simulated population, with results plotted against prevalence (x-axis, log-scale). Results show
 284 the average of 200,000 trials, with individuals selected at random on each day in each trial.

285 Pooling designs are separated by the number of samples tested on a daily basis (individual
286 panels); the number of pools (color); and the number of pools into which each sample is split
287 (dashed versus solid line). Solid red line indicates results for individual testing. **(C)** Every
288 design was evaluated under constraints on the maximum number of samples collected
289 (columns) and average number of reactions that can be run on a daily basis (rows) over days
290 40-90. Text in each box indicates the optimal design for a given set of constraints (number of
291 samples per batch (N), number of pools (b), number of pools into which each sample is split
292 (q), average number of total samples screened per day). Color indicates the average number
293 of samples screened on a daily basis using the optimal design. Arrows indicate that the same
294 pooling design is optimal at higher sample collection capacities due to testing constraints. **(D)**
295 Fold increase in the number of positive samples identified relative to individual testing with the
296 same resource constraints. Error bar shows range amongst optimal designs.

297 **Pilot and validation experiments**

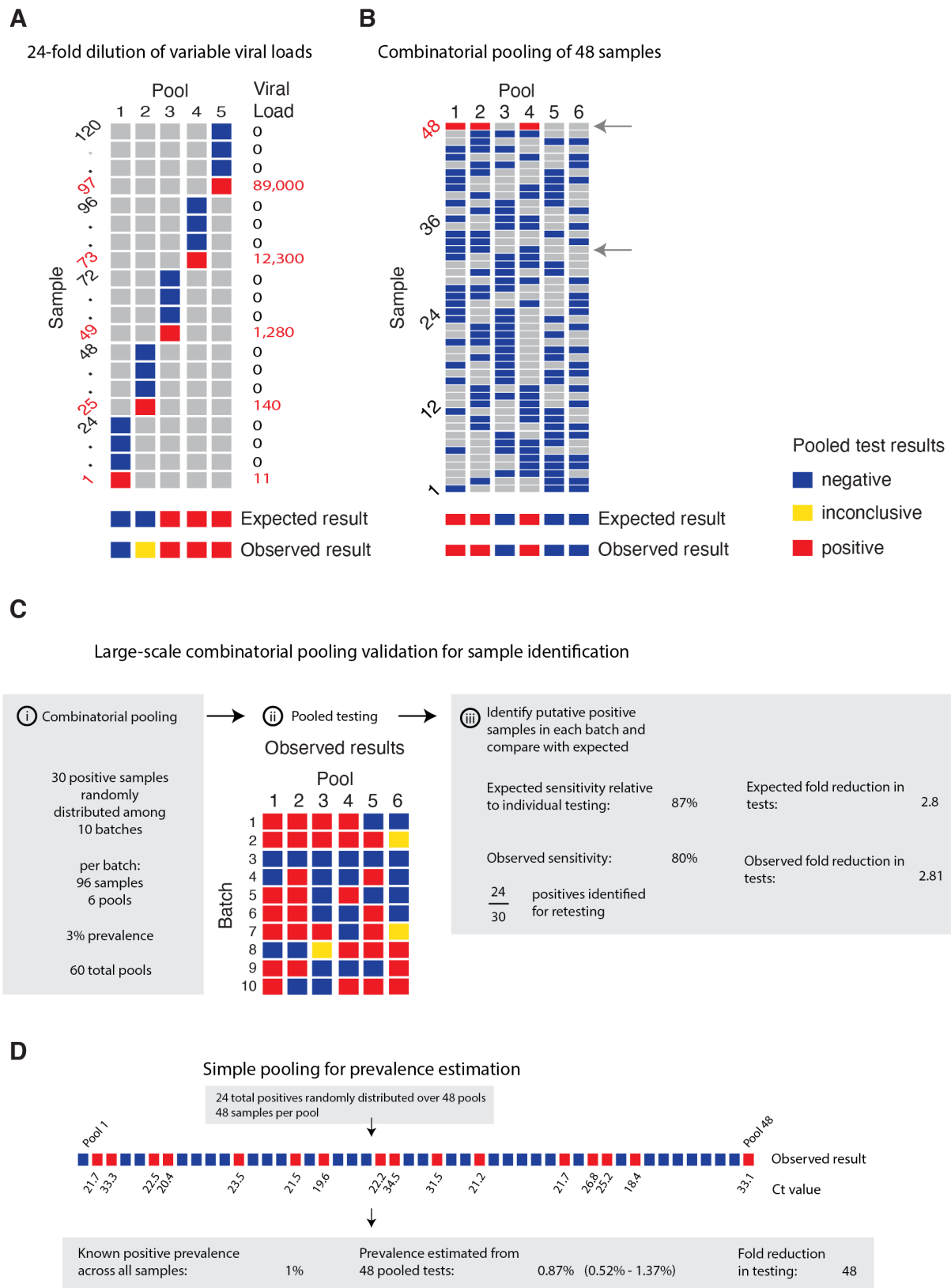
298 We validated our pooling strategies using anonymized clinical nasopharyngeal swab
299 specimens. To evaluate simple pooling across a range of inputs, we diluted 5
300 nasopharyngeal clinical swab samples with viral loads of 89000, 12300, 1280, 140 and 11
301 viral copies per ml, respectively, into 23 negative nasopharyngeal swab samples (pools of
302 24). Further details are provided in **Materials and Methods**, and **Supplementary Material**
303 **1, sections 7&8**. The results matched the simulated sampling results: the first three pools
304 were all positive, the fourth was inconclusive (negative on N1, positive on N2), and the
305 remaining pool was negative (**Fig. 5A, Table S2**). These results are as expected because
306 the EUA approved assay used has a limit of detection of ~100 virus copies per ml, such that
307 the last two specimens fall below the limit of detection given a dilution factor of 24 (i.e. 0.46
308 and 5.8 virus copies per ml once pooled).

309 We next tested combinatorial pooling, first using only a modest pooling design. We split 48
310 samples, including 1 positive, into 6 pools with each sample spread across three different
311 pools. The method correctly identified the three pools containing the positive specimen (**Fig.**
312 **5B, Table S2**). One negative sample was included in the same 3 pools as the positive
313 sample; thus, 8 total tests (6 pools + 2 validations) were needed to accurately identify the
314 status of all 48 samples, a 6x efficiency gain, which matched our expectations from the
315 simulations.

316 We next performed two larger validation studies (**Materials and Methods**, and
317 **Supplementary Material 1, section 8**). To validate combinatorial pooling, we used
318 anonymized samples representing 930 negative and 30 distinct positive specimens (3.1%
319 prevalence), split across 10 batches of 96 specimens each (**Table S3**). For each batch of 96,
320 we split the specimens into 6 pools and each specimen was spread across 2 pools (**Fig. 5C,**
321 **Table S4**). For this combinatorial pooling design and prevalence, our simulations suggest

322 that we would expect to identify 26 out of 30 known positives (87%) and would see a 2.81x
323 efficiency gain – using only 35% of the number of tests compared to no pooling. We
324 identified 24 of the 30 known positives (80%) and, indeed, required 35% fewer tests (341 vs
325 960, a 2.8x efficiency gain).

326 To further validate our methods for prevalence estimation, we created a large study
327 representing 2,304 samples with a (true) positive prevalence of 1%. We aimed to determine
328 how well our methods would work to estimate the true prevalence using 1/48th the number of
329 tests compared to testing samples individually. To do this, we randomly assigned 24 distinct
330 positive samples into 48 pools, with each pool containing 48 samples (**Table S3**; to create the
331 full set of pools, we treated some known negatives as distinct samples across separate pools).
332 By using the measured viral loads detected in each of the pools, our methods estimated a
333 prevalence of 0.87% (compared to the true prevalence of 1%) with a bootstrapped 95%
334 confidence interval of 0.52% - 1.37% (**Fig. 5D**), and did so using 48x fewer tests than without
335 pooling. This level of accuracy is in line with our expectations from our simulations. Notably,
336 the inference algorithm applied to these data used viral load distributions calibrated from our
337 simulated epidemic, which in turn had viral kinetics calibrated to samples collected and tested
338 on another continent, demonstrating robustness of the training procedure.



339

340 **Fig. 5:** Validation of simple and combinatorial pooling. Pooled testing of samples was validated

341 experimentally. (A) Five pools (columns of matrix), each consisting of 24 nasopharyngeal

342 swab samples (rows of matrix; 23 negative samples per pool and 1 positive, with viral load
343 indicated in red on right) were tested by viral extraction and RT+qPCR. Pooled results
344 indicated as: negative (blue), inconclusive (yellow), or positive (red). **(B)** Six combinatorial
345 pools (columns) of 48 samples (rows; 47 negative and 1 positive with viral load of 12,300)
346 were tested as above. Pools 1, 2, and 4 tested positive. Arrows indicate two samples that were
347 in pools 1, 2, and 4: sample 32 (negative), and sample 48 (positive). **(C)** Previously tested de-
348 identified samples were pooled using a combinatorial design with 96 samples, 6 pools, and 2
349 pools per sample. 30 positive samples were randomly distributed across 10 batches of the
350 design. Viral RNA extraction and RT-qPCR were performed on each pool, with the results
351 used to identify potentially positive samples. **(D)** Samples were pooled according to a simple
352 design (48 pools with 48 samples per pool). 24 positive samples were randomly distributed
353 among the pools (establishing a 1% prevalence). The pooled test results were used with an
354 MLE procedure to estimate prevalence (0.87%), and bootstrapping was used to estimate a
355 95% confidence interval (0.52% - 1.37%).

356 Discussion

357 Our results show that group testing for SARS-CoV-2 can be a highly effective tool to
358 increase surveillance coverage and capacity when resources are constrained. For
359 prevalence testing, we find that fewer than 40 tests can be used to accurately infer
360 prevalences across four orders of magnitude, providing large savings on tests required. For
361 individual identification, we determined an array of designs that optimize the rate at which
362 infected individuals are identified under constraints on sample collection and daily test
363 capacities. These results provide pooling designs that maximize the number of positive
364 individuals identified on a daily basis, while accounting for epidemic dynamics, viral kinetics,
365 viral loads measured from nasopharyngeal swabs or sputum, and practical considerations of
366 laboratory capacity.

367 While our experiments suggest that pooling designs may be beneficial for SARS-CoV-2
368 surveillance and identification of individual specimens, there are substantial logistical
369 challenges to implementing theoretically optimized pooling designs. Large-scale testing
370 without the use of pooling already requires managing thousands of specimens per day,
371 largely in series. Pooling adds complexity because samples must be tracked across multiple
372 pools and stored for potential re-testing. These complexities can be overcome with proper
373 tracking software (including simple spreadsheets) and standard operating procedures in
374 place before pooling begins. Such procedures can mitigate the risk of handling error or
375 specimen mix-up. In addition, expecting laboratories to regularly adapt their workflow and
376 optimize pool sizes based on prevalence may not be feasible in some settings. (8,16) A
377 potential solution is to follow a simple, fixed protocol that is robust to a range of prevalences.
378 **Supplementary Material 2** provides an example spreadsheet guiding a technician receiving
379 96 labeled samples to create 6 pools, enter the result of each pool and be provided a list of
380 putative positives to be retested.

381 Enhancing efficiency at the expense of sensitivity must be considered depending on the
382 purpose of testing. For prevalence testing, accurate estimates can be obtained using very
383 few tests if individual identification is not the aim. For individual testing, although identifying
384 all positive samples that are tested is of course the objective, increasing the number of
385 specimens tested when sacrificing sensitivity may be a crucially important tradeoff. This
386 tradeoff is particularly pertinent because the specimens most likely to be lost due to dilution
387 are those samples with the lowest viral loads already near the limit of detection (**Fig.**
388 **S3D&E**). Although there is a chance that the low viral load samples missed are on the
389 upswing of an infection – when identifying the individual would be maximally beneficial – the
390 asymmetric course of viral titers over the full duration of positivity means that most false
391 negatives would arise from failure to detect late-stage, low-titer individuals who are less
392 likely to be infectious. (20) Optimal strategies and expectations of sensitivity should also be
393 considered alongside the phase of the epidemic and how samples are collected, as this will
394 dictate the distribution of sampled viral loads. For example, if individuals are under a regular
395 testing regimen or are tested due to recent exposure or symptom onset, then viral loads at
396 the time of sampling will typically be higher, leading to higher sensitivity in spite of dilution
397 effects.

398 Testing throughput and staffing resources should also be considered. If a testing facility can
399 only run a limited number of tests per day, it may be preferable to process more samples at a
400 slight cost to sensitivity. Back-logs of individual testing can result in substantial delays in
401 returning individual test results, which can ultimately defeat the purpose of identifying
402 individuals for isolation - potentially further justifying some sensitivity losses. (23,24) Choosing
403 a pooling strategy will therefore depend on target population and availability of resources. For
404 testing in the community or in existing sentinel surveillance populations (e.g., antenatal
405 clinics), point prevalence is likely to be low (<0.1 - 3%), which may favor strategies with fewer
406 pools. (6,25–27) Conversely, secondary attack rates in contacts of index cases may vary from

407 <1% to 17% depending on the setting (e.g., casual vs. household contacts), (28–30) and may
408 be even higher in some instances, (31,32) favoring more pools. These high point prevalence
409 sub-populations may represent less efficient use cases for pooled testing, as our results
410 suggest that pooling for individual identification is inefficient once prevalence reaches 10%.
411 However, group testing may still be useful if testing capacity is severely limited – for example,
412 samples from all members of a household could be tested as a pool and quarantined if tested
413 positive, enabling faster turnaround than testing individuals. This approach may be even more
414 efficient if samples can be pooled at the point of collection, requiring no change to laboratory
415 protocols.

416 Our modelling results have a number of limitations and may be updated as more data
417 become available. First, our simulation results depend on the generalizability of the
418 simulated Ct values, which were based on viral load data from symptomatic patients.
419 Although some features of viral trajectories, such as viral waning, differ between
420 symptomatic and asymptomatic individuals, population-wide data suggest that the range of
421 Ct values do not differ based on symptom status. (20,33) Furthermore, we have assumed a
422 simple hinge function to describe viral kinetics. Different shapes for the viral kinetics
423 trajectory may become apparent as more data become available. Nonetheless, our
424 simulated population distribution of Ct values is comparable to existing data and we
425 propagated substantial uncertainty in viral kinetics parameters to generate a wide range of
426 viral trajectories. For prevalence estimation, the MLE framework requires training on a
427 distribution of Ct values. Such data can be available based on past tests from a given
428 laboratory, but care should be taken to use a distribution appropriate for the population
429 under consideration. For example, training the virus kinetics model on data skewed towards
430 lower viral loads (as would be observed during the tail end of an epidemic curve) may be
431 inappropriate when the true viral load distribution is skewed higher (as might be the case
432 during the growth phase of an epidemic curve). Nevertheless, we used our simulated

433 distribution of Ct, which were fit to virus kinetics in published reports from distinct labs from
434 across the world and obtained highly accurate results throughout. Thus, despite the
435 limitations just mentioned, this shows that the virus kinetics models are quite robust and may
436 not, in practice, require new fitting to the individual laboratory or population. In addition, while
437 we assume that individuals are sampled from the population at random in our analysis, in
438 practice samples that are processed together are also typically collected together, which
439 may bias the distribution of positive samples among pools.

440 We have shown that simple designs that are straightforward to implement have the potential
441 to greatly improve testing throughput across the time course of the pandemic. These principles
442 likely also hold for pooling of sera for antibody testing, which remains an avenue for future
443 work. There are logistical challenges and additional costs associated with pooling that we do
444 not consider deeply here, and it will therefore be up to laboratories and policy makers to decide
445 where these designs are feasible. Substantial coordination will therefore be necessary to make
446 group testing practical but investing in these efforts could enable community screening where
447 it is currently infeasible and provide epidemiological insights that are urgently needed.

448 **Materials and Methods**

449 **Simulation model of infection dynamics and viral load kinetics**

450 We developed a population-level mathematical model of SARS-CoV-2 transmission that
451 incorporates realistic within-host virus kinetics. Full details are provided in **Supplementary**
452 **Material 1, sections 1-4**, but we provide an overview here. First, we fit a viral kinetics model
453 to published longitudinally collected viral load data from nasopharyngeal swab and sputum
454 samples using a Bayesian hierarchical model that captures the variation of peak viral loads,
455 delays from infection to peak and virus decline rates across infected individuals (**Fig. 5A&B;**
456 **Fig. S8**). (19) By incorporating estimated biological variation in virus kinetics, this model allows
457 random draws each representing distinct within-host virus trajectories. We then simulated
458 infection prevalence during a SARS-CoV-2 outbreak using a deterministic Susceptible-
459 Exposed-Infected-Removed (SEIR) model with parameters reflecting the natural history of
460 SARS-CoV-2 (**Fig. 5D**). For each simulated infection, we generated longitudinal virus titers
461 over time by drawing from the distribution of fitted virus kinetic curves, using distributions
462 derived using either nasopharyngeal swab or sputum data (**Fig. S4**). All estimated and
463 assumed model parameters are shown in **Table S5**, with model fits shown in **Fig. S8**. Posterior
464 estimates and Markov chain Monte Carlo trace plots are shown in **Fig. S9** and **Fig. S10**. We
465 accounted for measurement variation by: i) transforming viral loads into Ct values under a
466 range of Ct calibration curves, ii) simulating false positives with 1% probability, and,
467 importantly, iii) simulating sampling variation. We assumed a limit of detection (LOD) of 100
468 RNA copies / ml.

469 **Estimating prevalence from pooled test results**

470 We adapted a statistical (maximum likelihood) framework initially developed to estimate HIV
471 prevalence with pooled antibody tests to estimate prevalence of SARS-CoV-2 using pooled
472 samples. (12,13) The framework accounts for the distribution of viral loads (and uncertainty

473 around them) measured in pools containing a mixture of negative and potentially positive
474 samples. By measuring viral loads from multiple such pools, it is possible to estimate the
475 prevalence of positive samples without individual testing. See **Supplementary Material 1,**
476 **section 5** for full details.

477 We evaluated prevalence estimation under a range of sample availabilities (N total samples;
478 $N=288$ to $\sim 18,000$) and pooling designs. We varied the pool size of combined specimens (n
479 samples per pool; $n=48, 96, 192,$ or 384) and the number of pools ($b=6, 12, 24,$ or 48). For
480 each combination, we estimated the point prevalence from pooled tests on random samples
481 of individuals drawn during epidemic growth (days 20-120) and decline (days 155-300).

482 Because the data is realistic but simulated, we used ground truth prevalence in the
483 population and, separately, in the specific set of samples collected from the overall
484 population to assess accuracy of our estimates (see for example **Fig. 3B**). We calculated
485 estimates for 100 entirely distinct epidemic simulations.

486 **Pooled tests for individual sample identification**

487 Using the same simulated population, we evaluated a range of simple and combinatorial
488 pooling strategies for individual positive sample identification (**Supplementary Material 1,**
489 **section 6**). In simple pooling designs, each sample is placed in one pool, and each pool
490 consists of some pre-specified number of samples. If a pool tests positive, all samples that
491 were placed in that pool are retested individually (**Fig. 1A**). For combinatorial pooling, each
492 sample is split into multiple, partially overlapping pools (**Fig. 1B**). (9,10) Every sample that
493 was placed in any pool that tested negative is inferred to be negative, and the remaining
494 samples are identified as potential positives. Here, we consider a very simple form of
495 combinatorial testing, where identified potential positive samples are individually tested in a
496 validation stage.

497 A given pooling design is defined by three parameters: the total number of individuals to be
498 tested (N); the total number of pools to test (b); and the number of pools a given sample is
499 included in (q). For instance, if we have 50 individuals (N) to test, we might split the 50 samples
500 into four pools (b) of 25 samples each, where each sample is included in two pools. Note that,
501 by definition, in simple pooling designs each sample is placed in one pool ($q=1$).

502 To identify optimal testing designs under different resource constraints, we systematically
503 analyzed a large array of pooling designs under various sample and test kit availabilities. We
504 evaluated different combinations of between 12 and ~6,000 available samples/tests per day.
505 The daily testing capacity shown is the daily average, though we assume that there is some
506 flexibility to use fewer or more tests day to day (*i.e.*, there is a budget for period of time under
507 evaluation).

508 For each set of resource constraints, we evaluated designs that split N samples between 1 to
509 96 distinct pools, and with samples included in $q=1$ (simple pooling), 2, 3, or 4 (combinatorial
510 pooling) pools (**Table S1**). To ensure robust estimates (especially at low prevalences of less
511 than 1 in 10,000), we repeated each simulated pooling protocol at each time point in the
512 epidemic up to 200,000 times.

513 In each scenario, we calculated: i) the sensitivity to detect positive samples when they existed
514 in the pool; ii) the efficiency, defined as the total number of samples tested divided by the total
515 number of tests used; iii) the total number of identified true positives (total recall); and iv) the
516 effectiveness, defined as the total recall relative to individual testing.

517 **Pilot experiments**

518 For validation experiments of our simulation-based approach, we used fully de-identified,
519 discarded human nasopharyngeal specimens obtained from the Broad Institute of MIT and
520 Harvard. In each experiment, sample aliquots were pooled before RNA extraction and qPCR

521 and pooled specimens were tested using the Emergency Use Authorization (EUA) approved
522 SARS-CoV-2 assay performed by the Broad Institute CLIA laboratory. The protocol is
523 described in full detail in **Supplementary Material 1, section 7**. Specifics of each pooling
524 approach is detailed alongside the relevant results and in **Supplementary Material 1, section**
525 **8**.

526 **References**

- 527 1. Coronavirus disease (COVID-19) Weekly Epidemiological Update Global
528 epidemiological situation. [cited 2020 Aug 28]. Available from:
529 [https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200824-](https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200824-weekly-epi-update.pdf?sfvrsn=806986d1_4)
530 [weekly-epi-update.pdf?sfvrsn=806986d1_4](https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200824-weekly-epi-update.pdf?sfvrsn=806986d1_4)
- 531 2. World Health Organization (WHO). Critical preparedness, readiness and response
532 actions for COVID-19. World Heal Organ. 2020;
- 533 3. Cheng MP, Papenburg J, Desjardins M, Kanjilal S, Quach C, Libman M, et al.
534 Diagnostic Testing for Severe Acute Respiratory Syndrome–Related Coronavirus-2.
535 *Ann Intern Med.* 2020;
- 536 4. Yoo JH, Chung MS, Kim JY, Ko JH, Kim Y, Kim YJ, Kim JM, Chung YS, Kim HM, Han
537 MG KSY. Report on the epidemiological features of coronavirus disease 2019 (covid-
538 19) outbreak in the republic of korea from january 19 to march 2, 2020. *J Korean Med*
539 *Sci.* 2020;
- 540 5. Lee VJ, Chiew CJ, Khong WX. Interrupting transmission of COVID-19: lessons from
541 containment efforts in Singapore. *J Travel Med.* 2020;
- 542 6. Gudbjartsson DF, Helgason A, Jonsson H, Magnusson OT, Melsted P, Norddahl GL,
543 et al. Spread of SARS-CoV-2 in the Icelandic Population. *N Engl J Med.* 2020;
- 544 7. Grassly NC, Pons-Salort M, Parker EPK, White PJ, Ferguson NM, Ainslie K, et al.
545 Comparison of molecular testing strategies for COVID-19 control: a mathematical
546 modelling study. *Lancet Infect Dis.* 2020 Aug [cited 2020 Aug 20];0(0). Available from:
547 <https://linkinghub.elsevier.com/retrieve/pii/S1473309920306307>
- 548 8. Dorfman R. The Detection of Defective Members of Large Populations. *Ann Math*
549 *Stat.* 1943;
- 550 9. Du D, Hwang FK, Hwang F. Combinatorial group testing and its applications. Vol. 12.
551 World Scientific; 2000.
- 552 10. Porat E, Rothschild A. Explicit non-adaptive combinatorial group testing schemes. In:
553 *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial*
554 *Intelligence and Lecture Notes in Bioinformatics).* 2008.
- 555 11. Centre for Disease Prevention E. Methodology for estimating point prevalence of
556 SARS-CoV-2 infection by pooled RT-PCR testing. 2020 [cited 2020 Aug 28]. Available

- 557 from: [https://www.ecdc.europa.eu/en/publications-data/methodology-estimating-point-](https://www.ecdc.europa.eu/en/publications-data/methodology-estimating-point-prevalence-sars-cov-2-infection-pooled-rt-pcr)
558 [prevalence-sars-cov-2-infection-pooled-rt-pcr](https://www.ecdc.europa.eu/en/publications-data/methodology-estimating-point-prevalence-sars-cov-2-infection-pooled-rt-pcr)
- 559 12. Zenios SA, Wein LM. Pooled testing for HIV prevalence estimation: Exploiting the
560 dilution effect. Vol. 17, *Statistics in Medicine*. John Wiley & Sons, Ltd; 1998 [cited
561 2020 Aug 14]. p. 1447–67. Available from:
562 [https://onlinelibrary.wiley.com/doi/full/10.1002/%28SICI%291097-](https://onlinelibrary.wiley.com/doi/full/10.1002/%28SICI%291097-0258%2819980715%2917%3A13%3C1447%3A%3AAID-SIM862%3E3.0.CO%3B2-K)
563 [0258%2819980715%2917%3A13%3C1447%3A%3AAID-SIM862%3E3.0.CO%3B2-K](https://onlinelibrary.wiley.com/doi/full/10.1002/%28SICI%291097-0258%2819980715%2917%3A13%3C1447%3A%3AAID-SIM862%3E3.0.CO%3B2-K)
- 564 13. Wein LM, Zenios SA. Pooled Testing for HIV Screening: Capturing the Dilution Effect.
565 Vol. 44, *Operations Research*. INFORMS; [cited 2020 Aug 14]. p. 543–69. Available
566 from: <https://www.jstor.org/stable/171999>
- 567 14. Hogan CA, Sahoo MK, Pinsky BA. Sample Pooling as a Strategy to Detect
568 Community Transmission of SARS-CoV-2. *JAMA*. 2020 May 19 [cited 2020 Jul
569 3];323(19):1967. Available from:
570 <https://jamanetwork.com/journals/jama/fullarticle/2764364>
- 571 15. Bilder CR, Iwen PC, Abdalhamid B, Tebbs JM, McMahan CS. Tests in short supply?
572 Try group testing. *Significance* (Oxford, England). 2020 Jun [cited 2020 Jul
573 8];17(3):15–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/32536952>
- 574 16. Ben-Ami R, Klochendler A, Seidel M, Sido T, Gurel-Gurevich O, Yassour M, et al.
575 Large-scale implementation of pooled RNA extraction and RT-PCR for SARS-CoV-2
576 detection. *Clin Microbiol Infect*. 2020 Sep 1;26(9):1248–53.
- 577 17. Shental N, Levy S, Wuvshet V, Skorniakov S, Shalem B, Ottolenghi A, et al. Efficient
578 high-throughput SARS-CoV-2 testing to detect asymptomatic carriers. *Sci Adv*. 2020
579 Aug 21 [cited 2020 Aug 28];eabc5961. Available from:
580 <https://advances.sciencemag.org/lookup/doi/10.1126/sciadv.abc5961>
- 581 18. Zou L, Ruan F, Huang M, Liang L, Huang H, Hong Z, et al. SARS-CoV-2 viral load in
582 upper respiratory specimens of infected patients. *New England Journal of Medicine*.
583 2020.
- 584 19. Wölfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Müller MA, et al.
585 Virological assessment of hospitalized patients with COVID-2019. *Nature*. 2020;
- 586 20. Cevik M, Tate M, Lloyd O, Maraolo AE, Schafers J, Ho A. SARS-CoV-2 viral load
587 dynamics, duration of viral shedding and infectiousness: a living systematic review
588 and meta-analysis. *medRxiv*. 2020 Jul 29 [cited 2020 Aug 28];pre-
589 print:2020.07.25.20162107. Available from:

- 590 <http://medrxiv.org/content/early/2020/07/28/2020.07.25.20162107.abstract>
- 591 21. Williams TC, Wastnedge E, McAllister G, Bhatia R, Cuschieri K, Kefala K, et al.
592 Sensitivity of RT-PCR testing of upper respiratory tract samples for SARS-CoV-2 in
593 hospitalised patients: a retrospective cohort study. medRxiv. 2020 Jun 20 [cited 2020
594 Aug 28];2020.06.19.20135756. Available from:
595 <http://medrxiv.org/content/early/2020/06/20/2020.06.19.20135756.abstract>
- 596 22. Arevalo-Rodriguez I, Buitrago-Garcia D, Simancas-Racines D, Zambrano-Achig P, del
597 Campo R, Ciapponi A, et al. FALSE-NEGATIVE RESULTS OF INITIAL RT-PCR
598 ASSAYS FOR COVID-19: A SYSTEMATIC REVIEW. medRxiv. 2020 Aug 13 [cited
599 2020 Sep 25];2020.04.16.20066787. Available from:
600 <https://doi.org/10.1101/2020.04.16.20066787>
- 601 23. Larremore DB, Wilder B, Lester E, Shehata S, Burke JM, Hay JA, et al. Test
602 sensitivity is secondary to frequency and turnaround time for COVID-19 surveillance.
603 medRxiv. 2020 Jun 27 [cited 2020 Aug 28];2020.06.22.20136309. Available from:
604 <https://www.medrxiv.org/content/10.1101/2020.06.22.20136309v2>
- 605 24. Paltiel AD, Zheng A, Walensky RP. Assessment of SARS-CoV-2 Screening Strategies
606 to Permit the Safe Reopening of College Campuses in the United States. JAMA Netw
607 open. 2020 Jul 1 [cited 2020 Aug 28];3(7):e2016818. Available from:
608 <https://jamanetwork.com/>
- 609 25. Riley S, Ainslie K, Eales O, Jeffrey B, Walters C, Atchison C, et al. Community
610 prevalence of SARS-CoV-2 virus in England during May 2020: REACT study.
611 medRxiv. 2020 Jul 11 [cited 2020 Aug 28];2020.07.10.20150524. Available from:
612 <https://doi.org/10.1101/2020.07.10.20150524>
- 613 26. Pouwels KB, House T, Robotham J V, Birrell P, Gelman AB, Bowers N, et al.
614 Community prevalence of SARS-CoV-2 in England: Results from the ONS
615 Coronavirus Infection Survey Pilot. medRxiv. 2020 Jul 7 [cited 2020 Jul
616 8];2020.07.06.20147348. Available from:
617 <https://www.medrxiv.org/content/10.1101/2020.07.06.20147348v1>
- 618 27. Lavezzo E, Franchin E, Ciavarella C, Cuomo-Dannenburg G, Barzon L, Del Vecchio
619 C, et al. Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'.
620 Nature. 2020 Jun 30 [cited 2020 Aug 28];584(7821):425. Available from:
621 <https://doi.org/10.1038/s41586-020-2488-1>
- 622 28. Jing Q-L, Liu M-J, Zhang Z-B, Fang L-Q, Yuan J, Zhang A-R, et al. Household

- 623 secondary attack rate of COVID-19 and associated determinants in Guangzhou,
624 China: a retrospective cohort study. *Lancet Infect Dis*. 2020 Jun [cited 2020 Aug
625 28];0(0). Available from: <https://doi.org/10.1016/S1473-3099>
- 626 29. Li W, Zhang B, Lu J, Liu S, Chang Z, Cao P, et al. The characteristics of household
627 transmission of COVID-19. *Clin Infect Dis*. 2020 Apr 17 [cited 2020 Aug 28]; Available
628 from: [/pmc/articles/PMC7184465/?report=abstract](https://pubmed.ncbi.nlm.nih.gov/32553448/)
- 629 30. Huang Y-T, Tu Y-K, Lai P-C. Estimation of the secondary attack rate of COVID-19
630 using proportional meta-analysis of nationwide contact tracing data in Taiwan. *J*
631 *Microbiol Immunol Infect*. 2020 Jun 11 [cited 2020 Jul 8]; Available from:
632 <http://www.ncbi.nlm.nih.gov/pubmed/32553448>
- 633 31. Liu Y, Eggo RM, Kucharski AJ. Secondary attack rate and superspreading events for
634 SARS-CoV-2. Vol. 395, *The Lancet*. Lancet Publishing Group; 2020 [cited 2020 Aug
635 28]. p. e47. Available from: <http://ees.elsevier.com/thelancet/www.thelancet.com>
- 636 32. Hamner L, Dubbel P, Capron I, Ross A, Jordan A, Lee J, et al. High SARS-CoV-2
637 Attack Rate Following Exposure at a Choir Practice - Skagit County, Washington,
638 March 2020. *MMWR Morb Mortal Wkly Rep*. 2020 May 15 [cited 2020 Jul
639 8];69(19):606–10. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/32407303>
- 640 33. Lennon NJ, Bhattacharyya RP, Mina MJ, Rehm HL, Hung DT, Smole S, et al.
641 Comparison of viral levels in individuals with or without symptoms at time of COVID-
642 19 testing among 32,480 residents and staff of nursing homes and assisted living
643 facilities in Massachusetts. *medRxiv*. 2020 Jul 26 [cited 2020 Aug
644 14];2020.07.20.20157792. Available from:
645 <https://doi.org/10.1101/2020.07.20.20157792>
- 646 34. Wölfel R, Corman VM, Guggemos W, Seilmaier M, Zange S, Müller MA, et al.
647 Virological assessment of hospitalized patients with COVID-2019. *Nature*. 2020 May
648 28 [cited 2020 Aug 14];581(7809):465–9. Available from:
649 <https://doi.org/10.1038/s41586-020-2196-x>
- 650 35. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, et al. The Incubation
651 Period of Coronavirus Disease 2019 (COVID-19) From Publicly Reported Confirmed
652 Cases: Estimation and Application. *Ann Intern Med*. 2020 Mar 10 [cited 2020 Apr 3];
653 Available from: <http://www.ncbi.nlm.nih.gov/pubmed/32150748>
- 654 36. lazymcmc. [cited 2018 Jun 21]. Available from:
655 <https://github.com/jameshay218/lazymcmc>

- 656 37. Poon LLM, Chan KH, Wong OK, Cheung TKW, Ng I, Zheng B, et al. Detection of
657 SARS Coronavirus in Patients with Severe Acute Respiratory Syndrome by
658 Conventional and Real-Time Quantitative Reverse Transcription-PCR Assays. Clin
659 Chem. 2004 Jan [cited 2020 Aug 14];50(1):67–72. Available from:
660 <https://pubmed.ncbi.nlm.nih.gov/14709637/>

661 **Acknowledgements:** We thank Benedicte Gnanngnon, Rounak Dey, Xihong Lin, Edgar
662 Doriban, Rene Niehus and Heather Shakerchi for useful discussions. Work was supported
663 by the Merkin Institute Fellowship at the Broad Institute of MIT and Harvard (B.C.), by the
664 National Institute of General Medical Sciences (#U54GM088558; J.H. and M.M.); and by an
665 NIH DP5 grant (M.M.), and the Dean's Fund for Postdoctoral Research of the Wharton
666 School and NSF BIGDATA grant IIS 1837992 (D.H.).

667 **Author contributions:** B.C., J.H., A.R. and M.M. conceived the study. B.C. and J.H.
668 performed the simulations and modelling. B.B., M.H, M.C, J.B, and B.S performed the
669 pooled testing. B.C., J.H, A.R, and M.M. analyzed results and wrote the manuscript with
670 input from all authors.

671 **Competing interests:** A.R. is a co-founder and equity holder of Celsius Therapeutics, an
672 equity holder in Immunitas, and until July 31, 2020, was an SAB member of ThermoFisher
673 Scientific, Syros Pharmaceuticals, Asimov, and Neogene Therapeutics. From August 1,
674 2020, A.R. is an employee of Genentech.

675 **Code availability:** The SEIR model, viral kinetics model and MCMC were implemented in R
676 version 3.6.2. The remainder of the work was performed in Python version 3.7. The code
677 used for the MCMC framework is available at: <https://github.com/jameshay218/lazymcmc>. All
678 other code and data used are available at: <https://github.com/cleary-lab/covid19-group-tests>.

679 **Data availability:** Raw data generated in this study are available in **Table S2** and **Table S6**.
680 Simulated data can be regenerated using the accompanying code.

681 **Supplementary Material 1: additional materials and methods**

682 **1. Model of infection dynamics**

683 We implemented a deterministic compartmental model to describe the increase and
684 subsequent decline of SARS-CoV-2 infection incidence and prevalence. The model captured
685 the following progression: individuals begin fully susceptible to infection (S); become exposed
686 but not yet infectious (E); become infectious (I); and are finally removed (R). Our aim was to
687 capture changes in incidence over time (increasing incidence rate before the peak and
688 declining incidence thereafter) and the resulting population-level distribution of viral loads, and
689 we therefore did not model additional complexity such as deaths, pre-symptomatic
690 transmission or age-stratified outcomes. Although incorporating these mechanisms may alter
691 the simulated epidemic dynamics for a given set of transmission parameters, we do not expect
692 them to impact inference from the pooled testing analyses.

693 The model was defined by the following set of ordinary differential equations:

$$694 \quad \frac{dS}{dt} = \frac{-\beta SI}{N}$$

$$695 \quad \frac{dE}{dt} = \frac{\beta SI}{N} - \sigma E$$

$$696 \quad \frac{dI}{dt} = \sigma E - \gamma I$$

$$697 \quad \frac{dR}{dt} = \gamma I$$

698 where β is the transmission rate, scaled to give a basic reproductive number R_0 of 2.5; $1/\sigma$ is
699 the incubation period, assumed to have a mean of 6.4 days; $1/\gamma$ is the infectious period
700 assumed to have a mean of 7 days; and N is the population size, set to 12,500,000 to generate
701 at least 10,000,000 infected individuals. All model parameters and assumed values are

702 described in **Table S5**. Using this model, we simulated per capita daily infection probability
703 (i.e., the daily incidence) and prevalence.

704 **2. Viral load kinetics**

705 Following infection (t_{inc}) and after a latent period (t_g), viral load was assumed to increase
706 exponentially, peaking a short time (t_p) after the end of the latent period and before the onset
707 of symptoms (t_o). Viral load then decreased monotonically to undetectable levels, defined by
708 a time parameter (t_w) giving the number of days post symptom onset before crossing the limit
709 of detection. These assumptions are equivalent to linear growth and decay on a log scale. We
710 chose to model viral load waning with respect to symptom onset rather than peak viral load,
711 as almost all available time-series viral load data are presented with respect to symptom onset
712 time. The \log_{10} viral load over time, $v(t)$, was given by:

$$713 \quad v(t) = \begin{cases} 0, & t_{inf} < t \leq t_g \\ \frac{\alpha}{t_p} t, & t_g < t \leq t_p \\ \alpha - \frac{\alpha}{t_{inc} - t_p + t_w} (t - t_p), & t_g < t \leq t_p \end{cases}$$

714 where t_{inf} is the time of infection; t_{inc} is the incubation period for symptom onset; α is the
715 peak viral load; t_g is the latent period before viral growth; t_p is the time taken to reach peak
716 viral load post; and t_w is the number of days from symptom onset to becoming undetectable.
717 Note that all individuals are assigned a symptom onset time regardless of whether they show
718 symptoms or not, as the parameter is used here to describe viral kinetics.

719 **3. Fitting the viral kinetics model**

720 Time-series viral load data were obtained from a case series of 9 hospitalized COVID-19
721 patients from a single hospital in Munich, Germany. (34) These data provide regular
722 measurements of \log_{10} RNA copies per nasopharyngeal swab and per ml of sputum. Data

723 were extracted using a web plot digitizer (<https://automeris.io/WebPlotDigitizer/>). We used a
724 random-effects model to infer individual-level viral kinetics parameters alongside population-
725 level distributions. Under this model, each individual had their own parameter values for t_{inc} ,
726 α , t_g , t_p and t_w . α and t_w drawn from a multivariate normal distributed with estimated means
727 $(\bar{\alpha}, \bar{t}_w)$, standard deviations $(\sigma_\alpha, \sigma_{t_w})$ and correlation (ρ_{α, t_w}) (**Table S5**).

728 Although the viral load measurements were taken after the onset of symptoms, we are
729 interested in modelling the entire time course from the time of infection, including uncertainty
730 in the incubation period. We therefore estimated t_{inc} for each individual by placing a strongly
731 informative log-normal prior on the incubation period based on previous estimates. (35)
732 Because no observations were made before symptom onset, the posterior distribution for this
733 parameter matched the prior.

734 The model was fit separately to the swab and sputum data using a Markov chain Monte
735 Carlo (MCMC) framework, using a likelihood that assumes normally distributed observation
736 error with estimated standard deviation σ_{obs} (**Fig. S8**). The likelihood function accounted for
737 truncation at the lower limit of detection (LOD) by integrating over the left-hand tail of the
738 normal cumulative density function for model-predicted viral loads less than the LOD.

739 Furthermore, Wölfel et al. present data with a limit of detection of 0 \log_{10} copies / swab or /
740 ml, but a limit of quantification of 2 \log_{10} copies. We therefore considered that observed viral
741 loads between 0 and 2 had unknown between 0 and 2. We placed uniform priors on $\bar{\alpha}$ and
742 \bar{t}_w , half Cauchy priors on σ_α and σ_{t_w} with scale parameters of 1, a uniform prior on ρ_{α, t_w}
743 between -1 and 1 and informative uniform priors on t_g , and t_p .

744 Code to regenerate the MCMC fitting is available in the accompanying GitHub repository. (36)
745 Briefly, an adaptive Metropolis-Hastings algorithm was run for 3,000,000 iterations with the
746 first 1,000,000 iterations discarded as burn in to obtain posterior estimates for the model
747 parameters (**Fig. S9**). Convergence was assessed visually based on trace plots of 3

748 independent chains (**Fig. S10**). All estimated parameters had an effective sample size (ESS)
749 greater than 200 and $\hat{R} < 1.1$.

750 **4. Simulation framework**

751 We combined the estimates from the transmission model and viral kinetics model above to
752 simulate viral loads for a population of 12,500,000 individuals over time. We performed
753 separate simulations for swab and sputum data. For each individual, we simulated a time of
754 infection (or no infection) from the time-varying probability of infection, $p(t)$. For each infected
755 individual, an incubation period was drawn from a log normal distribution with mean (of the
756 natural log) = 1.621 and standard deviation (of the natural log) = 0.418 as estimated previously.
757 (35) This incubation period was used as a reference point for t_w , but we did not explicitly model
758 symptomatic vs. asymptomatic individuals. Next, we drew viral kinetics parameters for each
759 individual from the joint posterior distribution. Finally, we simulated the time-varying viral load
760 of each infected individual, starting at the time of infection. Because of the high variance of
761 the drawn kinetics parameters, some simulated viral loads reached very high values, and we
762 therefore truncated all simulated viral loads at 16 \log_{10} virus copies per ml. These simulations
763 provide a population of viral loads incorporating realistic variability in viral load kinetics and
764 between-individual parameters.

765 **5. Estimating prevalence from pooled test results**

766 In this section, we adapt a previously described maximum likelihood estimate (MLE)
767 framework to estimate prevalence of positive samples, p , amongst N samples using b pooled
768 tests, with each pooling containing $n = \frac{N}{b}$. (12,13) Briefly, the MLE is determined by computing
769 the conditional probability of an observed RT-qPCR result given that there were k positive
770 samples in the pool, and integrating over all values of k . To compute these conditional
771 distributions, we estimate an empirical distribution of cycle threshold (Ct) values that follows
772 from the distribution of viral loads in infected individuals derived from a subset of the simulated

773 viral load measurements described above. This is conceptually similar to using the empirical
774 distribution of all real measurements from a population taken to date.

775 Let $f_{k,n} = P(Y|k, n)$ represent the conditional probability of the observed Ct values in each
776 pool, $Y = [y_1, \dots, y_b]$, given that there were k positive samples in the pool and $n-k$ negative
777 samples. Computing these distributions will be the central task of generating prevalence
778 estimates.

779 To begin, we model Ct values as a function of the underlying viral loads:

$$780 \quad y(v) = y_0 - \log_2(10) \log_{10}(v),$$

781 where v is the viral load (in copies per ml). To capture variance in the LOD between labs or
782 across batches within a lab, we let the intercept, y_0 , be normally distributed with unknown
783 mean and standard deviation. We assume uniform priors over each y_0 ($\mu \sim Uniform(37,39)$,
784 $\sigma \sim Uniform(0.1,1)$). These ranges and the log-linear relationship are based on a calibration
785 of Ct values to a dilution curve of positive control SARS-CoV-2 RNA samples (data not shown)
786 and are consistent with prior measurements calibrated to SARS coronavirus. (37)

787 $f_{k,n}$ depends not only on the distribution of Ct values arising from a given viral load, but also
788 on the distribution of viral loads in infected and uninfected individuals. Here we assume that
789 the viral load in uninfected individuals is always zero (we do, however, allow for false
790 positive PCR results, as discussed below). Given access to the true distribution of viral
791 loads, $f_{k,n}$ can be generated by the convolution of k density functions. Since we do not have
792 access to the true distributions, we approximate $f_{k,n}$ through a kernel density estimate (KDE)
793 of empirical convolutions. Specifically, for a given value of n and each value of k from 1 to n ,
794 we generate 10,000 random combinations of k viral loads,

$$795 \quad V_{k,n}: v_i = \sum_{j=1}^k z_j \text{ for } i \text{ from } 1 \text{ to } 10,000,$$

796 where z_j are the randomly sampled viral loads. For each i , we sample $\mu \sim Uniform(37,39)$,
797 $\sigma \sim Uniform(0.1,1)$ and $y_0 \sim N(\mu, \sigma)$ and then generate Ct values:

798 $X_{k,n}: x_i = y(v_i)$ for i from 1 to 10,000.

799 We then approximate $f_{k,n} = KDE(X_{k,n})$ with a bandwidth parameter of 0.1.

800 Several adjustments to these density approximations are needed to account for false
801 positives and undefined Ct values (i.e., undetected viral RNA). To allow for Ct values that
802 might arise from false positives, we define $f_{0,n} = f_{1,n} r$, where r is an assumed false positive
803 rate of PCR (here, we set $r = 0.2\%$ and allow this to be misspecified when simulating false
804 positives below). When viral RNA is undetected and $k > 1$, we let $f_{k,n}(undetected) =$
805 $\int_{40}^{\infty} f_{k,n}(x) dx$, where a Ct value of 40 is used as a typical limit of detectable RNA. When $k =$
806 0 , $f_{0,n}(undetected) = 1 - r$.

807 Following the model of Zenios and Wein, (12) the MLE of prevalence, p , is defined as:

808
$$p_{estimated} = \frac{1}{nb} \sum_{i=1}^b \sum_{k=0}^n k P(k | y_i, p),$$

809 where n is the number of samples per pool, b , is the number of pools, and y_i is the Ct value
810 observed in pool i . We calculate $P(k | y_i, p)$ using the conditional densities above:

811
$$P(k | y_i, p) = \frac{\frac{n}{k} p^k (1-p)^{n-k} f_{k,n}(y_i - \log_2(10) \log_{10}(n))}{\sum_{j=0}^n \frac{n}{j} p^j (1-p)^{n-j} f_{j,n}(y_i - \log_2(10) \log_{10}(n))},$$

812 substituting the appropriate values of f when the Ct value is undefined. Note that each of
813 the observed Ct values is adjusted to account for n -fold dilution. Finally, we find $p_{estimated}$
814 through an iterative expectation-maximization algorithm, where at each iteration

815
$$p_{t+1} = \frac{1}{nb} \sum_{i=1}^b \sum_{k=0}^n kP(k | y_i, p_t).$$

816 Using this model, we evaluated prevalence estimation across time in our simulated
817 populations. We first partition the simulated population into training and testing sets. Observed
818 viral loads in 20% of individuals (selected at random) are used to train the KDEs above.
819 Furthermore, training data were divided into growth phase (days 20 to 120) or decline phase
820 (days 155 to 300) data, allowing us to ascertain the effect of training and estimating on data
821 from consistent or inconsistent phases of the epidemic. All remaining prevalence estimation
822 and analyses are done on the testing population. We simulate viral loads in b pools of samples
823 as $V: v_i = \sum_{j=1}^n \text{Poisson}(\frac{z_j}{n})$ for i from 1 to b , where Poisson sampling is used to model the
824 sampling of a small volume from each swab. From these we simulate Ct values, Y . At each
825 time point, we randomly select N individuals, partition them into b pools, and simulate the
826 pooled viral loads and Ct values, Y , with $y_0 \sim N(38.5, 1)$ sampled randomly in each trial and
827 applying the transformation described previously. To capture false positive PCR results at a
828 rate of 1%, with 1% probability in each pool we add a viral load of $\text{Poisson}(\frac{z_j}{n})$, with z_j selected
829 at random from all positive samples with $\frac{z}{n} > LOD$ on a given date. We then compute the true
830 prevalence in the population, p_{true} , the prevalence in the population of N samples, $p_{sampled}$,
831 and the prevalence estimated from b tests, $p_{estimated}$. To quantify accuracy we calculate the
832 mean and standard deviation of these values at each time point across 100 random trials and
833 summarize differences between $p_{estimated}$ and either p_{true} or $p_{sampled}$ with the normalized
834 root mean squared error (NRMSE; either $\frac{RMSE_{true}}{p_{true}}$ or $\frac{RMSE_{sampled}}{p_{sampled}}$; **Fig. S2**).

835 To demonstrate the importance of calibrating the correct empirical distribution of Ct values,
836 **Fig. S2** shows the accuracy of the framework in recovering the true population prevalence
837 during the growth and decline phases of the epidemic when the model is calibrated to Ct

838 values measured during the growth and decline phases. Estimation accuracy is highest when
839 the calibrated Ct values were measured during the same phase as is being measured.

840 **6. Group testing for sample identification**

841 We evaluated an array of pooling designs for individual identification as described in the main
842 text. For any given design we can simulate pooled testing results on the simulated population
843 described above. For each time point, we randomly select N individuals from the population.
844 Each individual is then randomly assigned to q out of b pools. For each pool, we then calculate
845 the net sampled viral load, $v_i = \sum_{j \in pool_i} Poisson(\frac{z_j}{a})$ using the viral loads z_j for each individual
846 included in $pool_i$, where Poisson sampling is used to model a small volume sampled from
847 each swab. If $v_i > LOD$ the pool is assigned a positive value, otherwise the pool is negative
848 (here, with $LOD=100$). To allow for false positive PCR results at a rate of 1%, with probability
849 1% we set the result of the pool to be positive, regardless of the viral load.

850 Simulations were run by repeating the random pooling, pooled testing, and decoding
851 procedures described in the main text. At each point in time, 200,000 trials are run selecting
852 N individuals at random in each trial. We then record the number of validation tests, s , in
853 each trial, corresponding to the number of putative positive samples. Average efficiency at a
854 point in time, t , is then calculated as $Efficiency_t = \frac{N}{b+s}$. Average sensitivity is calculated as

855 $Sensitivity_t = \frac{N_{putative}}{N_{positive}}$ where $N_{positive}$ is the total number of infected individuals (viral load >

856 1) that were sampled across all trials and $N_{putative}$ is the number of such individuals who

857 were identified as putative positives. From these values we calculate $Total Recall_t =$

858 $Efficiency_t Sensitivity_t$.

859 Based on these results we evaluated a large array of group testing designs (**Table S1**) under
860 a series of constraints on number of samples collected and number of reactions run per day
861 (**Fig. 4C**). For a given design, $D: (N, b, q)$ we calculate the average number of tests run, $s_{total} =$

862 $b + \underline{s}$, and the total number of times that D could be run on a daily basis, $r_D =$
863 $\min(\frac{\text{samples collected}}{N}, \frac{\text{testing kits}}{s_{total}})$. We then calculate the effective testing capacity on a daily
864 basis, $c_D = N r_D \text{Sensitivity}_D$, and average this value over days 40 to 90 (**Fig. 4C**) or 80 to
865 108 (**Fig. S6**) of the simulated epidemic. Notably, when sample collection is limiting, $c_D =$
866 $\text{samples collected} * \text{Sensitivity}_{(\text{samples collected}, \text{samples collected}, 1)}$, corresponding to individual
867 testing. Finally, the optimal design for a given combination of constraints is the design with the
868 greatest average effective testing capacity. In **Fig. 4C** and **Fig. S6** we show the relative
869 effectiveness, corresponding to effective capacity relative to individual testing. **Fig. S7** shows
870 the same results as **Fig. 4** and **Fig. S6**, but when samples were simulated based on viral
871 trajectories using the sputum data. **Fig. S5** shows the same results again, but for the decline
872 phase of the epidemic.

873 **7. Broad Institute CRSP SARS-CoV-2 Real-time Reverse Transcriptase (RT)-PCR** 874 **Diagnostic Assay**

875 RNA is isolated from nasopharyngeal and oropharyngeal specimens in 50ul VTM/UTM using
876 MagMAX™-96 Viral RNA Isolation Kits (Thermo Fisher Scientific, AMB18365), performed on
877 an Agilent Bravo Liquid Handling Platform running VWorks (Build 11.4.0.1233), and is reverse
878 transcribed to cDNA and subsequently amplified in the Applied Biosystems® ViiA7 Real-Time
879 PCR Instrument with QuantStudio version 1.3 software. In the process, the probe anneals to
880 a specific target sequence located between the forward and reverse primers. During the
881 extension phase of the PCR cycle, the 5' nuclease activity of Taq polymerase degrades the
882 probe, causing the reporter dye to separate from the quencher dye, generating a fluorescent
883 signal. With each cycle, additional reporter dye molecules are cleaved from their respective
884 probes, increasing the fluorescence intensity. Fluorescence intensity is monitored at each
885 PCR cycle by Applied Biosystems® ViiA7 Real-Time PCR System with QuantStudio version

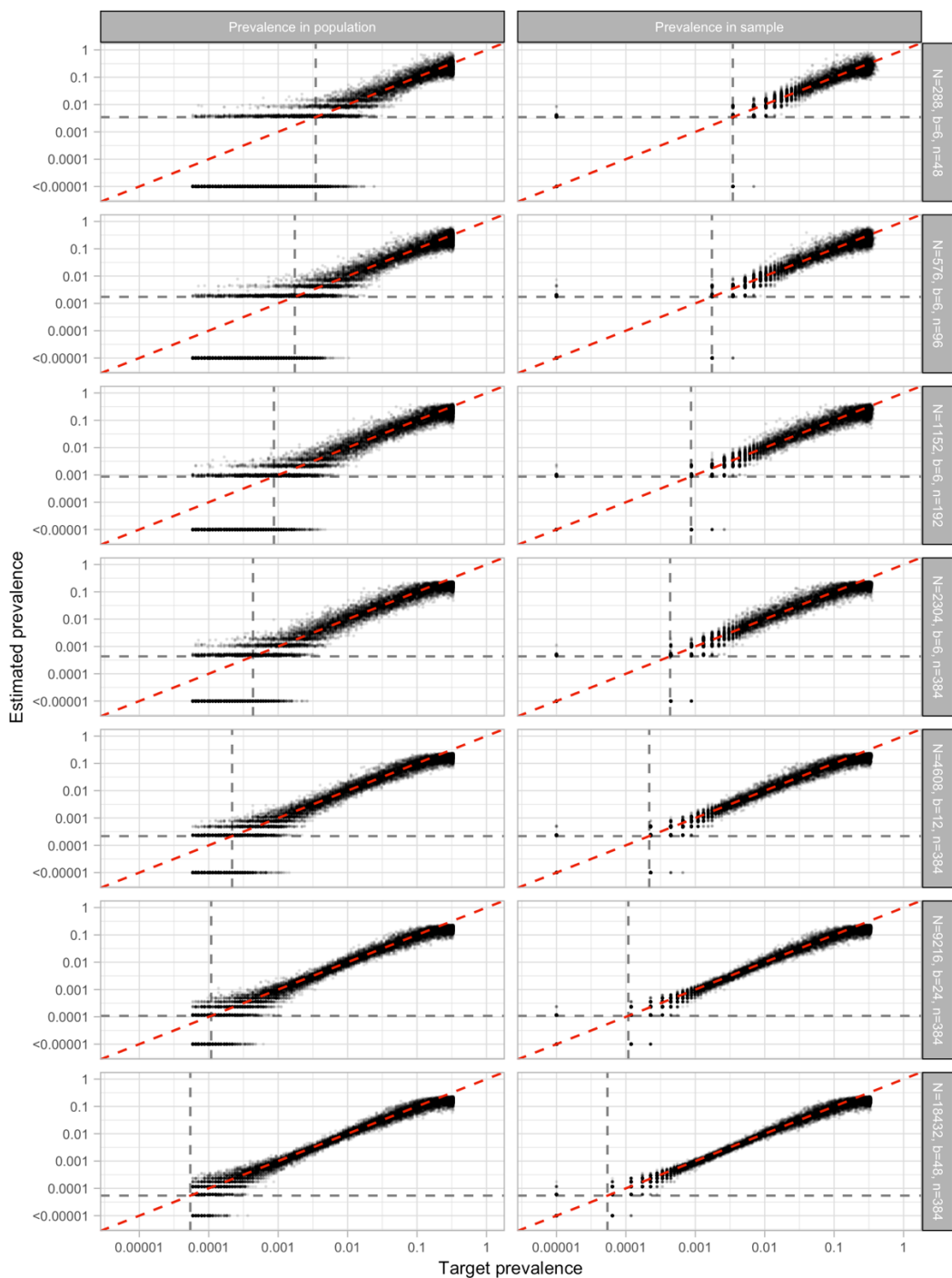
886 1.3 software. The software allows the fluorescence intensity to be monitored at each PCR
887 cycle to allow for the qualitative detection of nucleic acid from SARS-CoV-2.

888 To create the final pools for our larger validation studies (10 batches of 96 for sample
889 identification, and 48 pools of 48 samples for prevalence estimation), we treated a total of 96
890 known negative specimens as distinct samples across batches (in the case of sampled
891 identification study) or pools (for the prevalence study). Each positive sample in a batch or
892 pool was a distinct sample, and only used in one batch or pool. For the prevalence study, only
893 pools with one or more positive samples were tested using the assay above (the remaining
894 pools were assigned an “undefined” Ct value). All pools in all batches were tested in the
895 identification study, regardless of whether they contained a positive sample.

896 **8. Selecting pool compositions for large-scale sample identification validation**

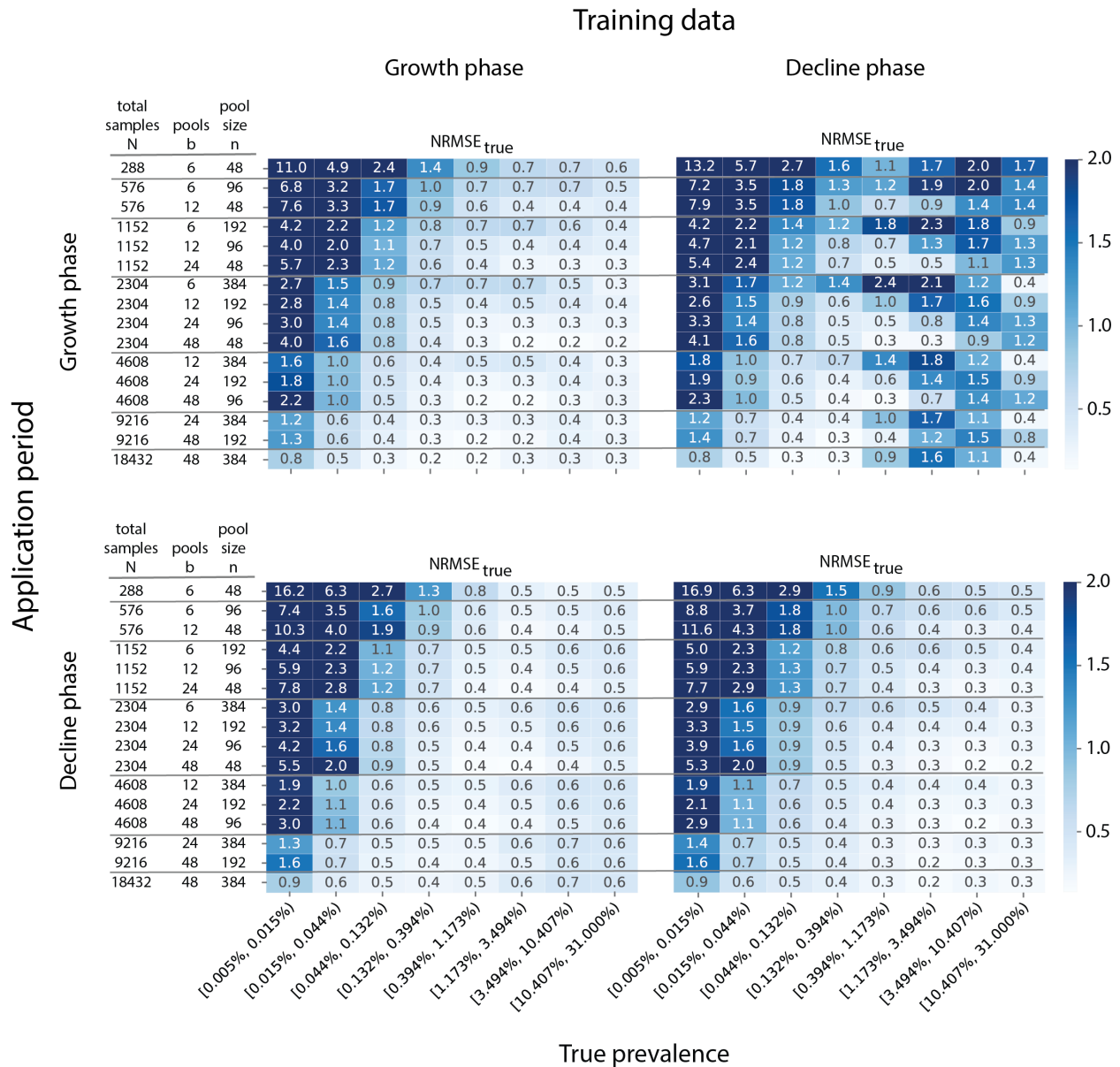
897 To form pools, we put each of the $N = 96$ individuals into $q = 2$ out of $b = 6$ pools (A-F) by
898 cycling through the following ordered list of pool pairs: AB, CD, EF, BC, DF, AE, BD, AF, CE,
899 BE, CF, AD, BF, DE, AC. Namely, individual 1 is put in pools A and B, individual 2 in C and D,
900 and so on; after individual 15, we return to the beginning of the list and cycle through again.
901 Finally, since each pair of pools is assigned to multiple individuals, we reordered the
902 individuals by what pair of pools they were put in. This final reordering simplifies the
903 presentation to one more conveniently arranged for manual pipetting.

904 A strength of the procedure is that it is simple and flexible; it can easily be carried out for any
905 number of individuals. Moreover, the resulting design here has pools with 32 individuals each,
906 so the final pool sizes are balanced. In addition, the first 6 pool pairs are each assigned to 7
907 individuals and the remaining 9 pairs are each assigned to 6 individuals, so the pairs of pools
908 are also approximately balanced. Indeed, the ordered list of pool pairs above was chosen to
909 achieve this balance; the list uses each pair of pools once and the five consecutive sets of
910 three pairs all use each pool once.

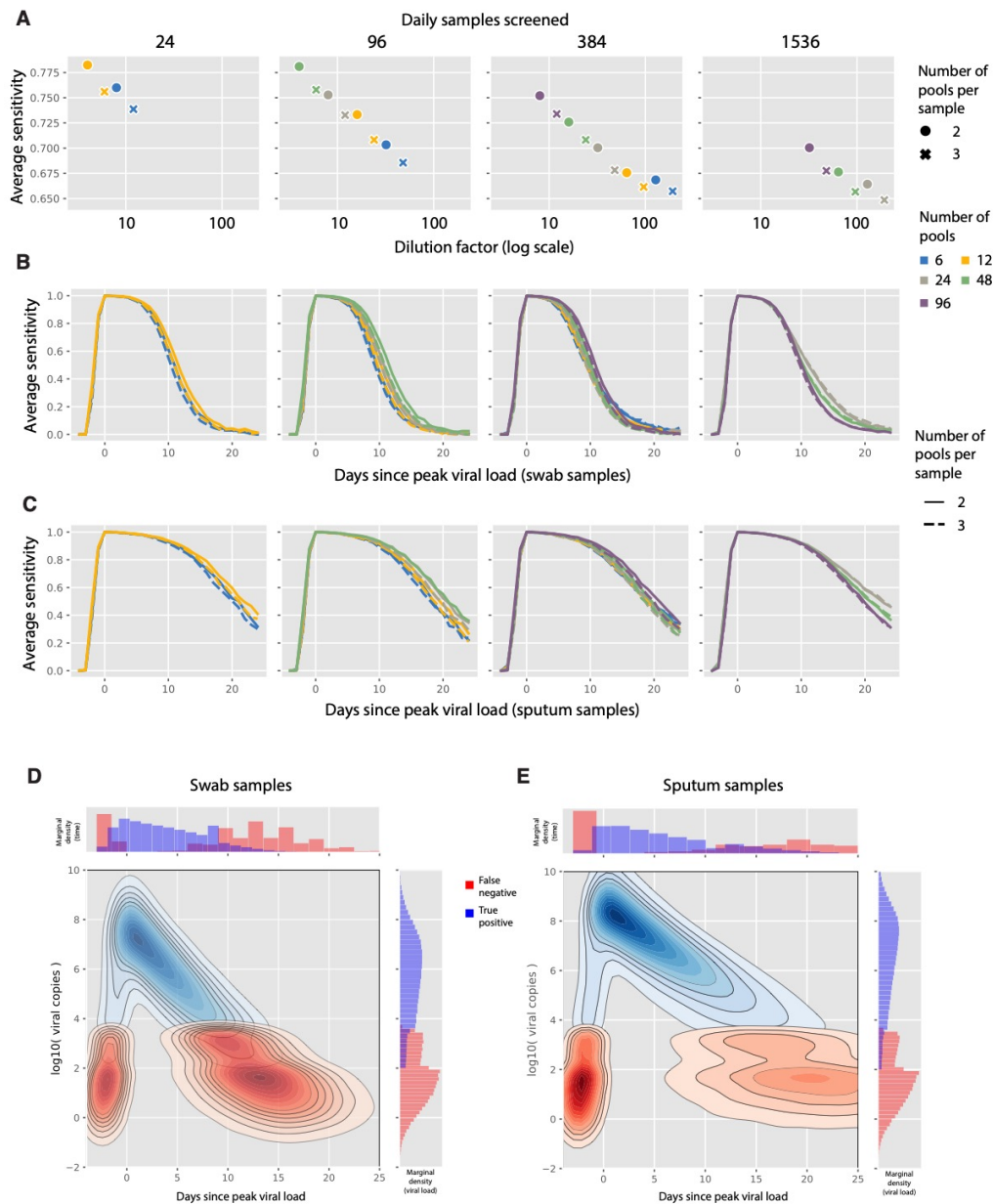


911

912 **Fig. S1:** Population prevalence (left) or prevalence in sample (right) against maximum
913 likelihood prevalence estimates. Population prevalence shown here are during the epidemic
914 growth phase. Results shown are from 100 independent trials at each day of the epidemic.
915 Each facet shows the pooling design with the fewest pools (tests used) for each sample size.
916 Dashed grey lines show one divided by the sample size, N .



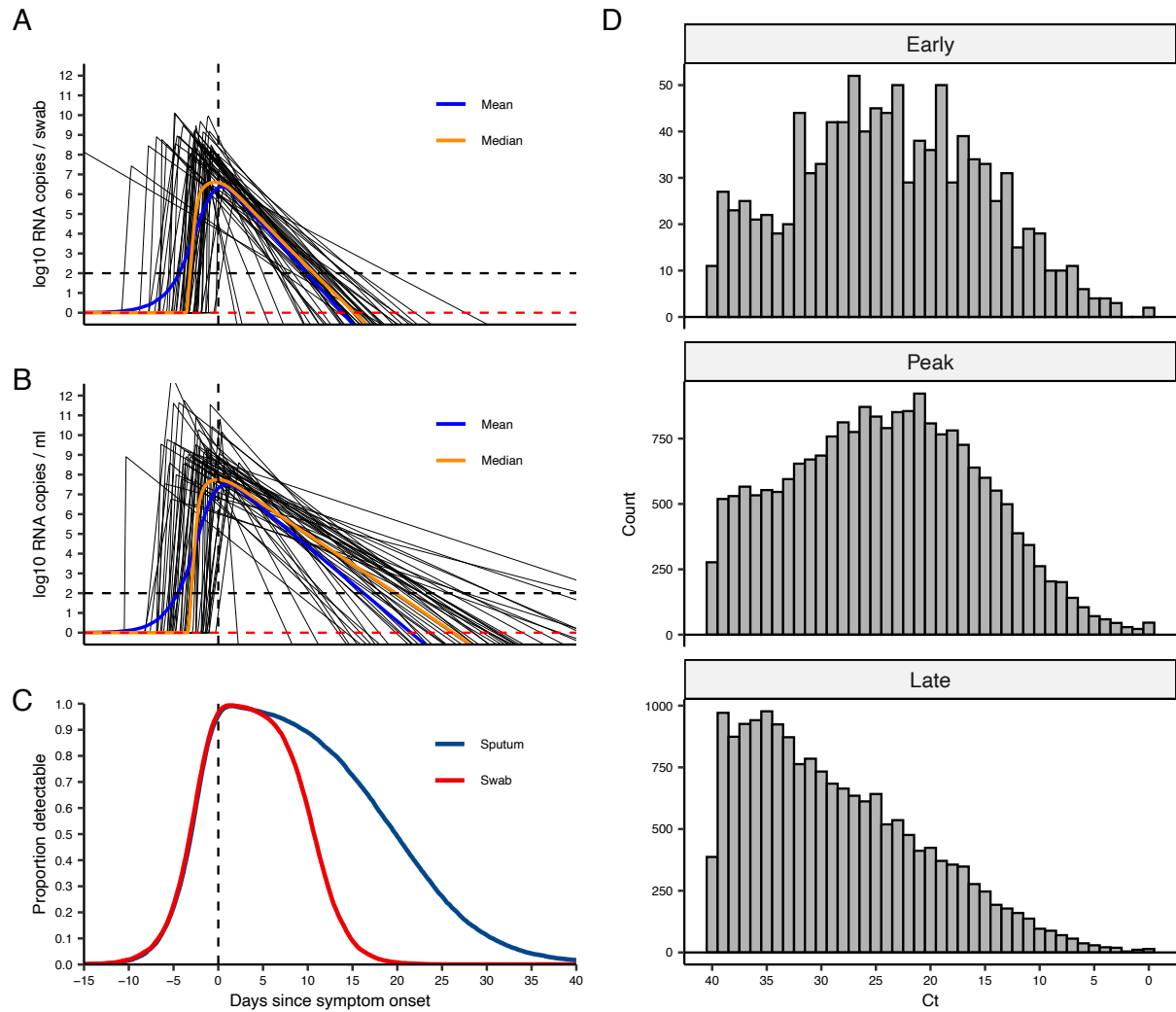
925 (two columns of panels) and application (two rows of panels) period. Colorbar and annotated
926 values within boxes indicate NRMSE.



927

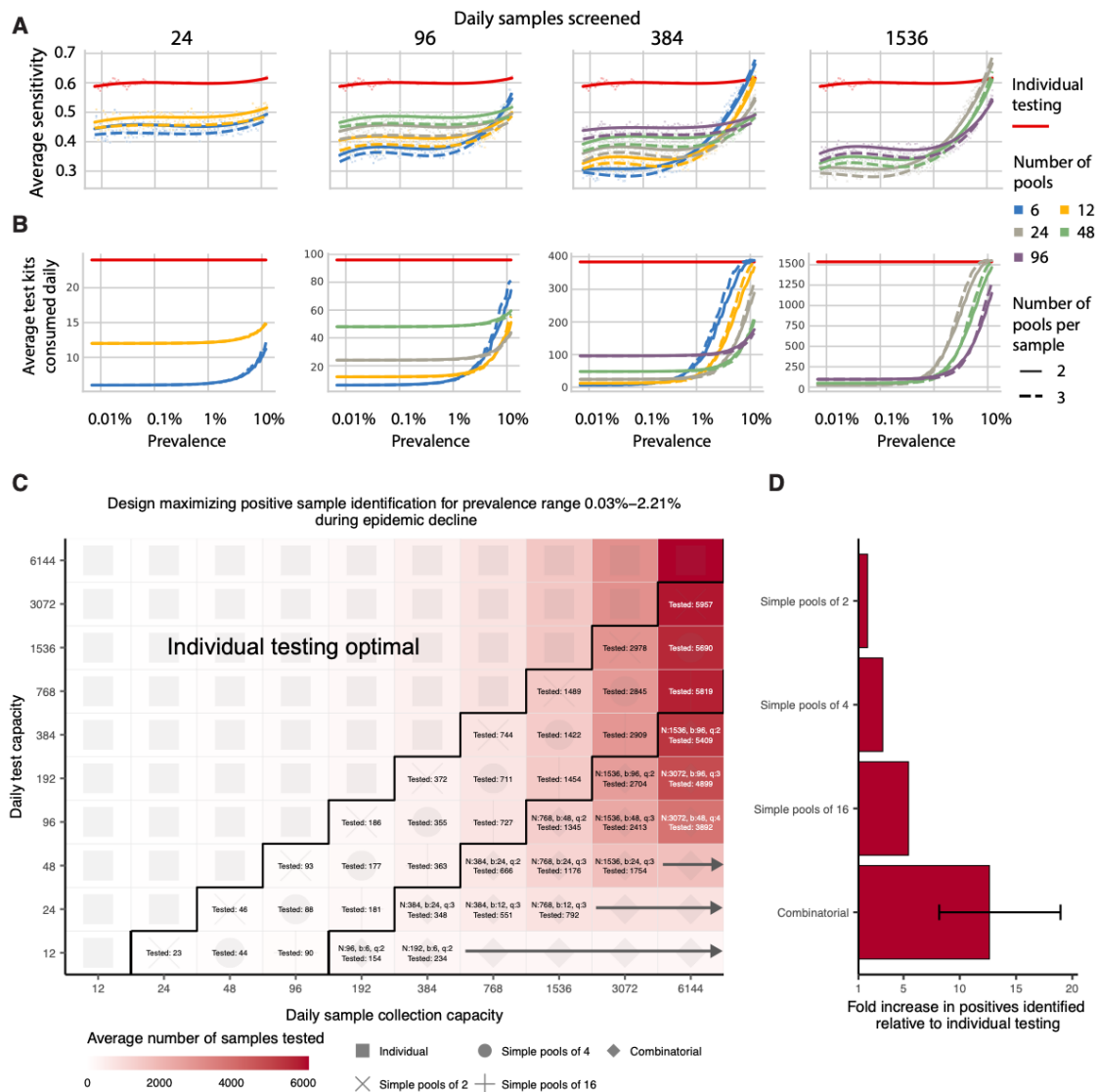
928 **Fig. S3:** Sensitivity of sample identification relative to dilution factor and time since peak viral
 929 load. **(A)** The sensitivity (y-axis) of different designs (individual points) is plotted against the
 930 dilution factor of each design (x-axis, log scale). Pooling designs are separated by the
 931 number of swab samples tested on a daily basis (individual panels); the number of pools
 932 (color); and the number of pools into which each sample is split (circle versus cross). **(B)** and
 933 **(C)** As in **(A)**, with sensitivity plotted against days since peak viral load (x-axis) for swab **(B)**

934 and sputum **(C)** samples. **(D and E)** The density of false negative (red) and true positive
935 (blue) results is shown as a function of days since peak viral load (x-axis) and viral load (y-
936 axis, *log*10 scale) at the time of sample collection and pooled testing for swab **(D)** and
937 sputum **(E)** samples. Contour plots depict 2-d density, and histograms show marginal
938 density over time (top) and viral load (right).



939

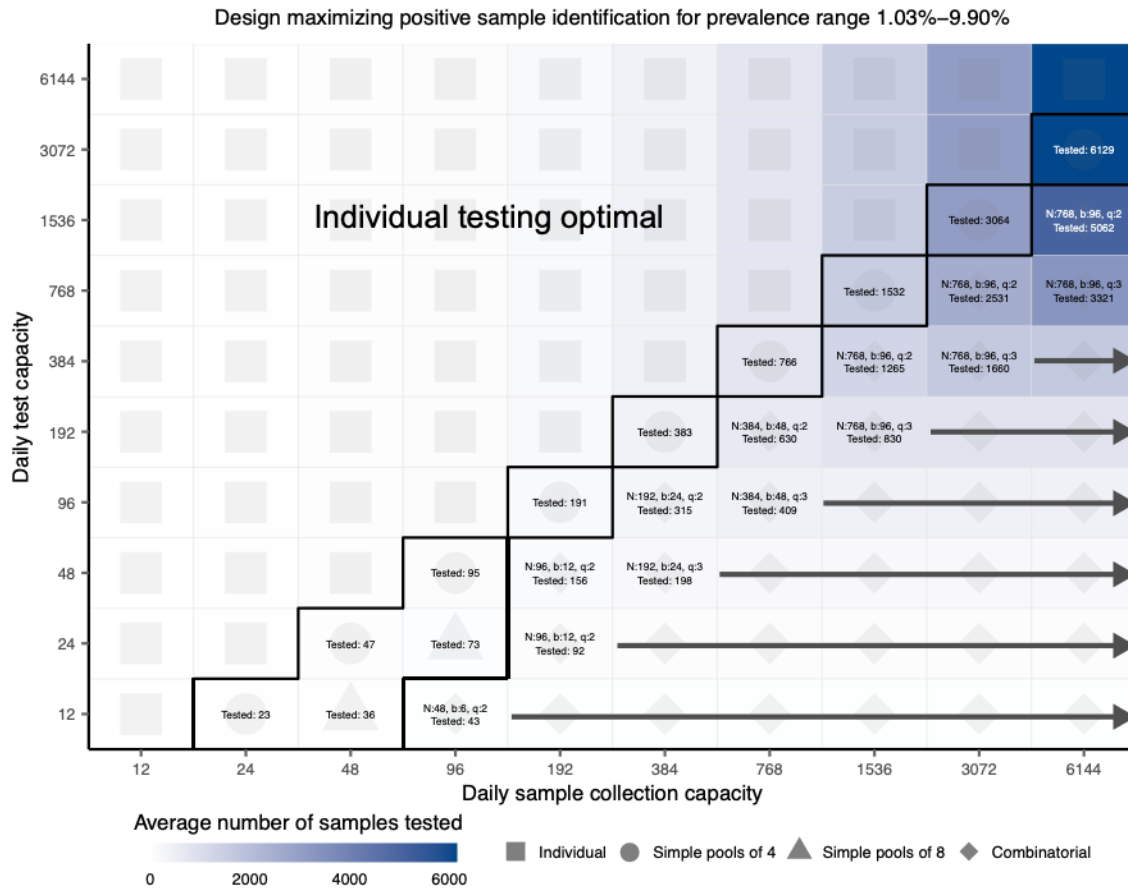
940 **Fig. S4:** Simulated viral loads. **(A)** Black lines show 50 randomly drawn viral load trajectories
 941 based on parameters estimated from fitting to swab data. Vertical dashed line shows time of
 942 symptom onset. Horizontal dashed lines show limit of detection assumed in subsequent
 943 analyses (black line) and limit of detection reported by Wölfel et al (red line). Blue and
 944 orange line shows median viral load on each day with respect to symptom onset. **(B)** as in
 945 **(A)**, but fitted to sputum data. **(C)** Proportion of true viral loads above the limit of detection
 946 ($LOD = 2 \log_{10}$ RNA copies / swab or ml) on each day with respect to symptom onset for
 947 simulated swab (red) and sputum (blue) data. Note that the distribution for observed viral
 948 loads will differ slightly due to the addition of observation and sampling error. **(D)** Distribution
 949 of simulated cycle threshold values from swab samples at different stages of the epidemic.



950

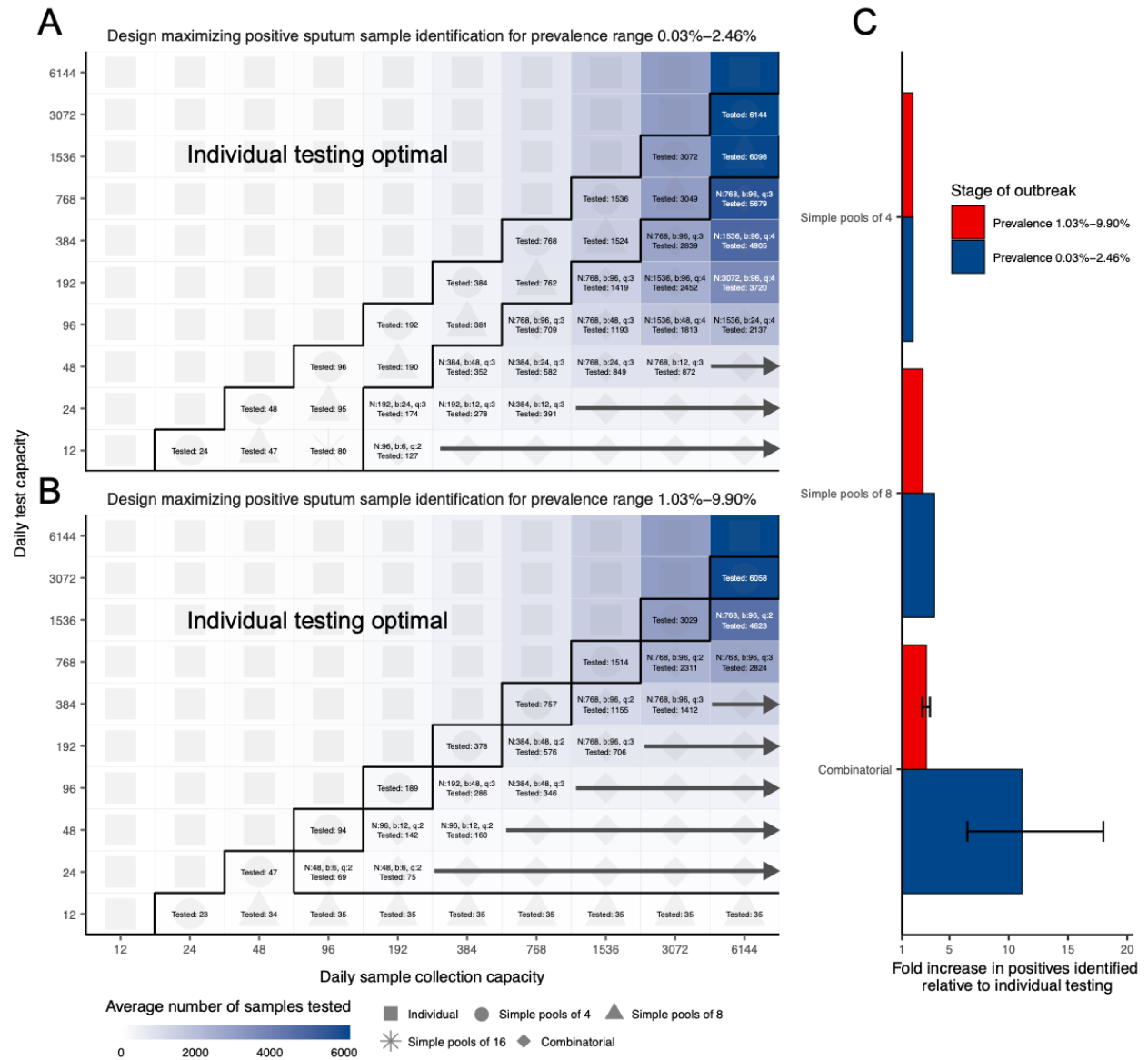
951 **Fig. S5:** Group testing for sample identification during epidemic decline. Results shown are
 952 based on the same testing strategies as in **Fig. 4**. We evaluated a variety of group testing
 953 designs for sample identification (**Table S1**) on the basis of sensitivity (**A**) shows sensitivity
 954 and (**B**) shows efficiency. total number of positive samples identified (**C**) and the fold increase
 955 in positive samples identified relative to individual testing (**D**). (**A** and **B**) The average
 956 sensitivity (**A**, y-axis, individual points and spline) and average number of tests needed to
 957 identify individual positive samples (**B**, y-axis) using different pooling designs (individual lines)
 958 were measured over days 20-110 in our simulated population, with results plotted against

959 prevalence (x-axis, log-scale). Results show the average of 200,000 trials, with individuals
960 selected at random on each day in each trial. Pooling designs are separated by the number
961 of samples tested on a daily basis (individual panels); the number of pools (color); and the
962 number of pools into which each sample is split (dashed versus solid line). Solid red line
963 indicates results for individual testing. **(C)** Every design was evaluated under constraints on
964 the maximum number of samples collected (columns) and average number of reactions that
965 can be run on a daily basis (rows) over days 40-90. Text in each box indicates the optimal
966 design for a given set of constraints (number of samples per batch (N), number of pools (b),
967 number of pools into which each sample is split (q), average number of total samples screened
968 per day). Color indicates the average number of samples screened on a daily basis using the
969 optimal design. Arrows indicate that the same pooling design is optimal at higher sample
970 collection capacities due to testing constraints. **(D)** Fold increase in the number of positive
971 samples identified relative to individual testing with the same resource constraints. Error bar
972 shows range amongst optimal designs.



973

974 **Fig. S6:** Effectiveness of optimal testing design under resource constraints at high
 975 prevalence. As in **Fig. 4**, every design was evaluated under constraints on the maximum
 976 number of samples collected (columns) and average number of reactions that can be run on
 977 a daily basis (rows), here from days 80-108. Text in each box indicates the optimal design
 978 for a given set of constraints (number of samples per batch (N), number of pools (b), number
 979 of pools into which each sample is split (n), average number of total samples screened per
 980 day). Color indicates the average number of samples screened on a daily basis using the
 981 optimal design. Arrows indicate that the same pooling design is optimal at higher sample
 982 collection capacities.



983

984 **Fig. S7:** Effectiveness of optimal testing design under resource constraints using sputum data.

985 **(A)** and **(B)** show optimal designs at lower and higher prevalence ranges respectively. Text in

986 each box indicates the optimal design for a given set of constraints (number of samples per

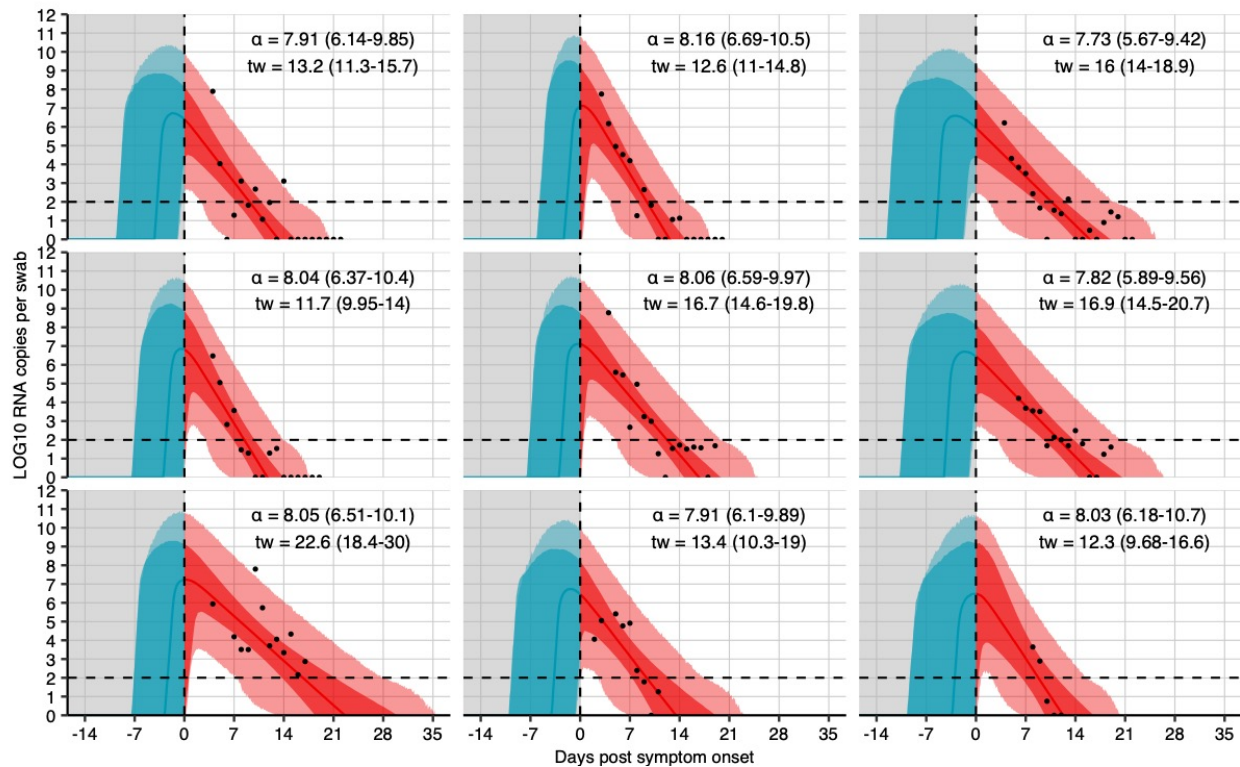
987 batch (N), number of pools (b), number of pools into which each sample is split (n), average

988 number of total samples screened per day). Color indicates the average number of samples

989 screened on a daily basis using the optimal design. Arrows indicate that the same pooling

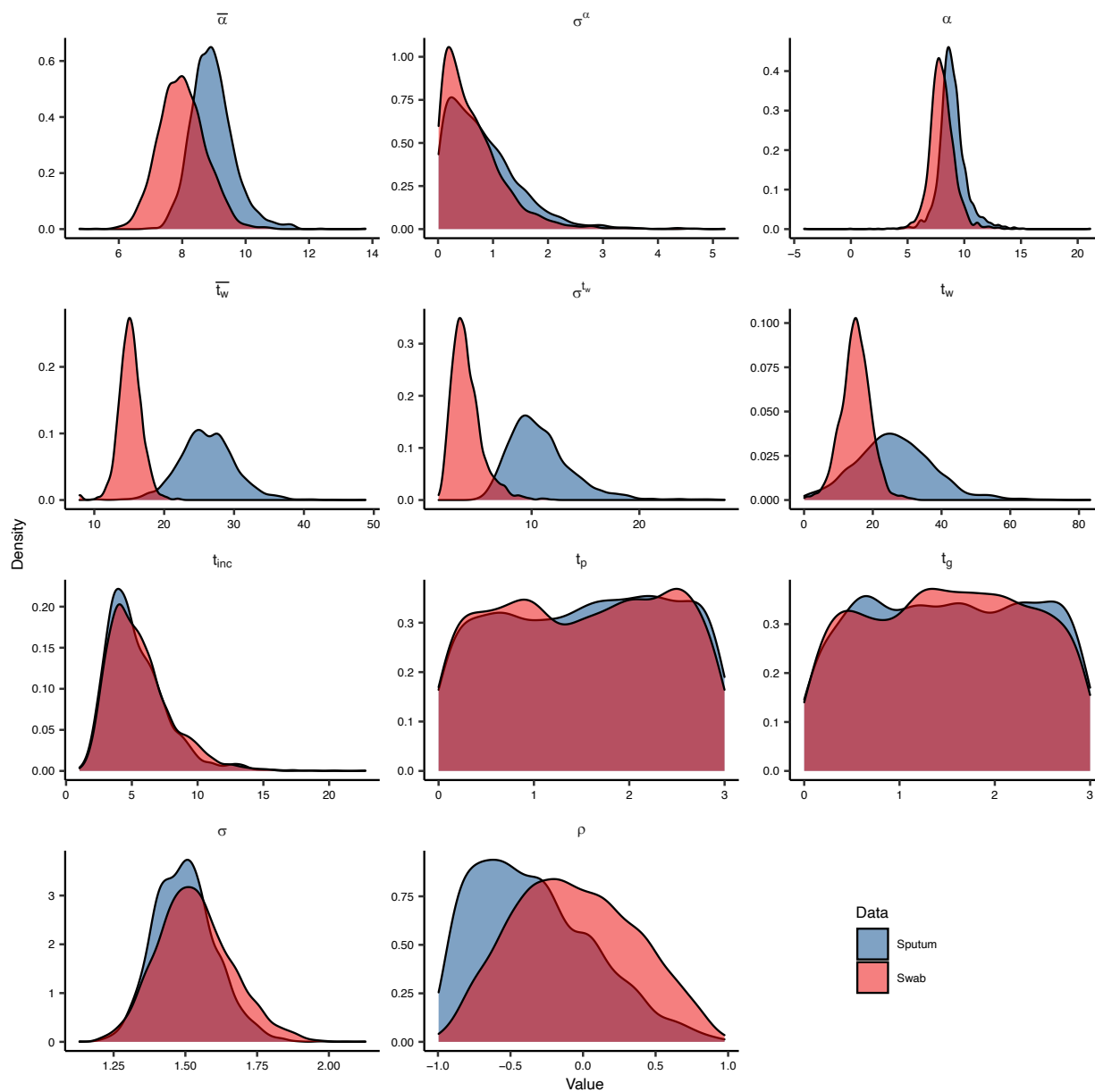
990 design is optimal at higher sample collection capacities. **(C)** Fold increase in the number of

- 991 positive samples identified relative to individual testing with the same resource constraints.
- 992 Error bar shows range amongst optimal designs.



993

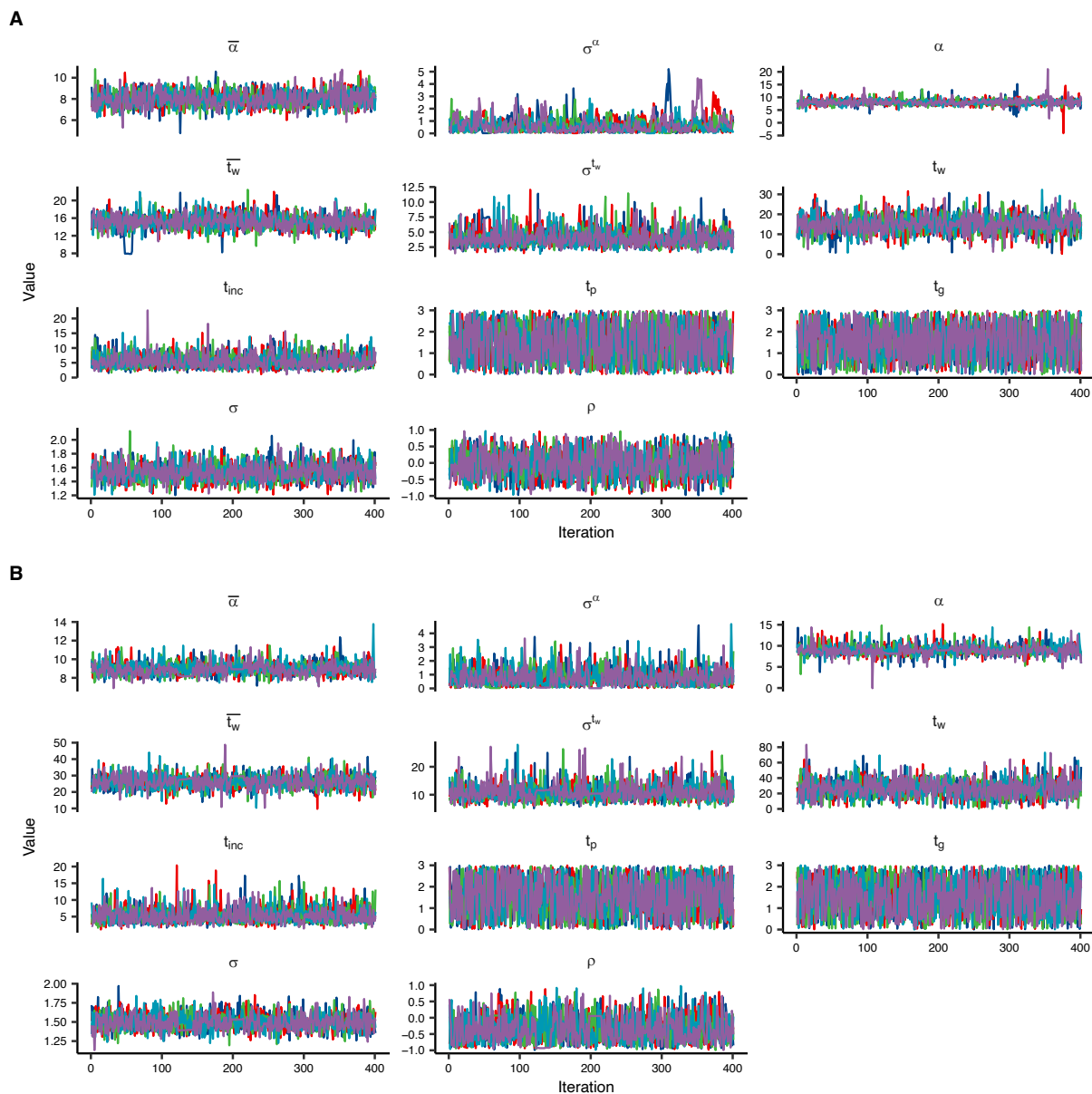
994 **Fig. S8:** Model fits to swab viral loads. Data were extracted from Wölfel et al. 2020. (34) The
995 black dots show observed \log_{10} RNA copies/swab; solid lines show posterior median
996 estimates; dark shaded regions show 95% credible intervals (CI) on model-predicted latent
997 viral loads; light shaded regions show 95% CI on simulated viral loads with added
998 observation noise. The vertical dashed line shows the time of symptom onset. The horizontal
999 dashed line shows the limit of quantification and the y-axis shows limit of detection reported
1000 by Wölfel et al. Inset text shows posterior median and 95% credible intervals of the fitted
1001 parameters.



1002

1003 **Fig. S9:** Posterior distributions of estimated parameters fitted to swab and sputum data.

1004 Parameters are described in **Table S5**.



1005

1006 **Fig. S10:** Markov chain Monte Carlo trace plots from fitting to swab and sputum data. Each
1007 color shows an independent chain. Trace plots showing the path of the Markov chain Monte
1008 Carlo sampler, demonstrating convergence to the same stationary distribution for all chains.
1009 Parameters are described in **Table S5**. **(A)** Swab data. **(B)** Sputum data.

1010 **Table S1.** List of all group test designs for sample identification.

Number of pools	Number of samples	Number of pools per sample
1	2	1
1	4	1
1	8	1
1	16	1
1	32	1
1	64	1
6	12	2
6	12	3
6	24	2
6	24	3
6	48	2
6	48	3
6	96	2
6	96	3
6	192	2
6	192	3
6	384	2
6	384	3
12	24	2
12	24	3
12	24	4
12	48	2
12	48	3
12	48	4
12	96	2
12	96	3
12	96	4
12	192	2
12	192	3
12	192	4
12	384	2
12	384	3
12	384	4
12	768	2
12	768	3
12	768	4
24	48	2

24	48	3
24	48	4
24	96	2
24	96	3
24	96	4
24	192	2
24	192	3
24	192	4
24	384	2
24	384	3
24	384	4
24	768	2
24	768	3
24	768	4
24	1536	2
24	1536	3
24	1536	4
48	96	2
48	96	3
48	96	4
48	192	2
48	192	3
48	192	4
48	384	2
48	384	3
48	384	4
48	768	2
48	768	3
48	768	4
48	1536	2
48	1536	3
48	1536	4
48	3072	2
48	3072	3
48	3072	4
96	192	2
96	192	3
96	192	4
96	384	2
96	384	3
96	384	4
96	768	2

96	768	3
96	768	4
96	1536	2
96	1536	3
96	1536	4
96	3072	2
96	3072	3
96	3072	4
96	6144	2
96	6144	3
96	6144	4

1011

1012 **Table S2:** Cycle threshold values from qPCR on pooled samples with variable viral load.
 1013 Five positive samples with variable viral loads (89,000, 12,300, 1,280, 140, and 11, for
 1014 samples 1-5, respectively) were each pooled with 23 negative samples (“Dil_1” to “Dil_5”).
 1015 Values in bottom table show the Ct values for each of five pools for three primer pairs (N1,
 1016 N2, and RP), along with a final call (positive, negative, or inconclusive).

Sample Name	N1 CT	N2 CT	RP CT	Call
24_Dil_1	28.93	29.58	28.62	POS
24_Dil_2	31.07	31.73	28.71	POS
24_Dil_3	33.29	33.91	28.37	POS
24_Dil_4	Undetermined	35.9	29.17	Inconclusive
24_Dil_5	Undetermined	Undetermined	28.62	Negative

Pool_1	30.49	31.18	28.3	POS
Pool_2	31.49	31.93	28.75	POS
Pool_3	Undetermined	Undetermined	29.62	Negative
Pool_4	31.71	31.66	29.34	POS
Pool_5	Undetermined	Undetermined	30.01	Negative
Pool_6	Undetermined	Undetermined	29.99	Negative

1017

1018 **Table S3:** Positive sample distribution within validation pools. 30 positive samples were
1019 randomly distributed across batches of combinatorial pooling (B1-B10) and randomly
1020 assigned sample numbers within batches. 24 separate positive samples were randomly
1021 assigned to 1 of 48 simple pools ("Prevalence", pool number 2...48).

Positive sample id	Batch	Sample number within batch
1	B5	46
2	B1	77
3	B5	21
4	B10	35
5	B2	10
6	B8	89
7	B1	5
8	B7	17
9	B6	6
10	B1	8
11	B7	91
12	B8	79
13	B2	20
14	B10	53
15	B8	30
16	B6	15
17	B5	50
18	B5	67
19	B9	7
20	B1	47
21	B9	6
22	B4	66
23	B2	46
24	B5	78
25	B2	96
26	B8	72

27	B3	83
28	B10	39
29	B8	20
30	B7	24
Positive sample id	Batch	Pool number
31	Prevalence	2
32	Prevalence	3
33	Prevalence	6
34	Prevalence	7
35	Prevalence	12
36	Prevalence	12
37	Prevalence	16
38	Prevalence	16
39	Prevalence	18
40	Prevalence	22
41	Prevalence	23
42	Prevalence	25
43	Prevalence	26
44	Prevalence	26
45	Prevalence	28
46	Prevalence	29
47	Prevalence	29
48	Prevalence	30
49	Prevalence	35
50	Prevalence	37
51	Prevalence	38
52	Prevalence	38
53	Prevalence	40
54	Prevalence	48

1023 **Table S4:** Pool design for combinatorial test with 96 samples. Each of 96 samples was split
1024 in equal volume into 2 out of 6 pools.

Sample	Pools	
1	A	B
2	A	B
3	A	B
4	A	B
5	A	B
6	A	B
7	A	B
8	C	D
9	C	D
10	C	D
11	C	D
12	C	D
13	C	D
14	C	D
15	E	F
16	E	F
17	E	F
18	E	F
19	E	F
20	E	F
21	E	F
22	B	C
23	B	C
24	B	C
25	B	C
26	B	C
27	B	C
28	B	C
29	D	F
30	D	F
31	D	F
32	D	F
33	D	F
34	D	F
35	D	F
36	A	E
37	A	E

38	A	E
39	A	E
40	A	E
41	A	E
42	A	E
43	B	D
44	B	D
45	B	D
46	B	D
47	B	D
48	B	D
49	A	F
50	A	F
51	A	F
52	A	F
53	A	F
54	A	F
55	C	E
56	C	E
57	C	E
58	C	E
59	C	E
60	C	E
61	B	E
62	B	E
63	B	E
64	B	E
65	B	E
66	B	E
67	C	F
68	C	F
69	C	F
70	C	F
71	C	F
72	C	F
73	A	D
74	A	D
75	A	D
76	A	D
77	A	D
78	A	D
79	B	F

80	B	F
81	B	F
82	B	F
83	B	F
84	B	F
85	D	E
86	D	E
87	D	E
88	D	E
89	D	E
90	D	E
91	A	C
92	A	C
93	A	C
94	A	C
95	A	C
96	A	C

1025

1026 **Table S5:** Description of all parameters used in the viral kinetics and transmission models.

1027 Values shown are posterior median estimates and 95% credible intervals or fixed value.

Parameter	Description	Posterior mean (95% CI)	Prior (parameters)
Viral kinetics model (swab)			
$\bar{\alpha}$	Mean of peak viral load parameters	7.98 (6.66-9.47)	Uniform(-100,100)
σ^α	Standard deviation of peak viral load parameters	0.498 (0.0254-2.13)	Half-Cauchy (1)
α_i	Individual peak viral load parameters	Varied	Multivariate normal
\bar{t}_w	Mean time from peak viral load to undetectable	15.1 (12-18.4)	Uniform(-100,100)
σ^{t_w}	Standard deviation of viral waning parameters	3.72 (2.02-7.55)	Half-Cauchy (1)
t_w	Individual time from peak viral load to undetectable	Varied	Multivariate normal
t_{inc}	Time from infection to symptom onset	5.52 (2.23-11.5)	Log normal (1.621, 0.418)
t_g	Duration of pre-viral growth latent period	Varied	Uniform(0,3)
t_p	Time from growth initiation to peak viral load	Varied	Uniform(0,3)
σ	Standard deviation of observation process	1.52 (1.31-1.82)	Uniform(0,10)
ρ	Correlation between $\bar{\alpha}$ and \bar{t}_w	-0.0808 (-0.79-0.765)	LKJ correlation matrix ($\eta=2$)
Viral kinetics model (sputum)			
$\bar{\alpha}$	Mean of peak viral load parameters	8.91 (7.8-10.5)	Uniform(-100,100)
σ^α	Standard deviation of peak viral load parameters	0.674 (0.0417-2.32)	Half-Cauchy (1)
α_i	Individual peak viral load parameters	Varied	Multivariate normal
\bar{t}_w	Mean time from peak viral load to undetectable	26.0 (18.2-34.4)	Uniform(-100,100)
σ^{t_w}	Standard deviation of viral waning parameters	10.4 (6.51-18.2)	Half-Cauchy (1)
t_w	Individual time from peak viral load to undetectable	Varied	Multivariate normal
t_{inc}	Time from infection to symptom onset	5.52 (2.23-11.5)	Log normal (1.621, 0.418)
t_g	Duration of pre-viral growth latent period	Varied	Uniform(0,3)
t_p	Time from growth initiation to peak viral load	Varied	Uniform(0,3)
σ	Standard deviation of observation process	1.49 (1.31-1.73)	Uniform(0,10)
ρ	Correlation between $\bar{\alpha}$ and \bar{t}_w	-0.398 (-0.925-0.579)	LKJ correlation matrix ($\eta=2$)
SEIR Model			
R_0	Basic reproductive number	2.5 (fixed)	NA
$1/\gamma$	Infectious period	7 days (fixed)	NA
$1/\sigma$	Incubation period	6.4 days (fixed)	NA
I_0	Initial number of infected individuals	100 (fixed)	NA
N	Total population size	12500000 (fixed)	NA

1028

1029 **Table S6:** RT-qPCR results for pooling validations. 10 batches of combinatorial pools (B1-
1030 B10, pools A-F in each batch) and 19 simple pools (“Prevalence”, pools P2...P48) were
1031 tested by RT-qPCR. Test results (POS, NEG, INCONCLUSIVE, or INVALID) and Ct values
1032 for N1, N2, and RP primers are shown for each pool.

Batch	PoolID	Result	N1 Ct	N2 Ct	RP Ct
B1	A	POS	29.39	30.37	31.36
B1	B	POS	21.32	23.31	30.48
B1	C	POS	32.76	35.59	32.12
B1	D	POS	25.39	27.54	31.34
B1	E	NEG	Undetermined	Undetermined	32.47
B1	F	NEG	Undetermined	Undetermined	32.55
B2	A	POS	20.37	21.34	30.26
B2	B	POS	17.19	18.52	31.24
B2	C	POS	27.22	27.15	30.95
B2	D	POS	22.61	23.93	31.7
B2	E	POS	26.63	28.73	32.57
B2	F	INCONCLUSIVE	Undetermined	39.46	32.5
B3	A	NEG	Undetermined	Undetermined	32.45
B3	B	NEG	Undetermined	Undetermined	32.51
B3	C	NEG	Undetermined	Undetermined	32.37
B3	D	NEG	Undetermined	Undetermined	32.88
B3	E	NEG	Undetermined	Undetermined	31.97
B3	F	NEG	Undetermined	Undetermined	33.08
B4	A	NEG	Undetermined	Undetermined	31.31
B4	B	POS	23.46	25	30.38
B4	C	NEG	Undetermined	Undetermined	32.72
B4	D	NEG	Undetermined	Undetermined	32.26
B4	E	POS	23.52	24.98	30.63
B4	F	NEG	Undetermined	Undetermined	32.65
B5	A	POS	28.68	30.06	30.56
B5	B	POS	31.23	33.89	31.51

B5	C	NEG	Undetermined	Undetermined	31.92
B5	D	POS	27.17	28.71	31.3
B5	E	NEG	Undetermined	Undetermined	32.12
B5	F	NEG	Undetermined	Undetermined	30.86
B6	A	POS	32.09	32.96	30.56
B6	B	POS	30.3	32.43	30.22
B6	C	INVALID	Undetermined	Undetermined	Undetermined
B6	D	NEG	Undetermined	Undetermined	31.76
B6	E	POS	32.86	30.42	32.41
B6	F	NEG	Undetermined	Undetermined	32.17
B7	A	POS	22.44	22.61	31.35
B7	B	POS	25.52	27.41	31.98
B7	C	POS	29.26	29.22	31.37
B7	D	NEG	Undetermined	Undetermined	32.12
B7	E	POS	25.09	26.34	31.28
B7	F	INCONCLUSIVE	Undetermined	35.75	31.71
B8	A	NEG	Undetermined	Undetermined	30.74
B8	B	NEG	Undetermined	Undetermined	31.59
B8	C	INCONCLUSIVE	Undetermined	33.89	32.4
B8	D	POS	17.85	18.76	31.24
B8	E	POS	17.71	18.18	30.57
B8	F	POS	24.73	25.65	31.05
B9	A	POS	22.44	22.76	31.27
B9	B	POS	21.17	21.96	32
B9	C	NEG	Undetermined	Undetermined	32.88
B9	D	NEG	Undetermined	Undetermined	32.04
B9	E	NEG	Undetermined	Undetermined	31.93
B9	F	POS	32.67	34.41	32.69
B10	A	POS	23.36	23.73	30.19
B10	B	NEG	Undetermined	Undetermined	31.94
B10	C	NEG	Undetermined	Undetermined	32.56
B10	D	POS	27.69	26.08	32
B10	E	POS	23.34	24.5	32.56

B10	F	POS	27.15	25.63	31.56
Prevalence	P2	POS	21.68	23.21	32.06
Prevalence	P3	POS	33.31	34.66	32.95
Prevalence	P6	POS	22.49	23.25	31.85
Prevalence	P7	POS	20.39	21.43	31.51
Prevalence	P12	POS	23.54	24.93	30.93
Prevalence	P16	POS	21.46	23.02	31.97
Prevalence	P18	POS	19.55	19.63	31.24
Prevalence	P22	POS	22.22	23.61	31.42
Prevalence	P23	POS	34.54	35.93	32.34
Prevalence	P25	INCONCLUSIVE	Undetermined	36.76	32.39
Prevalence	P26	POS	31.53	33.14	32.49
Prevalence	P28	NEG	Undetermined	Undetermined	31.93
Prevalence	P29	POS	21.2	22.61	31.64
Prevalence	P30	NEG	Undetermined	Undetermined	34.92
Prevalence	P35	POS	21.65	22.79	32.19
Prevalence	P37	POS	26.83	28.97	32.36
Prevalence	P38	POS	25.26	26.24	31.01
Prevalence	P40	POS	18.35	19.65	32.31
Prevalence	P48	POS	33.09	34.09	32.79

1033