Evaluating COVID-19 reporting data in the context of testing strategies across 31 LMICs

Mollie M. Van Gordon, Ph.D.^{1,*}, Kevin A. McCarthy, Ph.D.¹, Joshua L. Proctor, Ph.D.¹, Brittany L. Hagedorn, MBA¹

Abstract

0.1. Background

COVID-19 case counts are the predominant measure used to track epidemiological dynamics and inform policy decision-making. Case counts, however, are influenced by testing rates and strategies, which have varied over time and space. A method to consistently interpret COVID-19 case counts in the context of other surveillance data is needed, especially for data-limited settings in low- and middle-income countries (LMICs).

0.2. Methods

We leverage statistical analyses to detect changes in COVID-19 surveillance data. We apply the pruned exact linear time change detection method for COVID-19 case counts, number of tests, and test positivity rate over time.

*Corresponding author: mvangordon@idmod.org; +1 (425) 526-3104

¹Institute for Disease Modeling at the Bill & Melinda Gates Foundation, 500 5th Ave N, Seattle, WA, 98109, USA

With this information, we categorize change points as likely driven by epidemiological dynamics or non-epidemiological influences such as noise.

0.3. Findings

Higher rates of epidemiological change detection are more associated with open testing policies than with higher testing rates. We quantify alignment of non-pharmaceutical interventions with epidemiological changes. LMICs have the testing capacity to measure prevalence with precision if they use randomized testing. Rwanda stands out as a country with an efficient COVID-19 surveillance system. Sub-national data reveal heterogeneity in epidemiological dynamics and surveillance.

0.4. Interpretation

Relying solely on case counts to interpret pandemic dynamics has important limitations. Normalizing counts by testing rate mitigates some of these limitations, and open testing policy is key to efficient surveillance. Our findings can be leveraged by public health officials to strengthen COVID-19 surveillance and support programmatic decision-making.

0.5. Funding

This publication is based on models and data analysis performed by the Institute for Disease Modeling at the Bill & Melinda Gates Foundation.

Keywords: COVID-19, Change detection, Disease surveillance

Research in Context

Evidence before this study

We searched for articles on the current practices, challenges, and proposals for COVID-19 surveillance in LMICs. We used Google Scholar with search terms including “COVID surveillance.” Existing studies were found to be qualitative, anecdotal, or highly location-specific.

Added value of this study

We developed a quantitative method that makes use of limited information available from LMICs. Our approach improves interpretation of epidemiological data and enables evaluation of COVID-19 surveillance dynamics across countries.

Implications of all the available evidence

Our results demonstrate the importance of open testing for strong surveillance systems, bolstering existing anecdotal evidence. We show strong alignment across LMICs between non-pharmaceutical interventions and epidemiological changes. We demonstrate the importance of considering sub-national heterogeneity of epidemiological dynamics and surveillance.

2 1. Introduction

3 The virus known as SARS-CoV-2 was first identified in Wuhan, China in
 4 December 2019. Since then, countries have scrambled to monitor the severity
 5 and trajectory of the COVID-19 outbreak and to control its progression using
 6 non-pharmaceutical interventions (NPIs). Disease surveillance has mostly
 7 relied on case counts to inform public health policies (WHO 2020). There
 8 has not, however, been a robust evaluation of case counts as a metric for
 9 epidemiological dynamics, nor the varied surveillance approaches used to
 10 track disease trajectories.

11 Case-based surveillance systems have known weaknesses, including the strong
 12 influence of testing rates, which vary widely across space and time (Haider
 13 et al. 2020). Case counts can be inconsistently measured, testing capacity
 14 limited, and eligibility policies variable. It is critical to understand the lim-
 15 itations of available data and to identify metrics that are robust to these
 16 challenges, particularly for low- and middle-income countries (LMICs).

17 There is general recognition that surveillance system performance can be
 18 a challenge in LMICs, and that understanding disease surveillance is key
 19 to system improvement and production of representative data (Petti et al.
 20 2006). Existing efforts to evaluate LMIC surveillance systems, however, are
 21 largely qualitative, country-specific, or based on commentary (Ibrahim 2020;
 22 Farahbakhsh et al. 2020; Alwan 2020). Further, most national-level studies
 23 of NPI impacts focus on high-income countries, but there is evidence that

24 these insights cannot be readily generalized to LMIC settings (Brauner et al.
25 2020; Flaxman et al. 2020; Hsiang et al. 2020; Dehning et al. 2020; Chen
26 et al. 2020; Islam et al. 2020; Haider et al. 2020). This leaves an important
27 knowledge gap in understanding how to evaluate and interpret COVID-19
28 epidemiological data from LMICs.

29 To address the gap in systematic interpretation and evaluation methods, we
30 leverage statistical analysis techniques to detect changes in underlying prop-
31 erties of COVID-19 time series surveillance data across 31 LMICs. With this
32 information, we categorize detected change points as likely driven by epi-
33 demiological changes or non-epidemiological influences such as noise. This
34 provides a quantitative and automated approach to analyzing epidemiolog-
35 ical surveillance data. We make use of imperfect information despite data
36 weaknesses, deriving insights from information available in LMICs that may
37 otherwise be overlooked. The approach is fast and highly portable, well
38 suited to looking across countries, and has minimal data requirements.

39 In this article, we first present the methods for our analysis, including the
40 statistical model, change point categorization, and evaluation of epidemio-
41 logical change co-occurrence with NPIs. We follow with validation of our
42 method, the usefulness of open testing, comparisons of country surveillance
43 characteristics, and consideration of sub-national dynamics. Finally, we elab-
44 orate on the significance of our results, broader conclusions, and relevance
45 for public health applications.

46 **2. Methods**

47 The methods are outlined in Figure 1 for two example countries: South Africa
48 and Bangladesh. Details about each step are presented in the following sub-
49 sections.

Figure 1: Methods Overview. Time series for cases (orange), tests (blue) and positivity (green) for case study countries South Africa and Bangladesh. Cases and tests are plotted in units per thousand people. Vertical lines indicate detected change points on each series. National changes in testing policy are shown as blue bars; see Section 2.1 for policy descriptions. Positivity change points are overlaid with case and test changes. Change points from the three time series are grouped in time; shading on positivity changes indicates grouping tolerance. Category labels for change point groups are shown above positivity and described in Figure 2. Black arrows indicate NPI changes; arrow direction indicates increase or decrease in stringency. For categories D and E, Y(es) and N(o) in boxes below positivity indicate whether there is a co-occurring NPI change inverse to the change in slope of positivity.

50 2.1. Data

51 We use national-level case and testing data as well as records on national
 52 policies for testing and NPIs (Roser et al. 2020; Hale et al. 2020). We cal-
 53 culate test positivity by dividing cases by tests. Testing policy is indicated
 54 by ordinal values: zero indicates no testing policy; one indicates testing of
 55 those with symptoms who meet specific criteria (e.g. known contact with a
 56 positive individual); two indicates testing of any symptomatic individuals;
 57 three indicates open public testing. For South Africa, we also use provincial-
 58 level data on COVID-19-confirmed deaths, cases, tests, and excess mortality
 59 (Mkhize 2020; National Institute for Communicable Diseases 2020; Statistics
 60 South Africa 2020; Bradshaw et al. 2020).

61 We selected countries for analysis based on three conditions: available case
 62 data, available testing data, and human development index (HDI) score. Of
 63 those with data, we included the countries in the lowest third of HDI score,
 64 all of which are considered low- or middle-income in 2020-2021 by the World
 65 Bank. All data used in this research are public. Further details on data and
 66 definitions found in Appendix A.1.

67 2.2. Change point detection

68 2.2.1. Pruned Exact Linear Time (PELT) change detection

69 Change point detection is a set of approaches for identifying points in time
 70 where the statistical properties of a time series change (Truong et al. 2020).

71 We apply change point detection to epidemiological time series (cases, tests,
72 and positivity) and national policy time series; details in Appendix A.2.
73 Without a priori knowledge of the appropriate number of changes, the PELT
74 algorithm must be assigned a penalty for the number of changes to iden-
75 tify. In the absence of an established method for this parameterization when
76 working across time series, we developed a novel systematic approach for
77 penalty selection which enables comparison among time series and countries;
78 described in Appendix A.3.

79 *2.2.2. Method validation*

80 We apply PELT to synthetic case count data generated by the stochastic
81 agent-based COVID-19 simulator (COVASIM), in order to test PELT as a
82 robust method for change detection in epidemiological time series (Kerr et al.
83 2021). The model scenario inputs include step-wise changes in contacts per
84 person per time which represent NPI implementation, as well as a change
85 in testing policy from symptomatic to asymptomatic testing. The model
86 generates a simulated time series of cases and tests per thousand people,
87 from which we calculate a positivity time series. The change point detection
88 methods described above are applied to the seven-day mean of the time series
89 to align with the data smoothing used with our empirical time series.

90 2.3. *Change type categorization*

91 Change detection identifies changes that may be related to data quality,
 92 stochasticity, and testing dynamics, in addition to epidemiological changes.
 93 We classify the likely cause of changes identified by the PELT algorithm based
 94 on the co-occurrence of changes from different time series. This categoriza-
 95 tion simplifies the interpretation of epidemiological surveillance, separates
 96 signal from noise, and enables broad comparison across countries and testing
 97 dynamics.

98 We combine detected change points across cases, tests, and positivity time
 99 series to create change point groups. The tolerance for temporal association
 100 is set at $\pm$ seven days to account for seven-day smoothing and weekly data
 101 reporting practices. These change groups are then categorized as shown in
 102 Figure 2, with details of the interpretation described in Appendix B. To
 103 capture all changes that may be epidemiological, we include both categories
 104 D and E as epidemiological change in our analysis. We note that these
 105 categories are heuristically defined, however they are informed both by val-
 106 idation using the COVASIM simulations and a qualitative understanding of
 107 epidemiological surveillance dynamics.

Label	Constituent time series	Description	Origin
(A)	C / T / P	Single variable change	Data issues/noise
(B)	C + T	Cases and tests move together	Likely non-epi
(C)	T + P	Tests drive positivity change	Likely non-epi
(D)	C + P	Cases drive positivity change	Likely epi
(E)	C + T + P	Cases, tests, positivity change	Confounded

Figure 2: Summary of change group categories as determined by their constituent time series changes. Cases, tests, and positivity time series are indicated as orange, blue, and green, respectively. Details of the category interpretations are described in Appendix B.

108 2.4. NPI alignment

109 Change points classified as epidemiological are then assessed for whether they
110 are associated with NPI changes. Timings of known NPIs in the empirical
111 data are lagged by nine days to account for virus incubation time and the
112 delay from symptom onset to test-seeking (Qin et al. 2020). We consider a
113 change point to be aligned with an NPI when two conditions are met: 1)
114 an epidemiological change co-occurs with an offset NPI and 2) the change
115 in NPI stringency is inverse to the concurrent change in positivity slope.
116 The second condition includes occasions when stringency was increased and
117 positivity decreased, as well as occasions when stringency decreased and pos-
118 itivity increased.

119 3. Results

120 *3.1. Synthetic modeling validates Pruned Exact Linear Time (PELT) is a*
 121 *robust method for change detection in epidemiological time series.*

122 Before applying the PELT method to the surveillance data, we validate its
 123 applicability for epidemiological systems. We apply PELT change detection
 124 to data from the transmission model described in Section 2.2.2. The sensi-
 125 tivity of change point detection to parameterization is illustrated in the top
 126 two case rate time series of Figure 3. The bottom positivity time series shows
 127 the detected change points for all time series, parameterized by the method
 128 described in Appendix A.3. PELT successfully identifies step changes in NPI
 129 and testing policies, as well as slope changes in cases, tests, and positivity.
 130 Further, the categories of change point groups are correctly identified in line
 131 with our classification scheme, labeled on the positivity time series in black
 132 and purple boxes and described in Figure 2.

Figure 3: Synthetic model time series; detected change points shown as vertical lines. Upper plot shows detected change points in the cases time series using a high penalty, which promotes sparse change point detection. Middle plot shows detected cases change points in the same time series as above, but using a low penalty. Bottom plot shows positivity time series overlaid with detected change points from all time series, as indicated in the change point key. Change group categories are indicated in boxes above plot as described in Figure 2.

133 3.2. Testing rates and policies impact how surveillance measures should be
134 interpreted.

135 We illustrate the relevance of testing rates and the influence of testing pol-
136 icy using time series for Bangladesh in the context of local events (Figure 1).
137 Cases peaked in early July, an apparent epidemiological turning point if cases
138 were considered alone. Simultaneously, however, there was a new policy im-
139 plemented to charge for testing and thus a decline in testing (Cousins 2020).
140 This resulted in no change in positivity and contradicts the interpretation

141 of the case reduction as a declining outbreak. Similarly, the dip in case rate
142 in early August was accompanied by a dip in testing during the Eid al-Fitr
143 holiday; again there is no change in positivity.

144 While this recommends positivity as a surveillance metric instead of case
145 counts alone, further consideration of testing policy complicates the picture.
146 Test eligibility in Bangladesh is based on symptoms rather than open testing,
147 meaning that positivity is influenced by the prevalence of both COVID-19
148 and other respiratory illnesses. This limits the potential for positivity to de-
149 tect epidemiological changes, and indeed, the positivity curve for Bangladesh
150 is largely flat. An elaboration of COVID-19 surveillance considerations ap-
151 pears in Appendix C.

152 *3.3. Epidemiological change detection is more influenced by testing policy*
153 *than by testing rate.*

154 For all 31 LMICs in our dataset, we apply PELT change detection and change
155 point categorization. We quantify surveillance system efficiency as the per-
156 centage of all detected change points classified as epidemiological, i.e. epi-
157 demiological change detection rate. We compare linear fits of epidemiological
158 change detection by tests per thousand people and by testing policy (Fig-
159 ure 4). Results indicate that the ability to identify epidemiological changes
160 has a stronger relationship with testing policy than with tests per thousand
161 people. Open testing is the only testing policy bin with a mean or median
162 epidemiological change detection rate as high as 50%, but with a wide range,

163 indicating that open testing policy is necessary but not sufficient for quality
164 surveillance (with outlier exceptions).

Figure 4: Percent of detected changes categorized as epidemiological for each country by tests per thousand people (left) and binned by testing policy (right) at the time of change detection. Linear regression shown as dotted line on left. Box and whisker plots on right show quartiles, range, and median with means plotted as gray diamonds. Note that binned calculations cause the maximum epidemiological change detection rate to differ between the two plots.

165 Further, LMICs have the testing capacity to measure prevalence with pre-
166 cision. Based on the 95th percentile of their daily testing rates, nearly all
167 LMIC countries could measure down to 1% prevalence with a margin of er-
168 ror no larger than $\pm 1\%$ if testing were randomly sampled (Figure 5). Only
169 three countries hover around the margin of error to prevalence ratio of one:

170 Malawi, the Democratic Republic of Congo, and Togo. Note that true ran-
 171 dom sampling is difficult to achieve in any setting, but open testing policies
 172 can approximate random sampling more closely than symptomatic testing.

Figure 5: Margin of error for random sampling of 1% prevalence plotted by 95th percentile of national testing rate and population of each country in our dataset. See Appendix D for details on standard error calculations.

173 3.4. Change detection rates and NPI alignment frequency vary across LMICs.

174 Figure 6A shows the wide variation of epidemiological change detection rates
 175 across LMIC countries, with Rwanda the highest and Ethiopia the lowest.
 176 The percentage of NPIs that are aligned with a detected epidemiological
 177 change is shown in Figure 6B, again led by Rwanda. Rwanda performs well

178 by these metrics regardless of change detection parameterization, Appendix
179 A.4. Nearly all countries in our analysis show at least one detected epi-
180 demiological change. Conversely, approximately half of the countries in our
181 analysis show zero alignment of any type of NPI with an epidemiological
182 change, although the number of NPIs implemented in these countries spans
183 a wide range.

Figure 6: Epidemiological change detection rates (A) and NPI alignment rates (B) by country.

184 *3.5. NPI alignment with detected epidemiological changes is bimodal and sig-*
 185 *nificant.*

186 We test the significance of NPI alignment with detected epidemiological
 187 changes through comparison with alignment rates when NPIs are assigned a
 188 random date. We find all types of NPIs measured in this study have signifi-
 189 cant rates of alignment with epidemiological changes when the zero-alignment
 190 mode is excluded (max p-value = 1.38e-10). We calculate the distributions
 191 of random NPI alignment by re-assigning random dates to NPIs by type and
 192 then finding alignment rates for N=150 bootstrapping. Across NPIs, the
 193 rate of random NPI alignment with epidemiological change has a mean of
 194 11.6% and a standard deviation of 3.64% (gray violin distributions, Figure
 195 7). When analyzed by country, nearly all NPI alignment rates are either
 196 higher or lower than the random date distributions (cyan circles, Figure 7).
 197 This indicates two modes of detected NPI alignment. Excluding the mode of
 198 zero NPI alignment, mean NPI alignment ranges from 50% for restrictions on
 199 internal movement, to 33% for restrictions on gathering (black squares, Fig-
 200 ure 7). Differences in alignment rates between NPI types are not significant.
 201 Potential differences in NPI alignment rates are confounded by synchronous
 202 implementation of NPIs, although there is some evidence to support the ef-
 203 fect strength of workplace closing and stay at home requirements (Appendix
 204 E).

Figure 7: Percentage of each type of NPI that is aligned with a detected epidemiological change. Cyan circles are data for individual countries, sized by total number of NPIs by type and country. Gray violins are distributions of NPI alignment for random NPI dates. Black squares indicate weighted mean of country data for NPI alignment greater than zero; error bars indicate standard error. NPIs include both easing and tightening of policy restrictions.

205 3.6. National-level results obscure sub-national heterogeneity in epidemiolog- 206 ical dynamics and surveillance.

207 To investigate sub-national heterogeneity, we conduct the same analyses as
208 above, but at the province level in South Africa. Figure 8A shows substantial
209 variability in provinces both by NPI alignment rate and by epidemiological
210 change detection rate. In line with results from national-level data, epidemi-
211 ological change detection rate is not correlated with mean tests per thousand
212 people. Because of reporting limitations, the NPIs here are national policies
213 only.

214 We select three edge cases from the scatter plot in Figure 8A (Limpopo,
215 Northern Cape, and Western Cape) to compare time series of positivity,
216 COVID-19-confirmed deaths, and total estimated excess mortality (Figure
217 8B). The differences in the timing and trajectories of the time series illustrate
218 strong sub-national variability in underlying epidemiological dynamics that
219 are may be overlooked when time series are aggregated to the national level.

220 Variation among provinces in the difference in magnitude between excess
221 mortality and COVID-19 deaths points to differences in their surveillance
222 systems. Western Cape is the only province where the magnitude of excess
223 deaths resembles that of COVID-19-confirmed deaths throughout the time
224 series. In Northern Cape, the peak of excess deaths is roughly a factor of
225 three higher than the COVID-19-confirmed deaths, suggesting substantial
226 under-reporting.

Figure 8: A) South African provinces by aligned NPI fraction versus epidemiological detection rate with color indicating the mean over time of tests per thousand people. B) Time series from three example provinces. Positivity shown in green on left y-axis. Deaths per thousand people shown on the right y-axis: excess mortality in black; COVID-19-confirmed deaths in brown.

227 4. Discussion

228 We have demonstrated a standardized and quantitative approach to ana-
 229 lyzing epidemiological surveillance time series that can be automated for
 230 improved interpretation and comparison between countries. We find that in-
 231 terpretation of epidemiological trajectories are more informative when cases
 232 are normalized by tests and highlight the disadvantages of symptomatic test-
 233 ing for outbreak tracking and public health purposes. These findings align
 234 with literature emphasizing the importance of positivity and test sampling
 235 strategies (Pearce et al. 2020; Hilborne et al. 2020). Our finding of strong
 236 alignments of NPIs with epidemiological changes is consistent with existing
 237 literature on global NPI impacts (Islam et al. 2020; Haug et al. 2020; Liu
 238 et al. 2020). When we apply our analysis of change types to evaluate the
 239 efficiency of national surveillance systems, we find that Rwanda stands out as
 240 a country with a strong surveillance system, which is consistent with current
 241 qualitative evaluation (WHO Regional Office, Africa 2020).

242 Our approach substantially broadens the scope of previous analyses of COVID-
 243 19 surveillance data in LMICs. We use statistical change detection methods
 244 on COVID-19 surveillance time series from 31 LMICs to differentiate epi-
 245 demiological changes from changes related to stochasticity, data quality and
 246 non-epidemiological dynamics. This maximizes the insights gained from lim-
 247 ited data, reduces erroneous interpretations of epidemiological time series,
 248 and enables quantitative comparisons of disease surveillance approaches. We

249 use epidemiological change detection rate as a proxy for surveillance system
250 efficiency, and show that epidemiological change detection is not as strongly
251 associated with testing rate as with open testing policies. There is substan-
252 tial variation in epidemiological and surveillance dynamics across countries
253 and in our sub-national analysis.

254 There are limitations in our analysis related to the data themselves as well
255 as our methods. Simultaneously, these data challenges are precisely the mo-
256 tivation for developing our methods: maximizing information with limited
257 data. Our data are potentially biased by unmeasured factors such as fluctu-
258 ations in testing capacity and undocumented population sampling strategies
259 over time, delays and temporal uncertainty due to reporting systems, and
260 incentives for case-finding. Defining co-occurrence when working with im-
261 precise time series is a challenge, which we partially mitigate by considering
262 uncertainty bounds when defining change groups. We emphasize, of course,
263 that co-occurrence does not establish causality. In PELT change detection,
264 the changes detected are influenced by the choice of the sparsity parameter.
265 In a sensitivity analysis of our novel parameterization approach, however,
266 we find that Rwanda remains the leader in surveillance system performance,
267 regardless of the parameterization choice.

268 Results from this analysis highlight that surveillance data must be used care-
269 fully to ensure proper programmatic responses. As a sufficient and less
270 resource-intensive approximation of random sampling, open testing would
271 enable better estimation of disease prevalence and examination of NPI im-

272 pacts in geographies without reliable hospitalization data, death records, or
 273 seroprevalence surveys. NPIs without epidemiological changes may indicate
 274 inefficacy of policy, but may also indicate shortfalls of surveillance systems,
 275 which undermines policy makers' ability to make evidence-based decisions.
 276 Our methods could be further developed and applied not just to COVID-19
 277 but also to surveillance interpretation for other poorly measured diseases, en-
 278 abling more informed decision-making and targeted improvements in surveil-
 279 lance systems.

280 **5. Declarations**

281 *5.1. Declaration of interests*

282 The authors declare no conflicts of interest.

283 *5.2. Role of funding source*

284 This publication is based on models and data analysis performed by the
 285 Institute for Disease Modeling at the Bill & Melinda Gates Foundation. The
 286 funder had no influence on the analysis or conclusions presented here.

287 *5.3. Ethical approval*

288 Not required.

289 REFERENCES

- 290 Alwan NA. Surveillance is underestimating the burden of the COVID-19 pan-
291 demic. The Lancet 2020;396:e24. [https://doi.org/10.1016/S0140-6736\(20\)31823-](https://doi.org/10.1016/S0140-6736(20)31823-7)
292 7.
- 293 Bradshaw D, Laubscher R, Dorrington R, Groenewald P, Moultrie T. Report
294 on Weekly Deaths in South Africa 2020.
- 295 Brauner JM, Mindermann S, Sharma M, Johnston D, Salvatier J, Gavenčiak
296 T, et al. Inferring the effectiveness of government interventions against
297 COVID-19. Science 2020. <https://doi.org/10.1126/science.abd9338>.
- 298 Chen X, Qiu Z. Scenario analysis of non-pharmaceutical interventions on
299 global COVID-19 transmissions. Covid Economics 2020.
- 300 Cousins S. Bangladesh's COVID-19 testing criticised. The Lancet 2020;396:591.
301 [https://doi.org/10.1016/S0140-6736\(20\)31819-5](https://doi.org/10.1016/S0140-6736(20)31819-5).
- 302 Dehning J, Zierenberg J, Spitzner FP, Wibral M, Neto JP, Wilczek M, et al.
303 Inferring change points in the spread of COVID-19 reveals the effectiveness of
304 interventions. Science 2020;369. <https://doi.org/10.1126/science.abb9789>.
- 305 Farahbakhsh M, Fakhari A, Azizi H, Davtalab-Esmaeili E. Structure, Char-
306 acteristics and Components of COVID-19 Surveillance System. J Mil Med
307 2020;22. <https://doi.org/10.30491/JMM.22.6.534>.

- 308 Flaxman S, Mishra S, Gandy A, Unwin HJT, Mellan TA, Coupland H, et al.
309 Estimating the effects of non-pharmaceutical interventions on COVID-19 in
310 Europe. *Nature* 2020;584:257–61. [https://doi.org/10.1038/s41586-020-2405-](https://doi.org/10.1038/s41586-020-2405-7)
311 7.
- 312 Haider N, Osman AY, Gadzekpo A, Akipede GO, Asogun D, Ansumana R,
313 et al. Lockdown measures in response to COVID-19 in nine sub-Saharan
314 African countries. *BMJ Global Health* 2020;5:e003319.
315 <https://doi.org/10.1136/bmjgh-2020-003319>.
- 316 Hale T, Boby T, Angrist N, Cameron-Blake E, Hallas L, Kira B, et al. Oxford
317 COVID-19 Government Response Tracker 2020.
- 318 Haug N, Geyrhofer L, Londei A, Dervic E, Desvars-Larrive A, Loreto V, et al.
319 Ranking the effectiveness of worldwide COVID-19 government interventions.
320 *Nature Human Behaviour* 2020;4:1303–12. [https://doi.org/10.1038/s41562-](https://doi.org/10.1038/s41562-020-01009-0)
321 020-01009-0.
- 322 Hilborne LH, Wagner Z, Cabrerios I, Brook RH. Linking Statistics With Test-
323 ing Policy to Manage COVID-19 in the Community. *Am J Clin Pathol*
324 2020;154:142–8. <https://doi.org/10.1093/ajcp/aqaa099>.
- 325 Hsiang S, Allen D, Annan-Phan S, Bell K, Bolliger I, Chong T, et al. The
326 effect of large-scale anti-contagion policies on the COVID-19 pandemic. *Na-*
327 *ture* 2020;584:262–7. <https://doi.org/10.1038/s41586-020-2404-8>.
- 328 Ibrahim NK. Epidemiologic surveillance for controlling Covid-19 pandemic:

types, challenges and implications. *Journal of Infection and Public Health* 2020;13:1630–8. <https://doi.org/10.1016/j.jiph.2020.07.019>.

Islam N, Sharp SJ, Chowell G, Shabnam S, Kawachi I, Lacey B, et al. Physical distancing interventions and incidence of coronavirus disease 2019: natural experiment in 149 countries. *BMJ* 2020;370:m2743. <https://doi.org/10.1136/bmj.m2743>.

Kerr CC, Stuart RM, Mistry D, Abeysuriya RG, Rosenfeld K, Hart GR, et al. Covasim: an agent-based model of COVID-19 dynamics and interventions. *MedRxiv* 2021:2020.05.10.20097469. <https://doi.org/10.1101/2020.05.10.20097469>.

Killick R, Fearnhead P, Eckley IA. Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association* 2012;107:1590–8. <https://doi.org/10.1080/01621459.2012.737745>.

Liu Y, Morgenstern C, Kelly J, Lowe R, Group CC-19 W, Jit M. The impact of non-pharmaceutical interventions on SARS-CoV-2 transmission across 130 countries and territories. *MedRxiv* 2020:2020.08.11.20172643. <https://doi.org/10.1101/2020.08.11.20172643>.

Mkhize Z. Latest Confirmed Cases of COVID-19 in South Africa (1 Nov 2020) 2020. NICD. Weekly Testing Summary 2020.

Ouedraogo B, Inoue Y, Kambiré A, Sallah K, Dieng S, Tine R, et al. Spatio-temporal dynamic of malaria in Ouagadougou, Burkina Faso, 2011–2015.

350 Malar J 2018;17:138. <https://doi.org/10.1186/s12936-018-2280-y>.

351 Pearce N, Vandenbroucke JP, VanderWeele TJ, Greenland S. Accurate Statis-
352 tics on COVID-19 Are Essential for Policy Guidance and Decisions. Am J
353 Public Health 2020;110:949–51. <https://doi.org/10.2105/AJPH.2020.305708>.

354 Petti CA, Polage CR, Quinn TC, Ronald AR, Sande MA. Laboratory Medicine
355 in Africa: A Barrier to Effective Health Care. Clinical Infectious Diseases
356 2006;42:377–82. <https://doi.org/10.1086/499363>.

357 Qin J, You C, Lin Q, Hu T, Yu S, Zhou X-H. Estimation of incubation period
358 distribution of COVID-19 using disease onset forward time: A novel cross-
359 sectional and forward follow-up study. Science Advances 2020;6:eabc1202.
360 <https://doi.org/10.1126/sciadv.abc1202>.

361 Roser M, Ritchie H, Ortiz-Ospina E, Hasell J. Coronavirus Pandemic (COVID-
362 19) 2020.

363 Sissoko MS, Sissoko K, Kamate B, Samake Y, Goita S, Dabo A, et al. Tem-
364 poral dynamic of malaria in a suburban area along the Niger River. Malaria
365 Journal 2017;16:420. <https://doi.org/10.1186/s12936-017-2068-5>.

366 Statistics South Africa. P0302 - Mid-year population estimates, 2020 2020.

367 Truong C, Oudre L, Vayatis N. Selective review of offline change point de-
368 tection methods. Signal Processing 2020;167:107299.
369 <https://doi.org/10.1016/j.sigpro.2019.107299>.

370 WHO. COVID-19 Weekly Epidemiological Update, 15 Dec 2020. World Health
371 Organization; 2020.

372 WHO Regional Office A. COVID-19 in Rwanda: A country's response. WHO
373 — Regional Office for Africa 2020.

374 **Appendix A. Methods**

375 *Appendix A.1. Data and definitions*

376 The case rate is defined as the number of individuals confirmed positive
377 for the SARS-CoV-2 virus per population, regardless of symptoms. The
378 testing rate per population is defined as the number of people tested (i.e.
379 excluding duplicate confirmatory tests) divided by the population, regardless
380 of the test outcome. To address the dependence of case rate on testing rate,
381 we normalize case counts by the number of tests conducted, creating the
382 alternate metric of test positivity rate.

383 For the purposes of comparing between countries and over time, we define the
384 ‘mean testing policy’ as the average over time of the ordinal value represent-
385 ing the national testing policy. Thus, lower values represent more restricted
386 testing over longer periods of time. Social distancing policies tracked in the
387 dataset include the following: closing schools, closing workplaces, cancelling
388 public events, restricting gathering sizes, closing public transport, stay at
389 home requirements, restricting in-country mobility, and restricting interna-

390 tional travel.

391 Weekly cases, testing, and death data are interpolated using a cubic spline.
 392 All daily cases, testing and death data are smoothed using a centered seven-
 393 day rolling average. Error bars on plots show standard error.

394 *Appendix A.2. PELT change detection*

395 The naive approach to generating an exact solution to time series segmenta-
 396 tion is to test all possible solutions. For an unknown number of changes, this
 397 also requires testing a sufficiently large set of possible number of changes.
 398 We use the Pruned Exact Linear Time (PELT) change detection method to
 399 address these computational tractability issues.

400 PELT minimizes the sum of costs from a criterion function across time series
 401 segments while balancing model complexity by implementing a linear penalty
 402 function and change point pruning. At each iteration of cost minimization
 403 for a potential set of change points, time points that cannot be a global min-
 404 ima are removed from future consideration. The PELT method, developed
 405 with applications in genetics and finance in mind, is increasingly used for
 406 climate and epidemiological applications (Killick et al. 2012; Sissoko et al.
 407 2017; Ouedraogo et al. 2018).

408 To detect changes in slope of the epidemiological time series, we use the
 409 first derivative as input for the PELT algorithm. For detection of discrete
 410 step changes of policy time series, we feed the data directly into the change

411 detection algorithm without taking a derivative. For all time series, we use
412 the radial basis function kernel for the PELT detection algorithm.

413 *Appendix A.3. PELT parameterization*

414 To date, there is no established method for parameterizing the PELT change
415 density penalty across time series when the number of changes is not known.
416 One of the ways to choose penalty values across time series would be to unify
417 the number of changes detected in each time series. This, however, imposes
418 the assumption that all time series exhibit the same general change frequency
419 and it is only the point in time of a change that is unknown, rather than the
420 number of changes.

421 We present here a novel approach for systematic parameterization when iden-
422 tifying an unknown number of changes in slope over many time series, as in
423 our case with multiple epidemiological time series across countries. To ac-
424 complish this, we first conduct change detection in a sweep over parameter
425 space. The change points detected using a given value in parameter space
426 slice the time series into segments, each of which is input into a linear re-
427 gression. The standard error for each of those linear regressions is calculated
428 and then averaged, weighted by segment length.

429 The mean standard error associated with each penalty value, when plotted
430 over parameter space, is characterized by a series of plateaus that correspond
431 to plateaus in the number of changes found with each penalty value, Figure

432 A.9, top row. Descending through penalty values in the penalty parameter
433 space, the lowest penalty associated with each plateau is selected to represent
434 that plateau.

435 Each time series is thus associated with a sparse set of penalty values, ordered
436 from largest penalty (low change point density) to smallest penalty (high
437 change point density). The penalty values are unique to each time series,
438 but represent the same ordered progression of plateaus. To illustrate, change
439 detection with different ranked penalties for South Africa and Bangladesh
440 are shown in green in Figure A.9.

441 Penalty values for each unique time series can then be chosen based on their
442 order in the ranked plateau list. This enables a principled approach to pa-
443 rameterization that creates change density parity across time series, allowing
444 for the likelihood that some time series are characterized by more changes
445 than others. Among all time series and countries in our analysis, the min-
446 imum number of plateaus detected is four. We therefore choose the fourth
447 penalty value for all time series.

Figure A.9: Top row: change detection results over parameter space for the positivity time series of South Africa (left) and Bangladesh (right). For each penalty value, the associated number of changes is plotted in purple on the left y-axis and the mean standard error of linear regressions of the time series segments are plotted on the right y-axis. The parameter values selected to represent plateaus are shown as brown dotted vertical lines. Bottom row: positivity time series for South Africa (left) and Bangladesh (right) plotted with detected changes as vertical lines for each of the four penalty values selected to represent plateaus in the top row.

448 *Appendix A.4. Parameterization sensitivity analysis*

449 To evaluate the influence of penalty selection on our analysis results, we con-
 450 duct a parameterization sensitivity analysis. We compare results of country
 451 ranking by epidemiological change detection rate for different penalty plateau
 452 selections. Skipping penalty rank one for which no changes may be detected
 453 (see examples in Figure A.9), we find that regardless of which penalty rank
 454 we use, Rwanda appears at the top of the list with the highest epidemiological
 455 change detection rate.

Figure A.10: Countries ranked by epidemiological change detection rate, as in Figure 6, shown for different choices of penalty parameterization. Order top to bottom follows order of time series of Figure A.9, without rank one for which there are often no changes.

456 Appendix B. Change categorization

457 Appendix B.1. Heuristic interpretation

458 Detected change points across cases, testing, and positivity time series are
459 combined into groups by temporal co-occurrence. These groups are then

categorized by their constituent time series, Figure 2. Dynamical interpretation of the constituent time series aids in the characterization of each change group category, as follows:

A) Single variable change: Because positivity is defined according to the arithmetic relationship, $Positivity = Cases/Tests$, a change in any one of the variables should be accompanied by a change in at least one of the other variables. A single change in only one of the variables indicates that the change arises from issues in the data or noise. These single variable changes often occur early in the time series, when the numbers of cases and tests are smaller, signal to noise ratios are lower, and confidence intervals are larger.

B) Cases and tests change: Tests and cases move up or down together. What might look like a significant change in cases is associated with a change in testing, likely not a change in epidemiology. Factors affecting testing include testing capacity, care-seeking behavior, and testing sampling policy. With this change category, the change in testing could be a change in capacity or care-seeking, but the lack of change in positivity indicates testing is still sampling the same population the same way, without changes in epidemiology.

C) Tests and positivity change: Positivity change is driven by testing change, not a change in cases. An increase or decrease in testing does not impact absolute numbers of detected cases, which suggests a change in test sampling. Dynamics that would produce this pattern include, for example, adding population with lower prevalence in the case of open testing, or

484 limiting testing to a higher-prevalence population in the case of symp-
 485 tomatic testing. It is also possible, however, that a change in testing
 486 sampling masks a simultaneous change in epidemiology. In this situa-
 487 tion, the change in testing would have to precisely offset the change in
 488 epidemiology to observe this category of change association. Category C
 489 is thus designated to likely indicate a non-epidemiological change.

490 D) Cases and positivity change: Positivity change is driven by a change in
 491 cases without a change in testing. This suggests a change in epidemiology,
 492 but the significance may be different under random vs. symptomatic test-
 493 ing. Under random testing, this type of change arises only with a change
 494 in SARS-CoV-2 epidemiology. Under symptomatic testing, the restric-
 495 tion of sampling to CLI means that a change in the epidemiology may
 496 be confounded by a change in CLI epidemiology. Note also that symp-
 497 tomatic testing captures changes only in symptomatic SARS-CoV-2 (i.e.
 498 cases of COVID-19). Another possible explanation for this combination
 499 of changes is a change in sampling without a change in the absolute num-
 500 ber of tests. This might occur, for example in a switch from symptomatic
 501 to open testing. For this reason, we categorize this change combination
 502 as likely instead of certainly epidemiological.

503 E) All three variables change: With a change in cases, tests, and positivity, it
 504 remains difficult to disentangle epidemiological from non-epidemiological
 505 factors. Category E can be considered a combination of categories C and
 506 D, and the testing and case changes may or may not be independent. To
 507 capture all changes that may be epidemiological, we consider categories

508 D and E to be epidemiological changes, and categories A, B, and C to
509 be non-epidemiological changes.

510 A principal component analysis (PCA) supporting the separability of change
511 categories is detailed in Appendix B.2.

512 *Appendix B.2. PCA analysis of change categories*

Figure B.11: PCA of countries by change category detection rate (i.e. number of changes in each category divided by total changes detected for each country). A) Explained variance ratio by PCA component number; B) PCA factor loadings by change categories.

513 In addition to the dynamical interpretation of constituent time series (Ap-
514 pendix B), we show the separability of change categories with a principal
515 component analysis (PCA). The surveillance results of different countries are
516 quantitatively characterized by a PCA of the relative frequency with which

517 they detect different categories of changes. The PCA establishes how cate-
518 gories do or don't represent axes of difference across countries.

519 Based on the curve of explained variance ratio by PCA components (Figure
520 B.11A), we choose the first three PCA components to examine factor loadings
521 (Figure B.11B). Each component is dominated by a single category, in PCA
522 component order: category D (epidemiological change); category B (testing
523 artifacts); and category E (confounded). Each of these PCA components
524 is anti-correlated with category A (noise). These relationships among the
525 different change categories is consistent with our dynamical interpretation.
526 Figure B.12 shows the frequencies of change categories for those categories
527 that dominate the factor loadings for all countries in our dataset.

Figure B.12: Change category frequencies by country for categories D, B, and E, chosen according to the dominant categories in the PCA factor loadings.

528 Appendix C. Surveillance considerations

529 Below we lay out in basic terms considerations for three components of SARS-
530 CoV-2 epidemiological surveillance: population, testing, and their role in
531 surveillance metrics. We demonstrate that the testing strategy of random
532 testing with the surveillance metric of positivity is the combination that best
533 represents SARS-CoV-2 prevalence. Here we use the terminology of SARS-

534 CoV-2 to include all asymptomatic and symptomatic infections.

535 Population is composed of people with and without SARS-CoV-2. Of those
536 with SARS-CoV-2, some are asymptomatic, some are symptomatic. Of those
537 without SARS-CoV-2, some are non-symptomatic, others have symptoms of
538 non-COVID-like-illness, and some have COVID-like-illness (CLI) symptoms.

539 Relevant components of testing include eligibility for testing under a given
540 testing framework, as well as testing rate and capacity. Under random sam-
541 pling, the general population is eligible for testing; symptomatic testing re-
542 stricts eligibility to people with CLI symptoms. Testing rate is a measure
543 of tests conducted per total population, while testing capacity indicates the
544 proportion of eligible individuals who are actually tested.

545 Detected cases as a surveillance metric is a function of number of tests, the
546 eligible testing pool, and the total cases within the testing pool. Positivity
547 is defined as detected cases per tests conducted.

548 Applying these formulations to surveillance metrics, we can see that detected
549 cases under symptomatic testing is not only a function of number of tests
550 conducted, but also of the number of individuals exhibiting CLI symptoms.
551 CLI in turn is a function of non-SARS-CoV-2 CLI and symptomatic SARS-
552 CoV-2.

553 Positivity under symptomatic testing is normalized for number of tests con-
554 ducted, measuring not general prevalence in the population, but the portion

555 of CLI that is symptomatic COVID-19. Metrics derived from symptomatic
556 testing do not account for asymptomatic SARS-CoV-2 and are confounded
557 by non-SARS-CoV-2 CLI.

558 As with symptomatic testing, detected cases under random testing are a
559 function of number of tests. The sampling, however, is taken from the gen-
560 eral population, and thus positivity under random testing is a metric that
561 represents prevalence.

562 As tests approach eligible under symptomatic testing, cases detected equals
563 CLI COVID-19 cases. Note, however, that testing coverage (i.e. $tests/eligible$)
564 is not only influenced by the number of tests processed, but also reporting
565 rate. Who shows up for testing is a subset of the people who would be eligible
566 for testing.

567 Assuming capacity to test all eligible individuals and perfect reporting rates,
568 symptomatic testing would still yield only the number of symptomatic COVID-
569 19 cases. For random testing, testing rate is equivalent to testing coverage,
570 and case counts depends on testing. The random testing positivity metric
571 does not depend on testing rate, and captures both symptomatic and asymp-
572 tomatic COVID. The relationship shown empirically in Section 3.3 wherein
573 increasingly open testing policies are associated with increasingly effective
574 epidemiological change monitoring supports the equation-based result that
575 random testing is more suited to epidemiological surveillance.

Population components:

$$\begin{aligned}
 Population &= SARS-CoV-2 + Non-SARS-CoV-2 \\
 SARS-CoV-2 &= SARS-CoV-2_{asympt} + SARS-CoV-2_{sympt} \\
 Non-SARS-CoV-2 &= Non-SARS-CoV-2_{non-sympt} \\
 &\quad + Non-SARS-CoV-2_{sympt_non-CLI} \\
 &\quad + Non-SARS-CoV-2_{sympt_CLI}
 \end{aligned}$$

Testing components:

$$\begin{aligned}
 Eligible_{rand} &= Population = SARS-CoV-2 + Non-SARS-CoV-2 \\
 Eligible_{sympt} &= CLI = Non-SARS-CoV-2_{sympt_CLI} + SARS-CoV-2_{sympt} \\
 testing_rate &= Tests/Population \\
 testing_coverage &= Tests/Eligible
 \end{aligned}$$

Surveillance metrics:

$$\begin{aligned} Cases_{total} &= SARS-CoV-2 \\ Cases_{detected} &= Tests * \frac{Cases}{Eligible} \\ Positivity &= \frac{Cases_{detected}}{Tests} = \frac{Cases}{Eligible} \end{aligned}$$

Symptomatic testing:

$$\begin{aligned} Cases_{detected_sympt} &= Tests_{sympt} * \frac{Cases_{sympt}}{CLI} \\ Positivity_{sympt} &= \frac{Cases_{sympt}}{CLI} \end{aligned}$$

Random testing:

$$\begin{aligned} Cases_{detected_rand} &= Tests_{rand} * \frac{Cases_{total}}{Population} \\ Positivity_{rand} &= \frac{Cases_{total}}{Population} = Prevalence \end{aligned}$$

576 Appendix D. Summary statistics

577 For the purposes of understanding the sensitivity of a given level of testing,
578 we define the standard error for positivity, as number of positive tests per
579 total number of tests, we calculate standard error as follows:

$$SE = \sqrt{\frac{p(1-p)}{N} * \frac{N-n}{N-1}}$$

Where n equals total number of tests, N equals population, and p equals the number of positive tests per the total number of tests. The corresponding margin of error equals one-half the confidence interval, and when calculated at the 95% confidence level is as follows:

$$ME(95\%) = 1.96 * SE$$

Note that this formulation of confidence interval is not reliable when number of tests is very small, or probabilities are very close to zero or one. Under the condition of true random testing, positivity is a direct measure of prevalence. At any given prevalence, margin of error can be calculated for the number of tests administered and the total population. This calculation is carried out for all LMIC countries in our dataset. Margin of error is then normalized by the given prevalence rate. Based on these relationships, $ME(95\%)/Prevalence$ is higher at lower prevalence. In other words, precise measurement becomes increasingly more difficult as prevalence decreases.

Appendix E. NPI correlation

Figure E.13 illustrates aligned NPI co-occurrence. Correlation score is an indicator of how often a given type of aligned NPI is implemented simultaneously with another aligned NPI. Although alignment of an NPI with an epidemiological change can be established, in the case of co-occurrence of two or more aligned NPIs, it is challenging to separate possible effects between the two types. Nonetheless, low correlation score accompanied by high frequency

600 may be indicative of NPIs more likely to be the dominant forcing. This is
601 the case with stay at home requirements and workplace closing, bottom plot
602 of Figure E.13. Conversely, high correlation associated with low frequency
603 indicates NPIs that do not often align with epidemiological change indepen-
604 dently from other aligned NPI types. The NPIs of cancelling public events
605 and restrictions on internal movement are examples of this case.

Figure E.13: Top: Correlation matrix of co-occurrence of NPIs aligned with epidemiological changes. Correlations are normalized along the y-axis: counts of co-occurring NPIs are divided by NPI counts on the diagonal. Bottom: Correlation score and frequency of aligned NPIs by type. Correlation score is the sum of the correlation matrix along the y-axis, normalized to one. Normalized frequency is the count of aligned NPI by type normalized to one.