

1 **Tracking changes in reporting of epidemiological data during the COVID-19 pandemic in**  
2 **Southeast Asia: an observational study during the first wave**

3 Arianna Maever L. Amit MAS <sup>1,2,3</sup>, Veincent Christian F. Pepito MSc <sup>2,3</sup>, Bernardo Gutierrez MSc <sup>4,5</sup>,  
4 Thomas Rawson MMath <sup>4</sup>

- 5 1. Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA
- 6 2. School of Medicine and Public Health, Ateneo de Manila University, Pasig, Philippines
- 7 3. College of Medicine, University of the Philippines Manila, Manila, Philippines
- 8 4. Department of Zoology, University of Oxford, Oxford, United Kingdom
- 9 5. School of Biological and Environmental Sciences, Universidad San Francisco de Quito USFQ,  
10 Quito, Ecuador

11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27

Corresponding author:

Arianna Maever L. Amit MAS

College of Medicine, University of the Philippines Manila, Manila, Philippines

[alamit@up.edu.ph](mailto:alamit@up.edu.ph)

**NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.**

## 28 **Summary**

## 29 **Background**

30 When a new pathogen emerges, consistent case reporting is critical for public health surveillance.  
31 Tracking cases geographically and over time is key for understanding the spread of an infectious disease  
32 and how to effectively design interventions to contain and mitigate an epidemic. In this paper we  
33 describe the reporting systems on COVID-19 in Southeast Asia during the first wave in 2020, and  
34 highlight the impact of specific reporting methods.

35

## 36 **Methods**

37 We reviewed key epidemiological variables from various sources including a regionally comprehensive  
38 dataset, national trackers, dashboards, and case bulletins for 11 countries during the first wave of the  
39 epidemic in Southeast Asia. We recorded timelines of shifts in epidemiological reporting systems. We  
40 further described the differences in how epidemiological data are reported across countries and  
41 timepoints, and the accessibility of epidemiological data.

42

## 43 **Findings**

44 Our findings suggest that countries in Southeast Asia generally reported precise and detailed  
45 epidemiological data during the first wave of the COVID-19 pandemic. However, changes in reporting  
46 were frequent and varied across data and countries. Changes in reporting rarely occurred for  
47 demographic data such as age and sex, while reporting shifts for geographic and temporal data were  
48 frequent. We also found that most countries provided COVID-19 individual-level data daily using  
49 HTML and PDF, necessitating scraping and extraction before data could be used in analyses.

50

## 51 **Interpretation**

52 Countries have different reporting systems and different capacities for maintaining consistent reporting  
53 of epidemiological data. As the pandemic progresses, governments may also change their priorities in  
54 data sharing. Our study thus highlights the importance of more nuanced analyses of epidemiological  
55 data of COVID-19 within and across countries because of the frequent shifts in reporting. Further, most  
56 countries provide data on a daily basis but not always in a readily usable format. As governments  
57 continue to respond to the impacts of COVID-19 on health and the economy, data sharing also needs to  
58 be prioritised given its foundational role in policymaking, and the implementation and evaluation of  
59 interventions.

60

## 61 **Funding**

62 The work was supported through an Engineering and Physical Sciences Research Council (EPSRC)  
63 (<https://epsrc.ukri.org/>) Systems Biology studentship award (EP/G03706X/1) to TR. This project was  
64 also supported in part by the Oxford Martin School.

65 The funders had no role in study design, data collection and analysis, decision to publish, or preparation  
66 of the manuscript.

67

68

69

70

71

72

## 73 **Research in context**

74 Ongoing research into the epidemiology of SARS-CoV-2 depends entirely on access to regularly  
75 updated, factor-rich data. The benefits and importance of data-sharing practices have been well-  
76 documented during previous outbreaks, however the scale of the global pandemic currently faced  
77 presents its own unique challenges. The majority of countries are routinely reporting the number of  
78 confirmed cases and deaths attributed to COVID-19, with the country-wide cumulative totals readily  
79 accessible from databases such as the one curated by Johns Hopkins University.<sup>9</sup> However, the breadth  
80 of further information reported by each country is less understood. Access to demographic and  
81 geographic information of cases in particular is critically important in the context of informing response  
82 policy, as these provide greater insights into how subgroups of the population in different areas are  
83 affected by the disease. Understanding how and when these data is provided is critical to ensuring that  
84 modelling efforts and government response are well-informed.

85

## 86 **Evidence before this study**

87 The current research into the quality of data reporting is severely limited, with studies focussing  
88 primarily around specific locations, time periods, or population sub-groups. One research group has  
89 examined the data availability for 507 patients reported in January, finding that the majority of  
90 information was provided by social media and news outlets. Other than this example, there is no other  
91 work that we are aware of that investigates the issues surrounding data availability. Our work is the first  
92 to explore the scale of data reporting across the broader pandemic timeline.

93

## 94 **Added value of this study**

95 Our research provides wider insight into the data pipeline from government to researchers, and how it  
96 has adapted over time. This timeline provides greater context to the specific findings of subsequent

97 data-driven research, highlighting areas and time periods where particular data feeds are likely to be  
98 particularly biased or data-sparse. We are also able to recommend, based upon our findings, prioritising  
99 the use of the early-case histories of specific countries for the calculation of demographic-specific  
100 disease parameters. By highlighting particular regions where specific data are available, such as travel  
101 history, hospitalisation times and symptom-tracking, we are also able to identify ideal further topics of  
102 research in the ongoing attempts to fight the spread of COVID-19.

103

#### 104 **Implications of all the available evidence**

105 We reviewed the data of the Open COVID-19 Data Working Group's centralised repository and other  
106 relevant data sources to compare the reporting systems of 11 Southeast Asian countries. We found that  
107 governments frequently changed the type of data and level of detail reported as the pandemic  
108 progressed. Our study thus highlights the importance of more nuanced analyses of epidemiological  
109 COVID-19 data within and across countries because of the frequent shifts in reporting. Further, most  
110 countries provide data on a daily basis but not always in a readily usable format. As governments  
111 continue to respond to the impacts of COVID-19 on health and the economy, data sharing also needs to  
112 be prioritised given its foundational role in policymaking, and the implementation and evaluation of  
113 interventions.

114

115

116

117

118

119

## 120 **Introduction**

121 In December 2019, an outbreak of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was  
122 reported in Wuhan, China and was determined to cause the novel coronavirus disease 2019 (COVID-  
123 19). The World Health Organization (WHO) declared the outbreak to be a Public Health Emergency of  
124 International Concern on 30 January 2020, and subsequently a pandemic on 11 March 2020.

125  
126 To effectively respond to public health emergencies, there is a need for timely and accurate reporting  
127 of statistics and data sharing as highlighted in the recent Ebola and Zika epidemics.<sup>1-4</sup> To this end, the  
128 Principles for Data Sharing in Public Health Emergencies consisting of timeliness, ethics, equitability,  
129 accessibility, transparency, fairness, and quality have been developed and introduced.<sup>3,5,6</sup> As of writing,  
130 only one study on data sharing during disease outbreaks among Southeast Asian countries has been  
131 carried out.<sup>7</sup> The study evaluated data quality and timeliness of outbreak reporting in Cambodia, Lao  
132 PDR, Myanmar, and Vietnam for dengue, food poisoning and diarrhea, severe diarrhea, diphtheria,  
133 measles, H5N1 influenza, H1N1 influenza, rabies, and pertussis. Further, it highlighted the broad  
134 differences observed in the data quality and timeliness between participating countries, concluding that  
135 any international data-curating attempts must be versatile enough to accommodate this.

136  
137 In the ongoing COVID-19 crisis, government organisations, public health agencies, and research groups  
138 are responding to the call for rapid data sharing by providing data and curating detailed real-time  
139 databases that are readily and publicly accessible.<sup>8-11</sup> Data from various groups have informed more  
140 than 100,000 papers on COVID-19.<sup>12</sup> Despite progress in reporting and sharing data, several challenges  
141 remain. First, there are ethical and privacy considerations that need to be balanced carefully against the  
142 potential impact of open data sharing. Second, there is a clear lack of capacity and often appropriate  
143 computational infrastructure that may make data sharing in real time unfeasible and burdensome.<sup>3,4</sup>  
144 Such challenges may result in changes in the quality and detail of data reporting between and within

145 countries over time as their respective health systems become increasingly overwhelmed.<sup>11</sup> These shifts  
146 in reporting provide a challenge for accurately comparing epidemiological situations between countries.  
147 In China, for example, it has been shown that changes in reporting have impacted modelling results of  
148 the transmission parameters of COVID-19.<sup>13</sup> Further, as the pandemic progresses and epidemiological  
149 information becomes increasingly less available, analyses of detailed case counts that cover the entire  
150 duration of the epidemic may not be feasible.<sup>10</sup>

151

152 To our knowledge, this is the first study to describe the ways in which various countries in a geographic  
153 region report COVID-19 data, and how the detail of data reporting changed over time. We reviewed  
154 detailed epidemiological data from Southeast Asian countries and tracked how countries' reporting of  
155 COVID-19 data has shifted. We further evaluated differences in reporting between countries and  
156 described the accessibility of epidemiological data during the first wave in 2020. This study is  
157 descriptive and does not seek to evaluate health systems reporting. However, by providing these types  
158 of information, researchers may be able to conduct better and more nuanced analyses of epidemiological  
159 data of COVID-19. By showing changes in reporting, we hope to provide insights on how data should  
160 be viewed and analysed.

161

## 162 **Methods**

### 163 **Study design**

164 We conducted an observational study to describe and track changes in reporting of epidemiological data  
165 during the COVID-19 pandemic in 11 countries in Southeast Asia, namely Brunei, Cambodia,  
166 Indonesia, Lao PDR, Malaysia, Myanmar, Philippines, Singapore, Thailand, Timor-Leste, and  
167 Vietnam. Such a design allows us to compare the data reporting practices between different countries  
168 through time as the pandemic progresses.<sup>14</sup>

## 169 **Data sources and compilation**

170 We focused on reporting mechanisms of individual level COVID-19 data from the aforementioned 11  
171 countries in Southeast Asia. The region is characterized by archipelagos and comprises more than 8·0%  
172 of the world's population. During the first wave of the pandemic, these 11 countries contributed about  
173 1·3% of the cases to the global count of more than 2·3 million cases on April 20.

174

175 The Open COVID-19 Data Curation Group's centralised repository of individual-level information on  
176 patients with laboratory-confirmed COVID-19 collected data on the following variables deemed  
177 essential in monitoring pandemics: (a) Key dates, which include the date of travel, date of onset of  
178 symptoms, date of confirmation of infection, date of admission to hospital, and date of outcome; (b)  
179 Demographic information including the age, sex, and occupation of cases; (c) Geographic information  
180 on domicile and travel history at the highest resolution available down to the district level; (d) Any  
181 additional information such as symptoms and 'contact tracing data' (i.e., a record of exposure to infected  
182 individuals).<sup>10</sup> Information about the occupation of confirmed cases was also included where available.  
183 The collection of data on these variables mirrors the minimum data to be collected for a line list of  
184 pandemic influenza cases obtained from surveillance systems, as suggested by the WHO.<sup>15</sup> Other  
185 sources, such as the interactive dashboard by Johns Hopkins University,<sup>9</sup> do not provide detailed  
186 individual-level information and hence were not used in this study.

187

188 We reviewed the data of the Open COVID-19 Data Working Group's centralised repository and other  
189 relevant data sources. These were primarily websites of the various governments and ministries of  
190 health in Southeast Asia, including but not limited to their COVID-19 trackers, dashboards and case  
191 bulletins. In addition, we reviewed data from news agencies, pre-prints and peer-reviewed research  
192 articles that contained information on COVID-19 cases in the country. We reviewed all possible  
193 publicly available data sources from the date when the first confirmed case was reported in the country,



194 and up to April 20. We documented how key information was reported and how it changed through  
195 time.

196

### 197 **Shifts in reporting of epidemiological data during the first wave**

198 We documented trends and changes in how key epidemiological variables were reported by 11  
199 Southeast Asian countries throughout the study period from January 23 to April 20. Time periods were  
200 defined by specific milestones in each country's data reporting. The first time period or 'first reporting  
201 of cases' for all countries was the date at which the country reported its first COVID-19 case.  
202 Meanwhile, the 'first change in reporting' was the time when the information format was changed from  
203 the first report based on available data during the study period. Any further changes in the level of  
204 detail, also referred in this paper as granularity for geographic data and precision for both demographic  
205 and temporal data, in the reporting of any of the epidemiological variables were considered as a 'change  
206 in reporting' and were noted as a subsequent time period (Table 1). The 'last observed change in  
207 reporting' was the last documented change up to April 20. We also noted the number of cases in each  
208 timepoint. In this paper, we only present results on the 'first reporting of cases' (T0), 'first observed  
209 change in reporting' (T1), and 'last observed change in reporting' (T2).

210

### 211 **Differences in the granularity and precision of reporting across countries**

212 We explored the differences in reporting of demographic, geographic, and temporal data across  
213 countries at three key timepoints: at the time they first reported cases (T0), at the time when the reporting  
214 first changed (T1), and at the last observed change in reporting (T2). Any change in the level of  
215 granularity or precision in reporting is noted. We present these differences for each epidemiological  
216 variable classified into: (a) demographic data; (b) geographic data; and (c) temporal data. Data for other  
217 epidemiological variables are presented in the Supplementary Appendix.

## 218 **Accessibility of COVID-19 data in the region**

219 We described the accessibility of COVID-19 data in the region by documenting and providing a list of  
220 sources for each country in different formats, including: government trackers and dashboards that report  
221 close to real-time data, downloadable PDF reports on cases, downloadable CSV files for cases, and  
222 Github repository for cases. In addition, we noted the frequency of data updates by government and  
223 public health agencies in the country.

224

## 225 **Role of the funding source**

226 The funders had no role in study design, data compilation, data analysis, data interpretation or writing  
227 of the report. All authors had access to the data, and had final responsibility for the decision to submit  
228 for publication.

229

## 230 **Results**

### 231 **Shifts in reporting of epidemiological data during the first wave**

232 The first Southeast Asian country to report a COVID-19 case was Thailand on January 23. Singapore,  
233 Malaysia, Cambodia, Vietnam and the Philippines subsequently reported cases on or before the WHO  
234 declared COVID-19 a Public Health Emergency of International Concern (PHEIC) on January 30.  
235 Indonesia, Brunei, Timor-Leste, Myanmar and Lao PDR reported their first cases of COVID-19 in  
236 March (Figure 1).

237

238 Malaysia had the shortest time between reporting of the first case and first change in reporting of  
239 epidemiological data. Only a day after their first reported case, more detailed reports on the occurrence  
240 of symptoms, and dates of symptom onset and hospitalisation were provided. Similar improvements in

241 terms of the level of granularity and precision in reporting data were also noted for the following  
242 countries: Philippines eventually reported comorbidities for some patients, Singapore and Vietnam  
243 eventually reported data on occupation, and Timor-Leste eventually reported travel history data. As  
244 case numbers increased, several countries provided less detailed information. By March 15, when 96  
245 cases had been identified, Indonesia ceased reporting individual-level data and switched to aggregate  
246 data (i.e., number of cases per day). Timor-Leste followed by April 15, when it had 8 recorded cases.  
247 The first and the last changes in reporting were the same for Indonesia and Brunei, while Myanmar was  
248 the only country that consistently reported individual-level COVID-19 epidemiologic data since  
249 reporting its first two cases on March 23 until April 20.

250

#### 251 **Differences in the granularity and precision of reporting across countries**

252 There were minimal changes in the reporting of demographic data among countries. The majority of  
253 countries reported age and sex except for Timor-Leste, and only Indonesia shifted from a more precise  
254 reporting of age and sex to less detailed reporting (Figures 2a and 2b). We observed more changes in  
255 the reporting of occupation (Figure 2c); Indonesia only provided occupation data at the time of reporting  
256 of first cases, while Singapore and Vietnam included data on occupation of COVID-19 patients at later  
257 timepoints.

258

259 Location information on domicile and travel history differed across countries and timepoints. While all  
260 11 countries provided domicile information (Figure 3a), only Singapore provided precise-level  
261 addresses. Both Indonesia and Malaysia initially provided city-level information and shifted to less  
262 granular reporting. For Indonesia, province-level data was being reported by March 15 when it reached  
263 96 cases. Meanwhile, Malaysia started reporting province-level data on March 21 when it reached 1183  
264 confirmed cases. On the other hand, the information coming from some countries initially presented  
265 less granularity or lower geographic resolution: Lao PDR initially reported country-level information,

266 Thailand initially reported province-level addresses and Vietnam initially reported city-level addresses;  
267 eventually all three countries reported precise address data. There were less differences observed for  
268 travel location data reporting across countries, but also more shifts observed across time (Figure 3b).  
269 Most (8 of 11) provided city-level information of the travel history; only Myanmar provided country-  
270 level information, while both Indonesia and Timor-Leste provided no information at the time of  
271 reporting their first cases. Only Timor-Leste shifted to a more precise level of reporting over time, while  
272 Brunei, Cambodia, Malaysia, Philippines, Singapore and Thailand reported less granular data. Lao PDR  
273 shifted reporting travel histories from city-level information when it reported its first two cases to no  
274 information being shared when it had six confirmed cases, and then to country-level travel history data  
275 when it had reported 11 cases.

276  
277 For all temporal variables, countries reported either precise dates or no dates at all. At the start of each  
278 country's first case, the majority of countries provided travel history dates except for Brunei, Indonesia,  
279 and Timor-Leste (Figure 4a). Only Brunei shifted to reporting dates for the succeeding timepoints while  
280 Malaysia, Philippines, and Singapore stopped reporting dates as cases increased. Lao PDR repeatedly  
281 shifted between reporting travel dates and excluding this information. The precision of reporting  
282 symptom onset dates also varied across countries and timepoints (Figure 4b). Cambodia, Indonesia and  
283 Timor-Leste never reported such information, while Brunei and Vietnam consistently reported specific  
284 dates when symptoms presented. Malaysia provided day information in the succeeding timepoints while  
285 Myanmar, Philippines, Singapore, and Thailand eventually stopped reporting the date of symptom  
286 onset. Lao PDR repeatedly shifted between reporting of dates to no reporting. Both date of confirmation  
287 and date of outcome showed consistent reporting in all countries except Thailand, which stopped  
288 reporting the date of confirmation when it reported 42 cases (Figure 4c). Most countries reported the  
289 date of admission except for Brunei, Cambodia, Indonesia, Malaysia, and Timor-Leste (Figure 4d).  
290 Only Thailand had a shift in reporting dates of discharge, recovery, or death - reporting this information  
291 only in late February when it had 42 cases (Figure 4e). Malaysia provided this information in the  
292 succeeding timepoints while Philippines, Singapore, and Thailand eventually stopped reporting the date

293 of symptom onset. Lao PDR repeatedly shifted between reporting of dates to providing no such  
294 information.

295

## 296 **Accessibility of COVID-19 data in the region**

297 Most of the governments in Southeast Asia instituted infrastructures and guidelines for the  
298 dissemination of COVID-19 data, and provided free access to the public. Brunei's COVID-19 data are  
299 publicly available but the individual requesting the information needs to provide passport information  
300 and contact number to gain access to the database. During the study period, all countries had  
301 government trackers and dashboards except for Timor-Leste (Table 2). Cambodia, Indonesia, Malaysia,  
302 Philippines, Singapore, Thailand, and Vietnam provide daily updates on their COVID-19 data. For the  
303 remaining five countries, the frequency of updating is unclear/irregular. Brunei, Philippines, and  
304 Singapore use PDFs to disseminate select patient-level data, and only Thailand provides downloadable  
305 individual-level CSV files.

306

## 307 **Discussion**

308 Responding to calls for data sharing and transparency, most governments in Southeast Asia established  
309 publicly available sources of COVID-19 individual-level information. This commitment to data sharing  
310 and reporting allowed the comparison of the different data reporting practices of the countries in the  
311 region. We found that countries in Southeast Asia have different reporting practices since the start of  
312 the pandemic and during the first months of its progression. Overall, reporting of epidemiological data  
313 in Southeast Asia is precise and detailed. Many variables were consistently maintained throughout the  
314 initial outbreak period, but those with changes in reporting started early with case counts as low as four  
315 to as high as 136. There was little to no change in reporting of demographic data while changes in  
316 reporting of geographic and temporal variables were frequent and unpredictable as the pandemic  
317 progressed. Further, we find that changes in the level of precision in reporting does not only depend on  
318 case numbers, but also on the policies and interventions implemented. Comparisons across countries

319 for different epidemiologic variables showed that national governments may shift to a less or more  
320 precise reporting of data as dictated by the burden of COVID-19 in the communities and/or their  
321 national response. As an example, Indonesia started reporting aggregate data less than two weeks after  
322 their first case was reported. Their government did not implement a nationwide lockdown, but rather  
323 focused on scaling up capacity, treatment of patients and supporting economic recovery. Conversely,  
324 Lao PDR, Thailand and Vietnam reported more precise demographic and geographic data at the end of  
325 the study period compared to how they reported their first cases. The national governments of these  
326 countries established mechanisms to quickly identify and isolate cases and their contacts requiring  
327 detailed contact tracing data. Our findings also show that most countries reported more precise  
328 information towards the end of the study period, but some variables such as travel history location were  
329 reported with less detail compared to the increased precision for domicile data. These trends in travel  
330 history data highlight the shift in priorities of the governments in the region towards managing local  
331 transmission. Southeast Asian countries implemented travel restrictions early, therefore having fewer  
332 imported cases and less need for precise travel history data.<sup>33</sup>

333

334 Data on dates of symptom onset, confirmation, admission, and death or discharge are important in  
335 estimating disease burden and forecasting health service needs. Dates of confirmation and death or  
336 discharge were reported consistently by most countries. This reflects the effective system of  
337 governments to register all confirmed patients in their database upon entry and exit in the healthcare  
338 system. However, we found that dates of symptom onset and hospital admission were no longer reported  
339 at the end of the observation period. The reporting of less precise dates could be attributed to the  
340 increasing incidence of COVID-19, which could have overwhelmed data reporting mechanisms of the  
341 countries, particularly because individual patient follow-up requires symptom onset dates to be  
342 accurately logged. Governments thus need to establish systems that allow accurate and fast reporting of  
343 detailed temporal data. Lack of precision could adversely affect the quality of mathematical models and  
344 other analyses, which are used to forecast demand for health services and make decisions. This  
345 consequently impacts the responses to COVID-19 at a national and subnational level, which is of greater  
346 concern among low- and middle-countries (LMICs) that have already fragile health systems. Our

347 findings provide insights on how different health systems respond to the pandemic. Consequently, these  
348 could be used to guide how publicly available data is analysed, used, and interpreted.

349

350 Most countries reported COVID-19 data daily, with unclear reporting frequencies only being observed  
351 for Brunei, Lao PDR, and Timor-Leste. These countries do not report new cases every day because of  
352 the low number of new daily cases leading to days where no additional cases are confirmed. As they  
353 only provide updates on days when new COVID-19 cases are confirmed, their frequency of providing  
354 data updates on COVID-19 is thus irregular. Countries primarily reported individual-level data in either  
355 HTML and PDF formats, which necessitates scraping and extraction before such data could be used in  
356 analyses. During the study period, only Thailand provided a downloadable CSV format of their data.  
357 Ready-to-use data formats are important as these allow the public and scientific community to rapidly  
358 view and analyse country-specific information.

359

360 An important limitation of this study is the absence of an assessment on data quality. This evaluation  
361 was not carried out because of the fast progression of the pandemic with corresponding rapid changes  
362 in data reporting. The lack of an up-to-date and complete line list also prevents a thorough assessment  
363 of data quality. Lastly and most importantly, an evaluation of data quality also requires the consideration  
364 of other indicators such as flexibility, representativeness, data security and system stability to provide  
365 a more accurate picture of health systems and disease surveillance systems.<sup>7</sup> These information are not  
366 readily available and require more resources to be collected. Despite such caveats, however, this study  
367 is the first to systematically describe and compare reporting of important epidemiological data for  
368 COVID-19 across countries during the first wave. Our findings will allow researchers to conduct more  
369 nuanced analyses using epidemiological data of COVID-19.

370

371 In conclusion, reporting systems in the region have been quickly established and countries provided  
372 detailed individual-level data during the first wave. This pandemic highlights the critical role of timely,  
373 accurate, and precise data sharing during outbreaks of global scale. Some concerns regarding data  
374 sharing remain, such as data privacy and public criticisms.<sup>3,4</sup> Given that sharing of data is needed for

375 evidence-informed policies and interventions, maintaining and strengthening data reporting systems  
376 should still be a priority of countries.<sup>34-36</sup> For the purposes of surveillance on emerging infectious  
377 diseases, we recommend that governments coordinate data collection and reporting so that data are as  
378 comparable as possible between countries. Countries may also benefit from reporting data in a fully  
379 open access format that is readily available and in machine-readable formats to accommodate new  
380 epidemics and context-specific information. Hopefully, more governments will come to share precise  
381 data to allow more nuanced analyses. This will provide an opportunity to better understand the disease  
382 and how best to respond to the pandemic.

383

### 384 **Author contributions**

385 All authors contributed to the study concept and design, data compilation, analysis, and interpretation.  
386 AMLA and VCFP wrote the initial draft of the manuscript with inputs from BG and TR. All authors  
387 critically revised the report and approved the final version for submission.

388

### 389 **Declaration of interests**

390 The authors declare no competing interests.

391

### 392 **Data sharing**

393 All data and code are available here: <https://github.com/beoutbreakprepared/nCoV2019>

394

### 395 **Acknowledgements**

396 We thank Moritz U.G. Kraemer and the Open COVID-19 Data Working Group for their support and  
397 insights. The full list of curators and contributors making up the Open COVID-19 Data Working Group  
398 is provided at: <https://github.com/beoutbreakprepared/nCoV2019>. This work was supported through an  
399 Engineering and Physical Sciences Research Council (EPSRC) (<https://epsrc.ukri.org/>) Systems  
400 Biology studentship award (EP/G03706X/1) to TR. This project was also supported in part by the  
401 Oxford Martin School.



## 402 **Tables and figures**

403 Table 1. Variables included in the study and the definitions of ‘change in reporting’

404 Table 2. Availability of government dashboards and trackers for COVID-19, frequency of updates, and raw data  
405 sharing formats used in countries in Southeast Asia as of April 20

406 Figure 1. Timeline of key events with corresponding number of cases during the first wave of the COVID-19  
407 pandemic.

408 Figure 2. Differences in the level of precision in reporting demographic data: (a) age, (b) sex, and (c) occupation  
409 over three timepoints.

410 Figure 3. Differences in the level of granularity in reporting geographic data: (a) domicile, and (b) travel history  
411 location over three timepoints.

412 Figure 4. Differences in the level of precision in reporting temporal data: (a) date of travel, (b) date of symptom  
413 onset, (c) date of confirmation, (d) date of hospital admission, and (e) date of outcome over three timepoints.

414

415

416

417

418

419

420

421

422

423

424

425

426 **Table 1. Variables included in the study and the definitions of ‘change in reporting’**

Variable	Definition and possible responses	Change in reporting
Date of reporting	Date at which first case of COVID-19 is reported or date at which there is change of reporting in any of the variables	Not applicable
Cases	Number of cases as of the ‘Date of reporting’	Not applicable
Data aggregation	Type of data aggregation used. “Individual” denotes that information is reported for each individual case, “Aggregated” if otherwise.	Country no longer reports individual-level data or vice versa
<i>Demographic</i>		
Age	Reporting of age of COVID-19 cases. “Exact” if age in years is reported for each individual case. “Age bracket” if the country reports age ranges. “None” if not reported.	Country’s level of reporting shifts between “Exact”, “Age bracket”, or “None”
Sex	Reporting of sex of COVID-19 cases. “With” if the country reports the sex of their individual confirmed cases. “Without” if not reported.	Country no longer reports data on sex after reporting it previously, or vice versa.
Occupation	Reporting of occupation of COVID-19 cases. “With” if the country reports the occupation of individual confirmed cases. “Without” if not reported.	Country no longer reports occupation data after reporting it previously or vice versa
<i>Geographic</i>		
Domicile	Reporting of domicile of COVID-19 cases. “Precise location” if street-level addresses or exact location information are available. “City” if city addresses are available. “Province” if province/state-level addresses are available. “Country” if only country-level addresses are available. “None” if not reported.	Country changes the level of precision at which it reports its domicile data (e.g., precise location to province)
Travel history location	Reporting of international travel history of COVID-19 cases. “Precise location” if street-level addresses or exact location information are available. “City” if city addresses are available. “Province” if province/state-level addresses are available. “Country” if only country-level addresses are available. “None” if not reported.	Country changes the level of precision at which it reports its domicile data (e.g., precise location to province)

<i>Temporal</i>		
Date of travel	Reporting of dates of international travel of COVID-19 cases. “Day” if the country reports the exact date of travel. “Month” if the country only reports the month of travel. “None” if it is not reported.	Country’s level of reporting shifts between “Day, “Month”, and “None”
Date of symptom onset	Reporting of date of symptom onset of COVID-19 cases. “Day” if the country reports the exact date of travel. “Month” if the country only reports the month of travel. “None” if it is not reported.	Country’s level of reporting shifts between “Day, “Month”, and “None”
Date of confirmation	Reporting of date of laboratory confirmation of COVID-19 cases. “Day” if the country reports the exact date of confirmation. “Month” if the country only reports the month of confirmation. “None” if it is not reported.	Country’s level of reporting shifts between “Day, “Month”, and “None”
Date of hospital admission	Reporting of date of hospital admission of COVID-19 cases. “Day” if the country reports the exact date of hospital admission. “Month” if the country only reports the month of hospital admission. “None” if it is not reported.	Country’s level of reporting shifts between “Day, “Month”, and “None”
Date of discharge, recovery, or death	Reporting of date of hospital discharge or recovery or death of COVID-19 cases. “Day” if the country reports the exact date of hospital discharge or death. “Month” if the country only reports the month of hospital discharge or death. “None” if it is not reported.	Country’s level of reporting shifts between “Day, “Month”, and “None”
<i>Other epidemiological variables</i>		
Travel history	Reporting of international travel history of COVID-19 cases. “With” if the country reports the occupation of individual confirmed cases. “Without” if not reported.	Country no longer reports travel history data after reporting it previously or vice versa
Symptoms	Reporting of any symptoms of COVID-19 cases. “With” if the country reports the occupation of individual confirmed cases. “Without” if not reported.	Country no longer reports symptom data after reporting it previously or vice versa
Comorbidities	Reporting of any chronic disease or comorbidity of COVID-19 cases. “With” if the country reports the occupation of	Country no longer reports comorbidities after reporting it previously or vice versa

	individual confirmed cases. “Without” if not reported.	
Cluster	Reporting of disease clusters or known contacts of COVID-19 cases. “With” if the country reports the occupation of individual confirmed cases. “Without” if not reported.	Country no longer reports cluster data after reporting it previously or vice versa
Outcome	Reporting of hospital discharge or recovery or death of COVID-19 cases. “With” if the country reports the occupation of individual confirmed cases. “Without” if not reported.	Country no longer reports outcome data after reporting it previously or vice versa

427

428

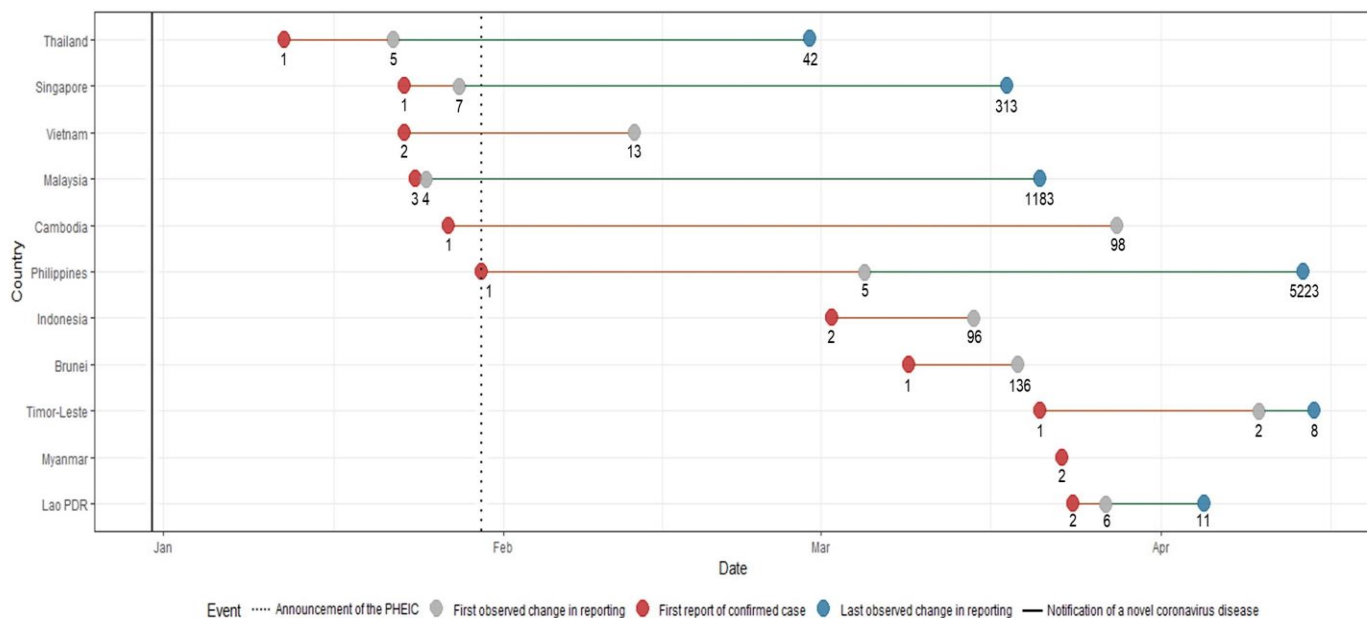
429 **Table 2. Availability of government dashboards and trackers for COVID-19, frequency of updates, and**  
 430 **raw data sharing formats used in countries in Southeast Asia as of April 20**

431

Country	Availability of government tracker/ dashboard	Frequency of update	Raw data sharing formats ever used			
			HTML	CSV	PDF	Other
Brunei	Available <sup>16</sup>	Unclear	Yes	No	Yes	None
Cambodia	Available <sup>17</sup>	Daily	No	No	No	None
Indonesia	Available <sup>18-20</sup>	Daily	No	No	No	None
Lao PDR	Available <sup>21</sup>	Unclear	Yes <sup>22</sup>	No	No	None
Malaysia	Available <sup>23</sup>	Daily	Yes <sup>24</sup>	No	No	None
Philippines	Available <sup>25</sup>	Daily	Yes <sup>25</sup>	No	Yes <sup>26</sup>	Google Sheets <sup>27</sup>
Singapore	Available <sup>28</sup>	Daily	Yes <sup>29</sup>	No	Yes <sup>29</sup>	None
Thailand	Available <sup>30</sup>	Daily	No	Yes <sup>31</sup>	No	None
Timor-Leste	Not available	Unclear	No	No	No	None
Vietnam	Available <sup>32</sup>	Daily	Yes <sup>32</sup>	No	No	None

432

433



434 Event ..... Announcement of the PHEIC ● First observed change in reporting ● First report of confirmed case ● Last observed change in reporting — Notification of a novel coronavirus disease

435 **Figure 1. Timeline of key events with corresponding number of cases during the first wave of the COVID-**  
 436 **19 pandemic.** The notification of a novel coronavirus disease in late December and the announcement of the  
 437 PHEIC are denoted by lines. Shifts in reporting are defined by a change in level of detail and precision in any of  
 438 the epidemiological variables. Each key reporting shift is denoted by a colored circle.

439

440

441

442

443

444

445

446

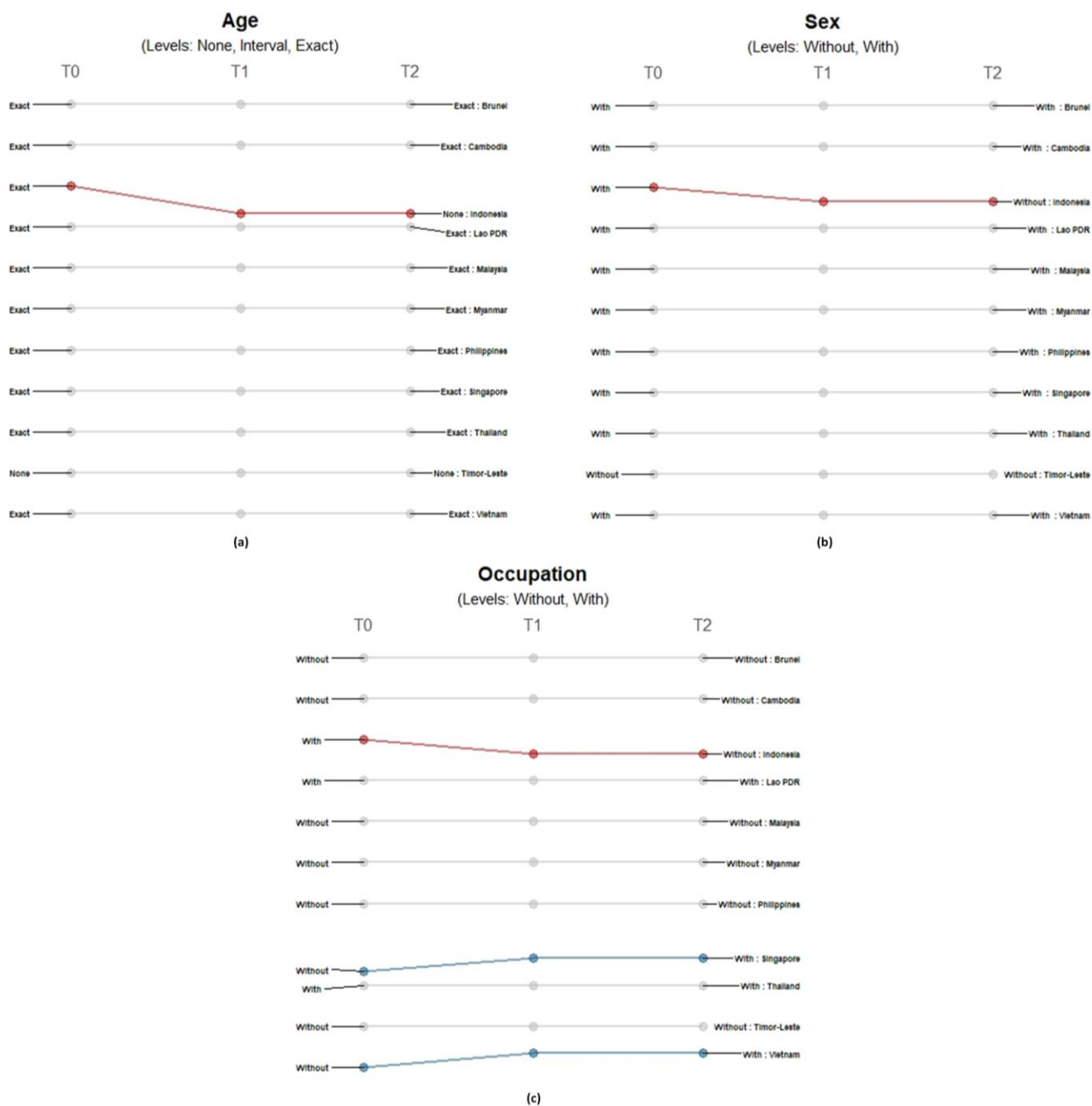
447

448

449

450

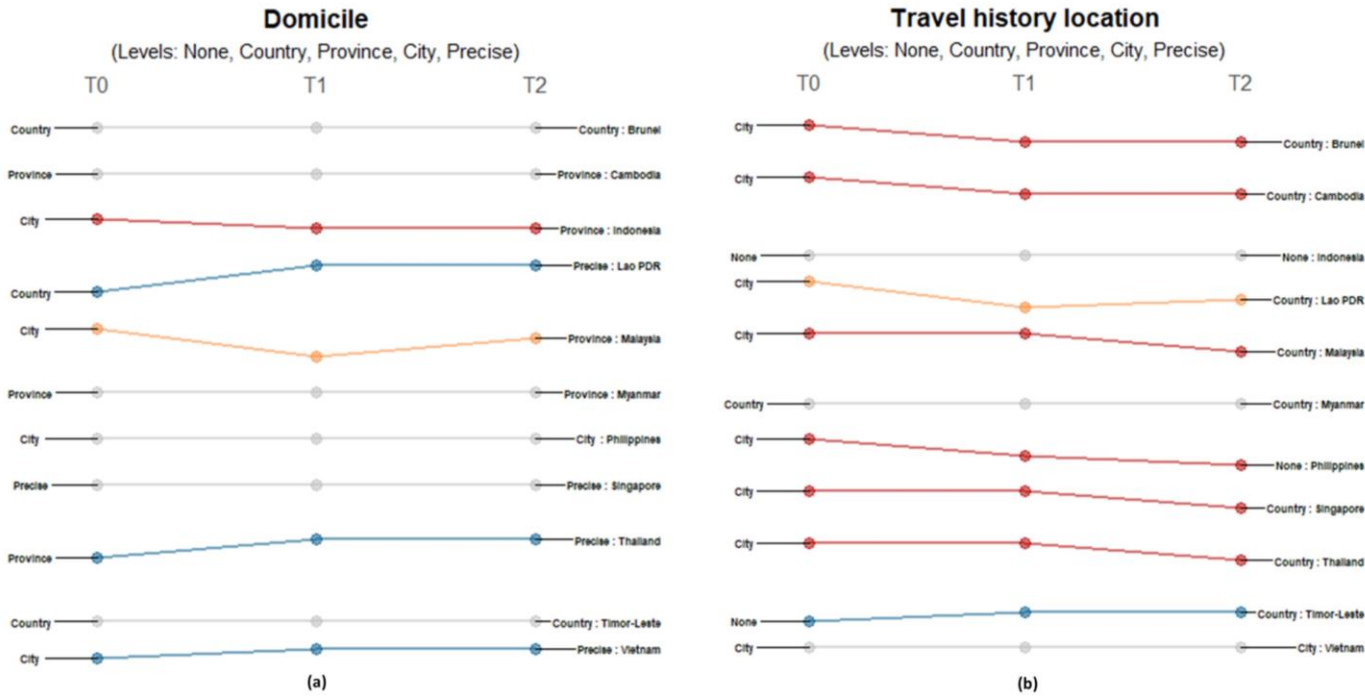
451



452

453 **Figure 2. Differences in the level of precision in reporting demographic data: (a) age, (b) sex, and (c)**  
 454 **occupation over three timepoints.** Only those countries with changes in the level of detail and precision of  
 455 reporting are highlighted. Each country may shift reporting at any timepoint: at the first reporting of cases' (T0),  
 456 'first observed change in reporting' (T1), and 'last observed change in reporting' (T2). Each country may report  
 457 less precise data indicated by a decreasing slope (red) or more precise data indicated by an increasing slope  
 458 (blue) consistently over time. Reporting may not be consistent across timepoints with shifts between different  
 459 levels of precision (yellow) or reporting may not have changed at all during the study period (grey). The levels  
 460 of precision are indicated for each epidemiological variable. Age has three levels while both sex and occupation  
 461 are binary variables.  
 462

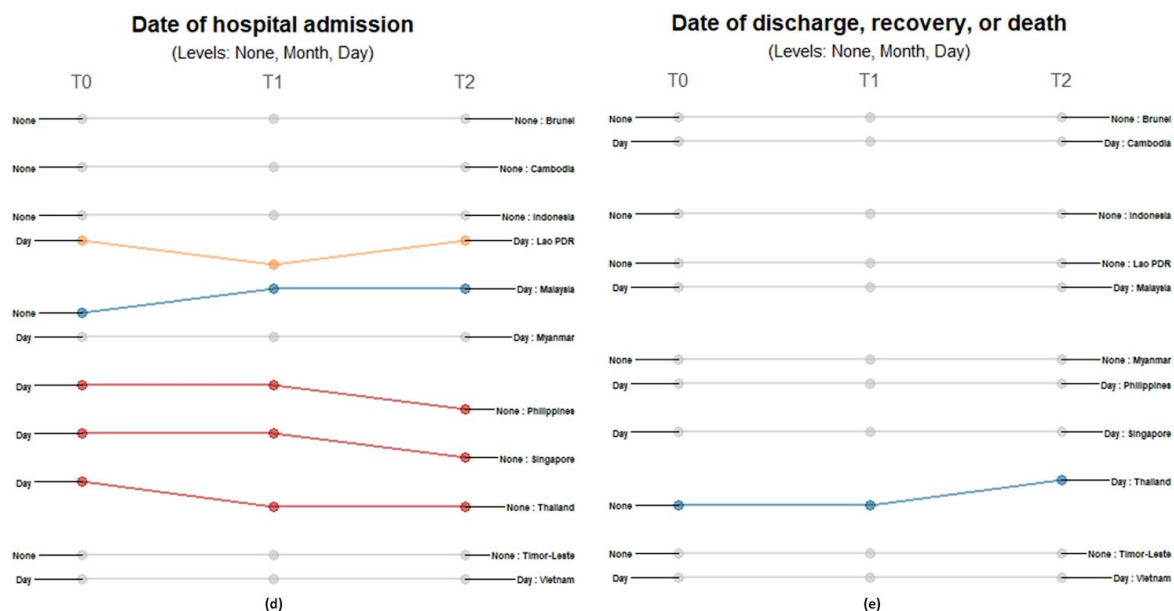
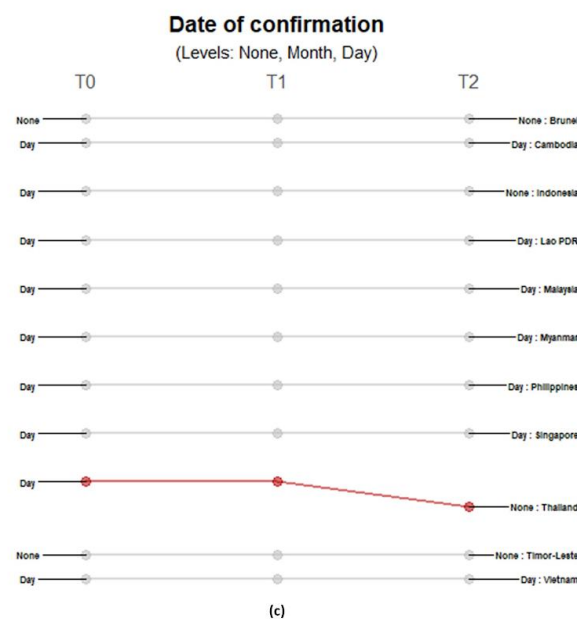
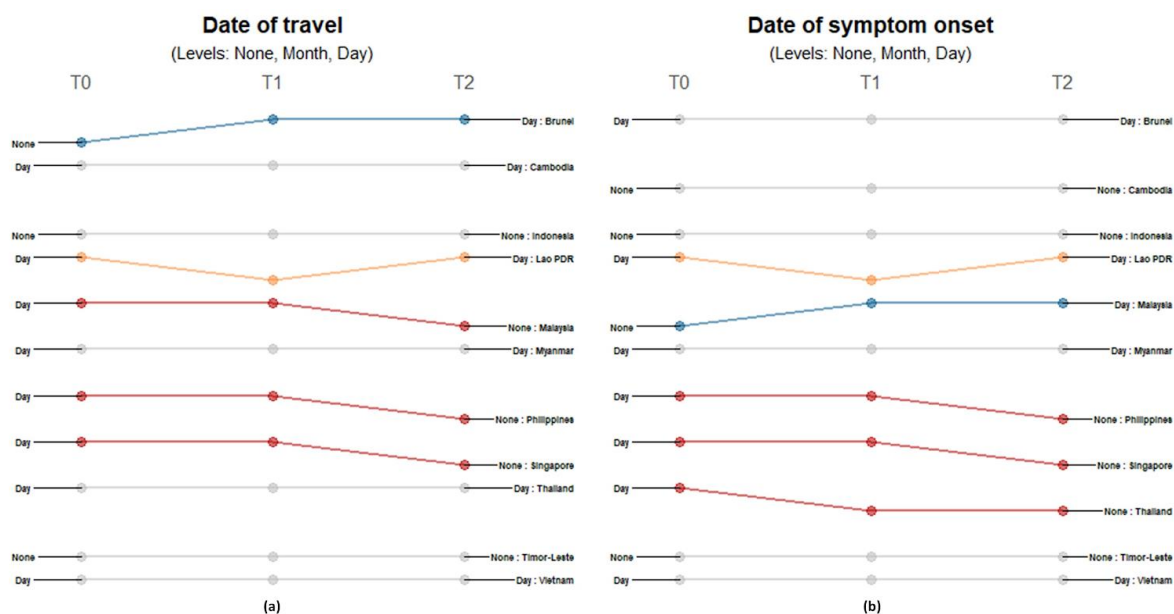
463



464

465 **Figure 3. Differences in the level of granularity in reporting geographic data: (a) domicile, and (b) travel**  
 466 **history location over three timepoints.** Only those countries with changes in the level of granularity or  
 467 geographic resolution of reporting are highlighted. Each country may shift reporting at any timepoint: at the first  
 468 reporting of cases' (T0), 'first observed change in reporting' (T1), and 'last observed change in reporting' (T2).  
 469 Each country may report less granular data indicated by a decreasing slope (red) or more granular data indicated  
 470 by an increasing slope (blue) consistently over time. Reporting may not be consistent across timepoints with shifts  
 471 between different levels of granularity (yellow) or reporting may not have changed at all during the study period  
 472 (grey). The levels of granularity are indicated for each epidemiological variable. All geographic data have five  
 473 levels of granularity/geographic resolution: none, country, province, city, and precise.  
 474

475





477 **Figure 4. Differences in the level of precision in reporting temporal data: (a) date of travel, (b) date of**  
478 **symptom onset, (c) date of confirmation, (d) date of hospital admission, and (e) date of outcome over**  
479 **three timepoints.** Only those countries with changes in the level of detail and precision of reporting are  
480 highlighted. Each country may shift reporting at any timepoint: at the first reporting of cases' (T0), 'first  
481 observed change in reporting' (T1), and 'last observed change in reporting' (T2). Each country may report less  
482 precise data indicated by a decreasing slope (red) or more precise data indicated by an increasing slope (blue)  
483 consistently over time. Reporting may not be consistent across timepoints with shifts between different levels of  
484 precision (yellow) or reporting may not have changed at all during the study period (grey). The levels of  
485 precision are indicated for each epidemiological variable. All date variables have three levels of precision: none,  
486 month, and day.  
487

488  
489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

## 504 **References**

- 505 1. Dye C, Bartolomeos K, Moorthy V, Kieny MP. Data sharing in public health emergencies: a call  
506 to researchers. *Bull World Health Organ*. 2016 Mar 1;94(3):158–158.
- 507 2. Pearce N, Vandenbroucke JP, VanderWeele TJ, Greenland S. Accurate Statistics on COVID-19  
508 Are Essential for Policy Guidance and Decisions. *Am J Public Health*. 2020 Apr 23:e1–3.
- 509 3. Littler K, Boon W-M, Carson G, Depoortere E, Mathewson S, Mietchen D, et al. Progress in  
510 promoting data sharing in public health emergencies. *Bull World Health Organ*. 2017 Apr 1;95(4):243–  
511 243.
- 512 4. Chretien J-P, Rivers CM, Johansson MA. Make Data Sharing Routine to Prepare for Public Health  
513 Emergencies. *PLOS Medicine*. 2016 Aug 16;13(8):e1002109.
- 514 5. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR  
515 Guiding Principles for scientific data management and stewardship. *Scientific Data*. 2016 Mar  
516 15;3(1):1–9.
- 517 6. GloPID-R. Data Sharing Working GroupGroup. Principles for Data Sharing in Public Health  
518 Emergencies [Internet]. 2017 [cited 2020 May 7]. Available from:  
519 [https://wellcome.figshare.com/articles/Principles\\_for\\_Data\\_Sharing\\_in\\_Public\\_Health\\_Emergencies/  
520 4733590](https://wellcome.figshare.com/articles/Principles_for_Data_Sharing_in_Public_Health_Emergencies/4733590).
- 521 7. Lawpoolsri S, Kaewkungwal J, Khamsiriwatchara A, Sovann L, Sreng B, Phommasack B, et al.  
522 Data quality and timeliness of outbreak reporting system among countries in Greater Mekong  
523 subregion: Challenges for international data sharing. *PLOS Neglected Tropical Diseases*. 2018 Apr  
524 25;12(4):e0006425.
- 525 8. Moorthy V, Henao Restrepo AM, Preziosi M-P, Swaminathan S. Data sharing for novel  
526 coronavirus (COVID-19). *Bull World Health Organ*. 2020 Mar 1;98(3):150.

- 527 9. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time.  
528 The Lancet Infectious Diseases. 2020 May 1;20(5):533–4.
- 529 10. Xu B, Kraemer MUG, Xu B, Gutierrez B, Mekaru S, Sewalk K, et al. Open access epidemiological  
530 data from the COVID-19 outbreak. The Lancet Infectious Diseases. 2020 May 1;20(5):534.
- 531 11. Sun K, Chen J, Viboud C. Early epidemiological analysis of the coronavirus disease 2019  
532 outbreak based on crowdsourced data: a population-level observational study. The Lancet Digital  
533 Health. 2020 Apr;2(4):e201–8.
- 534 12. Wang LL, Lo K, Chandrasekhar Y, Reas R, Yang J, Burdick D, et al. CORD-19: The COVID-19  
535 Open Research Dataset. ArXiv200410706 Cs [Internet]. 2020 Jul 10 [cited 2020 Oct 21]; Available  
536 from: <http://arxiv.org/abs/2004.10706>.
- 537 13. Tsang TK, Wu P, Lin Y, Lau EHY, Leung GM, Cowling BJ. Effect of changing case definitions  
538 for COVID-19 on the epidemic curve and transmission parameters in mainland China: a modelling  
539 study. The Lancet Public Health. 2020 May 1;5(5):e289–96.
- 540 14. Rothman K, Greenland S, Lash T. Modern Epidemiology. Third Edition. Philadelphia, PA:  
541 Lippincott Williams and Wilkins; 2008.
- 542 15. World Health Organization. WHO guidance for surveillance during an influenza pandemic  
543 [Internet]. 2017 [cited 2020 May 16]. Available from:  
544 [https://www.who.int/influenza/preparedness/pandemic/WHO\\_Guidance\\_for\\_surveillance\\_during\\_an\\_](https://www.who.int/influenza/preparedness/pandemic/WHO_Guidance_for_surveillance_during_an_influenza_pandemic_082017.pdf)  
545 [influenza\\_pandemic\\_082017.pdf](https://www.who.int/influenza/preparedness/pandemic/WHO_Guidance_for_surveillance_during_an_influenza_pandemic_082017.pdf).
- 546 16. Ministry of Health (Brunei). Welcome to HealthInfo [Internet]. 2020 [cited 2020 Jun 17].  
547 Available from:  
548 <https://www.healthinfo.gov.bn/smartguide/user/login?callback=https%3A%2F%2Fwww.healthinfo.g>  
549 [ov.bn%2F%23%2Fhome](https://www.healthinfo.gov.bn/smartguide/user/login?callback=https%3A%2F%2Fwww.healthinfo.gov.bn%2F%23%2Fhome).

- 550 17. Ministry of Health (Cambodia). ប្រព័ន្ធតាមដានករណីឆ្លងកូវីដ១៩ [Internet]. 2020 [cited 2020  
551 Jun 17]. Available from: [https://covid19-](https://covid19-map.cdcmoh.gov.kh/?fbclid=IwAR29DjPwYCLbkqtreXbXnRgsAVjOqEAA0HnX96NqnLLdeT4XKep_PDwH-3U)  
552 [map.cdcmoh.gov.kh/?fbclid=IwAR29DjPwYCLbkqtreXbXnRgsAVjOqEAA0HnX96NqnLLdeT4X](https://covid19-map.cdcmoh.gov.kh/?fbclid=IwAR29DjPwYCLbkqtreXbXnRgsAVjOqEAA0HnX96NqnLLdeT4XKep_PDwH-3U)  
553 [Kep\\_PDwH-3U](https://covid19-map.cdcmoh.gov.kh/?fbclid=IwAR29DjPwYCLbkqtreXbXnRgsAVjOqEAA0HnX96NqnLLdeT4XKep_PDwH-3U).
- 554 18. Ministry of Health (Indonesia). Home » Info Infeksi Emerging Kementerian Kesehatan RI  
555 [Internet]. Info Infeksi Emerging Kementerian Kesehatan RI. [cited 2020 Jun 17]. Available from:  
556 <https://covid19.kemkes.go.id/>.
- 557 19. Gugus Tugas Percepatan Penanganan COVID-19. Home Task Force for the Acceleration of  
558 Handling COVID-19 [Internet]. covid19.go.id. 2020 [cited 2020 Jun 17]. Available from:  
559 <https://covid19.go.id/>.
- 560 20. Ariansyah A. Badan Nasional Penanggulangan Bencana [Internet]. BNPB. 2020 [cited 2020 Jun  
561 17]. Available from: <https://bnpb.go.id>.
- 562 21. Ministry of Health (Laos). ຄະນະສະເພາະກົດ COVID-19 [Internet]. [cited 2020 Jun 17]. Available  
563 from: <https://www.covid19.gov.la/index.php>.
- 564 22. Ministry of Health (Laos). ກໍລະນີຕົວຈິງ [Internet]. 2020 [cited 2020 Jun 17]. Available from:  
565 <https://www.covid19.gov.la/index.php?r=site%2Fdetail&id=329>.
- 566 23. Evozi. COVID-19 | Malaysia Outbreak Monitor | Live Updates [Internet]. COVID-19 | Malaysia  
567 Outbreak Monitor | Live Updates. 2020 [cited 2020 Jun 17]. Available from: <https://www.outbreak.my>.
- 568 24. Evozi. Malaysia - Statistics | COVID-19 | Malaysia Outbreak Monitor | Live Updates [Internet].  
569 Malaysia - Statistics | COVID-19 | Malaysia Outbreak Monitor | Live Updates. 2020 [cited 2020 Jun  
570 17]. Available from: <https://www.outbreak.my/stats>.
- 571 25. Department of Health (Philippines). COVID-19 Tracker Philippines [Internet]. 2020 [cited 2020  
572 Jun 17]. Available from: <https://ncovtracker.doh.gov.ph/>.

- 573 26. Department of Health (Philippines). Updates of Novel Coronavirus Disease (COVID-19)  
574 [Internet]. [cited 2020 Mar 19]. Available from: <https://www.doh.gov.ph/2019-nCov>.
- 575 27. Department of Health (Philippines). DOH COVID-19 DataDrop - Google Drive [Internet]. 2020  
576 [cited 2020 Jun 10]. Available from:  
577 <https://drive.google.com/drive/folders/10VkiUA8x7TS2jkibhSZK1gmWxFM-EoZP>.
- 578 28. Ministry of Health (Singapore). COVID-19: Cases in Singapore [Internet]. 2020 [cited 2020 Jun  
579 17]. Available from: <http://www.gov.sg/article/covid-19-cases-in-singapore>.
- 580 29. Ministry of Health (Singapore). MOH | News Highlights [Internet]. 2020 [cited 2020 Jun 17].  
581 Available from: [https://www.moh.gov.sg/news-highlights/details/797-more-cases-discharged-151-](https://www.moh.gov.sg/news-highlights/details/797-more-cases-discharged-151-new-cases-of-covid-19-infection-confirmed)  
582 [new-cases-of-covid-19-infection-confirmed](https://www.moh.gov.sg/news-highlights/details/797-more-cases-discharged-151-new-cases-of-covid-19-infection-confirmed).
- 583 30. Digital Government Development Agency. ระดับความเสี่ยงและคำแนะนำในการปฏิบัติตน COVID19  
584 (Update ตามประกาศ กรมฯ วันที่ 7/4/2020) - Open Government Data of Thailand [Internet]. 2020 [cited  
585 2020 Jun 17]. Available from: <https://data.go.th/en/dataset/covid19>.
- 586 31. Digital Government Development Agency (Thailand). รายงาน COVID-19 ประจำวัน - Open  
587 Government Data of Thailand [Internet]. 2020 [cited 2020 Jun 17]. Available from:  
588 <https://data.go.th/en/dataset/covid-19-daily>.
- 589 32. Ministry of Health (Vietnam). TRANG TIN VỀ DỊCH BỆNH VIÊM ĐƯỜNG HÔ HẤP CẤP  
590 COVID-19 - Bộ Y tế - Trang tin về dịch bệnh viêm đường hô hấp cấp COVID-19 [Internet]. 2020 [cited  
591 2020 Jun 17]. Available from: <https://ncov.moh.gov.vn/>.
- 592 33. World Health Organization. Pandemic Influenza Preparedness and Response: A WHO Guidance  
593 Document [Internet]. Pandemic Influenza Preparedness and Response: A WHO Guidance Document.  
594 World Health Organization; 2009 [cited 2020 Aug 6]. Available from:  
595 <https://www.ncbi.nlm.nih.gov/books/NBK143061/>.

596 34. Goldacre B, Harrison S, Mahtani K, Heneghan C. WHO consultation on Data and Results Sharing  
597 During Public Health Emergencies [Internet]. Center for Evidence-Based Medicine, University of  
598 Oxford; 2015 [cited 2020 Jun 18]. Available from: [https://www.who.int/medicines/ebola-](https://www.who.int/medicines/ebola-treatment/background_briefing_on_data_results_sharing_during_phes.pdf?ua=1)  
599 [treatment/background\\_briefing\\_on\\_data\\_results\\_sharing\\_during\\_phes.pdf?ua=1](https://www.who.int/medicines/ebola-treatment/background_briefing_on_data_results_sharing_during_phes.pdf?ua=1).

600 35. Aljunid SM, Srithamrongsawat S, Chen W, Bae SJ, Pwu R-F, Ikeda S, et al. Health-Care Data  
601 Collecting, Sharing, and Using in Thailand, China Mainland, South Korea, Taiwan, Japan, and  
602 Malaysia. *Value in Health*. 2012 Jan 1;15(1, Supplement):S132–8.

603 36. Gewin V. Six tips for data sharing in the age of the coronavirus. *Nature* [Internet]. 2020 May 19  
604 [cited 2020 Jun 18]; Available from: <https://www.nature.com/articles/d41586-020-01516-0>.

605

606

607

608

609

610

611

612

613

614

615