

1 **Clustering of age standardised COVID-19 infection fatality ratios and death trajectories**

2 Thu-Lan Kelly<sup>1,2\*</sup>, Greer Humphrey<sup>1</sup>, Caroline Miller<sup>1,3</sup>, Jacqueline A Bowden<sup>3,1</sup>, Joanne

3 Dono<sup>1</sup>, Paddy A Phillips<sup>4</sup>

4

5 <sup>1</sup>South Australian Health and Medical Research Institute, South Australia, Australia

6 <sup>2</sup>University of South Australia, Clinical and Health Sciences, South Australia, Australia

7 <sup>3</sup>University of Adelaide, School of Public Health, South Australia, Australia

8 <sup>4</sup>SA Commission on Excellence and Innovation in Health, South Australia, Australia

9

10 \*Corresponding author

11 E-mail: [Lan.Kelly@sahmri.com](mailto:Lan.Kelly@sahmri.com) (TK)

12

## 13 **Abstract**

## 14 **Background**

15 An accurate measure of the impact of COVID-19 is the infection fatality ratio, or the  
16 proportion of deaths among those infected, which does not depend on variable testing rates  
17 between nations. The risk of mortality from COVID-19 depends strongly on age and current  
18 estimates of the infection fatality ratio do not account for differences in national age profiles.  
19 Comparisons of cumulative death trajectories allow the effect and timing of public health  
20 interventions to be assessed.

21 Our purpose is to (1) determine whether countries are clustered according to infection fatality  
22 ratios and (2) compare interventions to slow the spread of the disease by clustering death  
23 trajectories.

## 24 **Methods**

25 National age standardised infection fatality ratios were derived from age stratified estimates  
26 from China and population estimates from the World Health Organisation. The IFRs were  
27 clustered into groups using Gaussian mixture models. Trajectory analysis clustered  
28 cumulative death rates in two time windows, 50 and 100 days after the first reported death.

## 29 **Findings**

30 Infection fatality ratios from 201 nations were clustered into three groups: young, medium  
31 and older, with corresponding means (SD) of 0.20% (0.03%), 0.38% (0.11%) and 0.93%  
32 (0.21%).

33 At 50 and 100 days after the first reported death, there were two clusters of cumulative death  
34 trajectories from 113 nations with at least 25 deaths reported at 100 days. The first group had  
35 slowly increasing or stable cumulative death rates, while the second group had accelerating  
36 rates at the end of the time window. Fifty-two nations changed group membership between  
37 the time windows.

## 38 **Conclusion**

39 A cluster of younger nations have a lower estimated infection fatality ratio than older nations.  
40 The effect and timing of public health interventions in preventing the spread of the disease  
41 can be tracked by clustering death rate trajectories into stable or accelerating and comparing  
42 changes over time.

## 43 **Introduction**

44 The disease COVID-19 caused by the coronavirus SARS-CoV-2 was first described in  
45 Wuhan, China in December 2019 [1-2]. The impact the disease will have on the global  
46 population has been estimated through the case fatality rate (CFR), or the proportion of  
47 deaths among confirmed cases. Since the number of cases depends on testing rates which  
48 may include only symptomatic or severe cases, the CFR may be an over-estimate of the  
49 impact of the disease.

50 In contrast, the infection fatality ratio (IFR) is the proportion of deaths among infected  
51 individuals and is a more accurate estimate of disease mortality. However, it is difficult to  
52 determine the true number of infections in a population. Recent antibody prevalence studies  
53 have attempted to establish infection rates in the USA, Spain and elsewhere [3-5].

54 A recent meta-analysis calculated an IFR of 0.68 % (95% CI 0.53-0.82%), but with  
55 significant heterogeneity [6]. The authors concluded that different regions may experience  
56 different IFRs due to age structure and underlying co-morbidities and called for more  
57 research on age stratified IFRs. IFRs for China [7] and Italy [8] have been estimated both by  
58 age group and overall. Although the age stratified IFRs show a large increase in risk for older  
59 age groups in both countries, since Italy has an older population than China, overall IFRs  
60 vary from 0.66% (95% credible interval 0.39-1.33%) in China to 1.29 % (95% crI 0.89-  
61 2.01%) in Italy. An IFR of below 0.5% was ruled out in populations with more than 30%  
62 over 60 years old [8].

63 While age is a significant risk factor for mortality rates, the spread of infection through  
64 populations also depends on public health interventions such as physical distancing measures,  
65 mask wearing, testing, contact tracing, quarantine and border controls [9]. Comparing  
66 trajectories of death rates makes it possible to use real-world data to assess the impact of such

67 public health interventions on disease spread between nations. Clustering groups of countries  
68 with similar cumulative death rate trajectories over different time windows enables  
69 comparisons of the timing and effectiveness of such interventions.

70

## 71 **Methods**

72 Age standardisation of IFRs estimated from China by Verity et al [7] was performed using a  
73 weighting method (see Supporting Material). The overall estimate of the IFR from China  
74 (0.66% [7]) was multiplied by each nation's weight to obtain a point estimate of the age  
75 adjusted IFR.

76 Model based clustering with Gaussian mixture models was used to cluster groups of nations  
77 with IFRs arising from the same normal distribution, using the 'mclust' R package version  
78 5.4.6 [10]. Estimates of the mean IFR, SD and bootstrapped 95% confidence intervals of the  
79 mean were determined for each distribution.

80 To investigate factors related to death rates that are independent of national age profiles, such  
81 as public health interventions, the next stage was to analyse death rate trajectories. If  
82 infection rates are assumed to be equal across groups and age stratified IFRs relative to those  
83 in the 80+ age group are assumed to be the same for every nation, then the IFR weights can  
84 be used to age standardise the death rates per population. The age stratified IFRs relative to  
85 the 80+ age group in China are mostly in agreement with those estimated for Italy (S1 Table  
86 in the Supporting Material). Weighting reduced the effect of age on the trajectories so that  
87 other factors involved in the impact of the disease over time could be investigated. Trajectory  
88 analysis of cumulative death rates per population was used at two time points: 50 and 100  
89 days after the first reported death from COVID-19. To ensure stable trajectories, only

90 countries with at least 25 deaths reported by 100 days after the first death were included in  
91 the analysis. Rolling 14-day averages of weighted cumulative deaths rates were smoothed  
92 using splines to reduce the effect of outlying data points. The R package ‘traj’ version 1.2,  
93 which combines principal components of statistical measures of growth and cluster analysis,  
94 was used to cluster cumulative death rate trajectories into groups, without requiring the  
95 number of clusters to be determined a priori [11].

96 The R package ‘rworldmap’ version 1.3-6 [12] was used to visualise the IFR and trajectory  
97 clusters on a global scale. R software version 4.0.2 (R foundation for Statistical Computing,  
98 Vienna, Austria) was used for all analyses.

99

## 100 **Data**

101 Estimates of national populations in 2020 by five-year age groups from the World Health  
102 Organisation were available from [13]. Daily cumulative death rates compiled by the  
103 European Centre for Disease Control were obtained from the Our World in Data website  
104 [14]. All datasets and R code used to produce the results are available from  
105 <https://github.com/lan-k/COVID19>.

106

## 107 **Results**

108 Age adjusted IFRs were calculated for 201 countries. If the national IFRs were assumed to be  
109 from one normal distribution, the mean IFR would be 0.54% (SD 0.34%); however, a  
110 histogram of the IFRs showed the data were not drawn from a single normal distribution (Fig  
111 1).

112 **Figure 1:** Histogram of global Infection Fatality Ratios

113

114 Clustering of the IFRs produced three groups of nations with young, medium and older age  
115 profiles. Model fit diagnostics can be found in the Supporting Material (S1 and S2 Figs).

116 Mean IFRs (SD) from the three distributions are 0.20% (0.03%), 0.38% (0.11%) and 0.93%  
117 (0.21%) (Table 1). Bootstrapped 95% CIs for the mean and SD of the three normal

118 distributions and the minimum and maximum IFR in each cluster are also presented in Table

119 1. The countries included in each cluster are displayed in Fig 2. After excluding countries in

120 the ‘Young’ cluster, the mean IFR (95% CI) from the remaining countries, assuming the data

121 were from a single normal distribution, was 0.67% (0.62-0.72%), which is very close to the

122 meta-analysis estimate of 0.68% (95% CI 0.53-0.82%) [6].

123 **Table 1: Characteristics of the three IFR clusters.**

124

<b>Cluster (N)</b>	<b>Age profile</b>	<b>Mean IFR (%) (95% CI)</b>	<b>SD of distribution (%) (95% CI)</b>	<b>IFR range (%) (min, max)</b>
1 (55)	Young	0.20 (0.18-0.23)	0.03 (0.02-0.05)	(0.14-0.25)
2 (75)	Medium	0.38 (0.30-0.48)	0.11 (0.05-0.15)	(0.25-0.61)
3 (71)	Older	0.93 (0.78-1.03)	0.21 (0.13-0.28)	(0.62-1.51)

125

126 **Figure 2: National membership of three IFR clusters.** Countries with missing data are

127 shown in white.

128 For the trajectory analysis, cumulative death rates were available for 113 countries with at  
129 least 25 deaths 100 days after the first reported death, as of July 21, 2020. Trajectory analysis  
130 clustered countries based on the growth of weighted cumulative death rates over time and the  
131 groups are independent of the IFR clusters in Fig 2. Two groups were found; ‘Stable’ and  
132 ‘Accelerating’ (Table 2). Cluster group membership was based not on cumulative death rates  
133 at the end of the period, but on the shape of the trajectory. The first group, ‘Stable’, had  
134 cumulative death rates which had plateaued or slowly increased towards the end of the time  
135 window, while the second group, ‘Accelerating’, showed death rates which were rapidly  
136 increasing. Details of the statistical measures which defined the clusters at each time window  
137 and diagnostics can be found in S2 Table and S3 and S4 Figs in the Supporting Material. The  
138 median trajectories and interquartile ranges for each group at 50 and 100 days after the first  
139 reported death are shown in Fig 3. Large IQRs at 100 days compared with 50 days possibly  
140 reflect the timing of interventions between the time windows. Cluster group membership did  
141 not change in a sensitivity analysis using unweighted trajectories, since clusters were defined  
142 by the shape of the trajectories. Fig 4 shows national group membership of the two trajectory  
143 clusters at 50 and 100 days after the first reported death. Each trajectory cluster includes a  
144 mixture of young, medium and older nations, indicating that the spread of disease in a  
145 population over time depends on other factors besides age.

146

147 **Table 2: Median cumulative death rate and interquartile range for clusters at 50 and**  
148 **100 days after the first reported death.**

<b>Cluster</b>	<b>50 days (N)</b>	<b>100 days (N)</b>	<b>Median cumulative death rate at 50 days (IQR) (per 100k)</b>	<b>Median cumulative death rate at 100 days (IQR) (per 100k)</b>
----------------	------------------------	-------------------------	---	--

Stable	68	70	7.99 (3.72-15.8)	30.9 (10.1-110.7)
Accelerating	45	43	18.6 (3.90-49.6)	28.1 (11.8-83.1)

149

150 **Figure 3: Median cluster trajectories.** Median trajectories (solid line) and interquartile  
151 range (dashed line) for the ‘Stable’ (blue) and ‘Accelerating’ (red) clusters at (A) 50 days and  
152 (B) 100 days after the first reported death.

153

154 **Figure 4: National trajectory cluster group membership.** Cluster group membership of the  
155 ‘Stable’ (blue) and ‘Accelerating’ (red) clusters at (A) 50 days and (B) 100 days after the first  
156 reported death. Countries excluded from the analysis are shown in white.

157

158 Between the two time windows, 52 out of 113 nations (46%) changed group membership  
159 (Fig 5). 25 countries moved from ‘Stable’ to ‘Accelerating’ (the ‘Worse’ group), while 27  
160 changed from ‘Accelerating’ to ‘Stable’ (the ‘Improved’ group). Again, the change in group  
161 membership between the time points did not depend on the age profile of countries.

162

163 **Figure 5: Change in trajectory membership over time.** Change in group membership of  
164 trajectory clusters between 50 and 100 days after the first reported death. The ‘Stable’ (blue)  
165 and ‘Accelerating’ (red) groups remained unchanged between time points. Between 50 and  
166 100 days, ‘Improved’ (purple) changed from ‘Accelerating’ to ‘Stable’, while ‘Worse’  
167 (orange) changed from ‘Stable’ to ‘Accelerating’.

168

## 169 Discussion

170 The risk of death from COVID-19 is highly dependent on age. Current estimates of the IFR  
171 are between 0.5-1.3%, are difficult to calculate and assume a single value will describe the  
172 global impact of the disease. There have been few studies reporting IFRs for younger nations  
173 such as in Africa, possibly due to difficulties in testing, measuring the number of  
174 asymptomatic infections and reporting accurate death rates. It has been suggested the  
175 different regions will experience different IFRs due to age structure and co-morbidities [6,  
176 15]. A single estimate of the IFR for all nations may not capture the true global distribution.  
177 Our study has shown that national IFRs are not drawn from a single normal distribution, but  
178 from a mixture of three distributions with different means and standard deviations. This  
179 would explain some of the heterogeneity in the IFRs reported [6]. When countries from the  
180 ‘Young’ cluster were excluded, the mean of the remaining national IFRs, assuming a single  
181 normal distribution, is very close to the meta-analysis estimate. If data from younger  
182 countries become available, they may confirm our findings. While younger nations may have  
183 a lower IFRs from age alone, these countries may have less developed health systems and  
184 poorer health status, so the actual infection fatality ratio in these nations may be higher than  
185 that estimated due to age alone.

186 Mortality risk from COVID-19 does not depend on age alone. Apart from relatively  
187 immutable risk factors in a population, such as gender and co-morbidities, implementation of  
188 public health measures can make an important difference to the increase in deaths. Trajectory  
189 analysis showed the timing and effect (or lack) of implementation of public health  
190 interventions, independent of national age profiles. Each trajectory cluster included a mixture  
191 of nations from the IFR clusters, indicating that age alone does not explain the full impact of  
192 the disease on a population. At 50 days after the first reported death, North America, most of

193 Europe and Asia, parts of Central and South America and Australia were experiencing  
194 accelerating death rates. By 100 days, North America, China, Europe and Australia (the  
195 ‘Improved’ group in Fig 5) had stabilised their death trajectories through public health  
196 interventions such as lockdowns, increasing testing rates, contact tracing and border controls.  
197 In contrast, the number of countries with accelerating death rates in Central and South  
198 America, the Middle East and Africa (the ‘Worse’ group) had increased between 50 and 100  
199 days, while countries such as India, Pakistan and Russia had not slowed their trajectories.  
200 Future analysis with longer time windows may indicate whether some nations are  
201 experiencing subsequent waves of infections due to the easing of restrictions or other factors,  
202 including those who had successfully suppressed the disease previously.

203 Our study has some limitations. The assumption that infection rates are equal across age  
204 groups may be met only in nations with large outbreaks and high death rates or with high  
205 inter-generational mixing [16-17]. Age stratified IFRs relative to the 80+ age group may  
206 differ from those in China or Italy, particularly in countries where health system support is  
207 limited, overwhelmed or inequitable. However, cluster group membership in the trajectory  
208 analysis did not depend on weighting for age adjustment. COVID-19 mortality data may be  
209 under-reported and the calculated IFRs may be under-estimates. The IFR estimates were  
210 produced from data that was available in February, 2020, before large scale seroprevalence  
211 studies had been conducted [3-6]. If more up to date age stratified IFR estimates become  
212 available, the analysis can be updated.

213 Mortality also depends on gender, co-morbidities, ethnicity, obesity and other risk factors  
214 such as smoking [18-19], as well as access to health services. Increased risk of severe  
215 COVID-19 requiring hospitalisation due to underlying health conditions has been calculated  
216 at the national level [15]. As an extension to our analysis, the risk models developed for

217 infection hospitalisation ratios by Clark et al [15] could be adapted to the age standardised  
218 national IFR point estimates before clustering.

219

## 220 **Conclusion**

221 COVID-19 IFRs were clustered into three groups depending on national age profiles. A  
222 cluster of younger nations, predominantly in Africa, had a lower mean IFR than older  
223 nations. However, these countries may have less developed health systems and poorer overall  
224 health status, so the actual IFR in these nations may be higher than that estimated from age  
225 alone. Independently of risk due to age, clustering of death trajectories can track the effect of  
226 public health interventions in preventing the spread of the disease and changes over time.

227

## 228 **References**

- 229 1. Zhu N, Zhang D, Wang W, et al; China Novel Coronavirus Investigating and  
230 Research Team. A novel coronavirus from patients with pneumonia in China,  
231 2019. N Engl J Med. 2020;382(8):727-733. doi:[10.1056/NEJMoa2001017](https://doi.org/10.1056/NEJMoa2001017)
- 232 2. Novel Coronavirus Pneumonia Emergency Response Epidemiology Team.  
233 Vital surveillances: the epidemiological characteristics of an outbreak of 2019  
234 novel coronavirus diseases (COVID-19)—China, 2020. China CDC Weekly.  
235 2020;2(8):113-122. Accessed April 16, 2020.  
236 [http://weekly.chinacdc.cn/en/article/id/e53946e2-c6c4-41e9-9a9b-](http://weekly.chinacdc.cn/en/article/id/e53946e2-c6c4-41e9-9a9b-fea8db1a8f51)  
237 [fea8db1a8f51](http://weekly.chinacdc.cn/en/article/id/e53946e2-c6c4-41e9-9a9b-fea8db1a8f51)

- 238 3. Havers FP, Reed C, Lim T, et al. Seroprevalence of Antibodies to SARS-CoV-  
239 2 in 10 Sites in the United States, March 23-May 12, 2020. JAMA Intern Med.  
240 Published online July 21, 2020. doi:10.1001/jamainternmed.2020.4130
- 241 4. Pollán M, Pérez-Gómez B, Pastor-Barriuso R, Oteo J, Hernán MA, Pérez-  
242 Olmeda M, et al. Prevalence of SARS-CoV-2 in Spain (ENE-COVID): a  
243 nationwide, population-based seroepidemiological study. Lancet 2020;  
244 published online July 6. [http://dx.doi.org/10.1016/S0140-6736\(20\)31483-5](http://dx.doi.org/10.1016/S0140-6736(20)31483-5).
- 245 5. Stringhini S, Wisniak A, Piumatti G, Azman AS, Lauer SA, Baysson H, et al.  
246 Seroprevalence of anti-SARS-CoV-2 IgG antibodies in Geneva, Switzerland  
247 (SEROCoV-POP): a population-based study. Lancet 2020; published online  
248 June 11. [http://dx.doi.org/10.1016/S0140-6736\(20\)31304-0](http://dx.doi.org/10.1016/S0140-6736(20)31304-0)
- 249 6. Meyerowitz-Katz G, Merone L, A systematic review and meta-analysis of  
250 published research data on COVID-19 infection-fatality rates. Preprint  
251 published online July 07, 2020. medRxiv 2020.05.03.20089854; doi:  
252 <https://doi.org/10.1101/2020.05.03.20089854>
- 253 7. Verity R, Okell LC, Dorigatti I, et al. Estimates of the severity of coronavirus  
254 disease 2019: a model-based analysis. Lancet Infect Dis. Published online  
255 March 30, 2020. DOI:[https://doi.org/10.1016/S1473-3099\(20\)30243-7](https://doi.org/10.1016/S1473-3099(20)30243-7)
- 256 8. Rinaldi G, Paradisi M. An empirical estimate of the infection fatality rate of  
257 COVID-19 from the first Italian outbreak. Preprint published online May 18,  
258 2020. <https://www.ssrn.com/abstract=3582811>
- 259 9. Ferguson NM, Laydon D, Nedjati-Gilani G, Imai N, Ainslie K, Baguelin M, et  
260 al. Impact of non-pharmaceutical interventions (NPIs) to reduce COVID-19  
261 mortality and healthcare demand. Imperial College London. Published online  
262 16 March, 2020. doi: <https://doi.org/10.25561/77482>

- 263 10. Scrucca L, Fop M, Murphy TB and Raftery AE. mclust 5: clustering,  
264 classification and density estimation using Gaussian finite mixture models.  
265 The R Journal 2016; 8(1):289-317 <https://doi.org/10.32614/RJ-2016-021>
- 266 11. Sylvestre MP, McCusker J, Cole M, Regeasse. Classification of patterns of  
267 delirium severity scores over time in an elderly population. International  
268 Psychogeriatrics, 2006; 18(4):667-680. doi:10.1017/S1041610206003334.
- 269 12. South, A. rworldmap: A New R package for Mapping Global Data. The R  
270 Journal 2011; 3(1):35-43.
- 271 13. WHO World Population Prospects 2020 estimates. Accessed April 6, 2020.  
272 <https://population.un.org/wpp/Download/Standard/Population/>
- 273 14. Our World in Data, <https://covid.ourworldindata.org/data/owid-covid-data.csv>,  
274 extracted on July 21, 2020.
- 275 15. Clark A, Jit M, Warren-Gash C, Guthrie B, Wang HHX, Mercer SW, et al.  
276 Global, regional, and national estimates of the population at increased risk of  
277 severe COVID-19 due to underlying health conditions in 2020: a modelling  
278 study. Lancet Glob Health 2020; published online June 12.  
279 [http://dx.doi.org/10.1016/S2214-109X\(20\)30264-3](http://dx.doi.org/10.1016/S2214-109X(20)30264-3).
- 280 16. Dowd J, Andriano L, Brazel DM, Rotondi V, Block P, Ding X, et al.  
281 Demographic science aids in understanding the spread and fatality rates of  
282 COVID-19. Proceedings of the National Academy of Sciences. 2020; 117  
283 (18): 9696-9698; DOI: 10.1073/pnas.2004911117
- 284 17. Esteve A, Permanyer I, Boertien, Vaupel JW. National age and coresidence  
285 patterns shape COVID-19 vulnerability. Proceedings of the National Academy  
286 of Sciences. 2020; 117 (28) 16118-16120; DOI: 10.1073/pnas.2008764117

- 287                   18. Yang J, Zheng Y, Gou X, Pu K, Chen Z, Guo Q, et al. Prevalence of  
288                   comorbidities and its effects in coronavirus disease 2019 patients: A  
289                   systematic review and meta-analysis. *Int J Infect Dis.* 2020;94:91–95.  
290                   doi:10.1016/j.ijid.2020.03.017
- 291                   19. Richardson S, Hirsch JS, Narasimhan M, Crawford JM, McGinn T, Davidson  
292                   KW, et al. Presenting Characteristics, Comorbidities, and Outcomes Among  
293                   5700 Patients Hospitalized With COVID-19 in the New York City Area.  
294                   *JAMA.* Published online April 22, 2020. doi:10.1001/jama.2020.6775
- 295

## 296 **Supporting information captions**

297 **S1 Table: Relative death rates compared with the 80+ age group from China and Italy.**

298 Confidence limits for IFR/IFR 80+ were estimated using the upper and lower limits for IFR  
299 and dividing by the point estimate for IFR 80+ (IFR 81+ for Italy).

300

301 **S1 Figure: Probability density function and histogram.** Probability density function of  
302 fitted distributions from three clusters (solid line) and histogram of observed IFRs.

303

304 **S2 Figure: Model diagnostics.** Quantile-Quantile plots to test normality (left) and estimated  
305 and empirical cumulative density functions (CDF) (right).

306

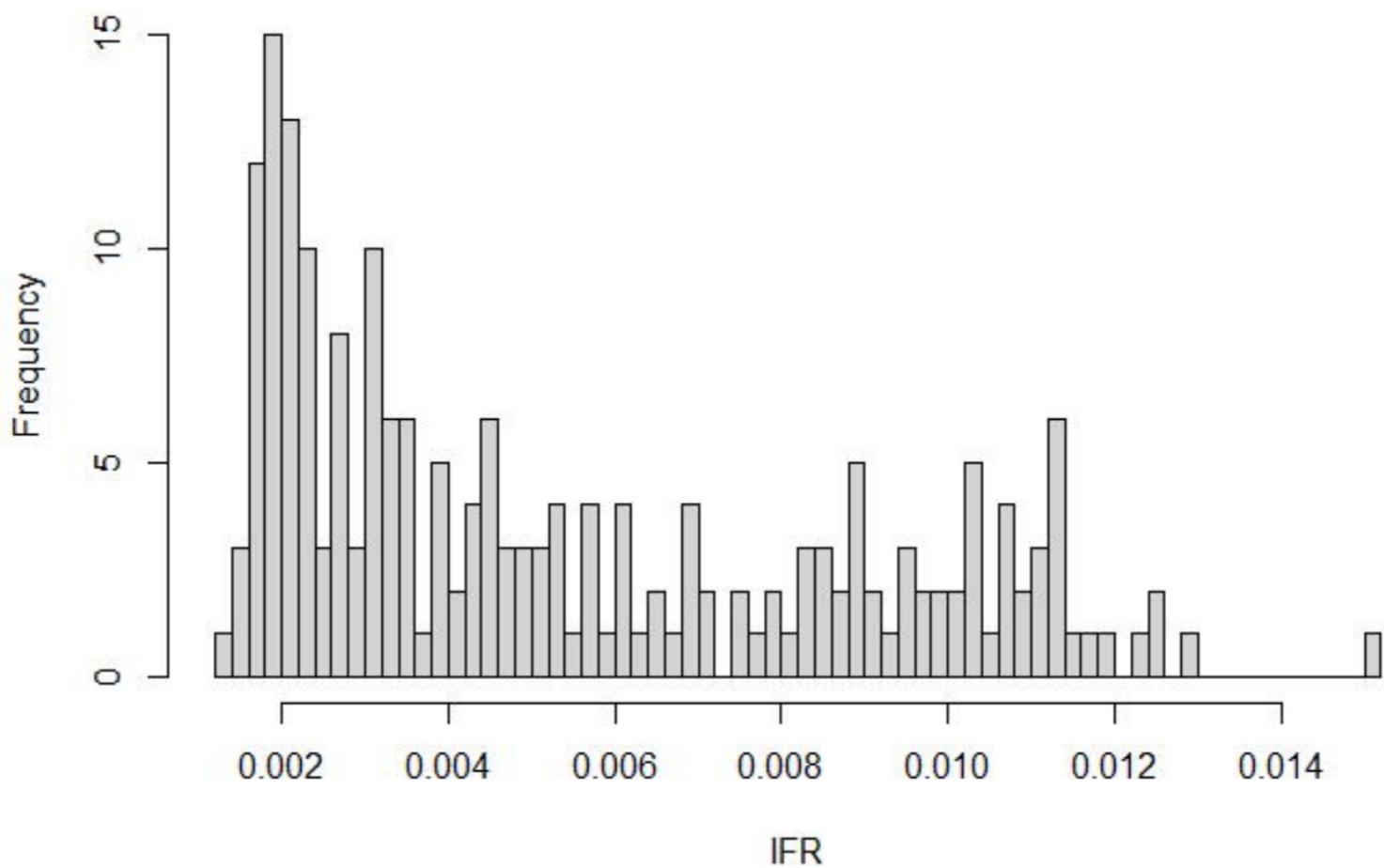
307 **S2 Table: Statistical measures of trajectory clusters.** Statistical measures selected by  
308 factor analysis to describe the 50 and 100 day trajectories by cluster.

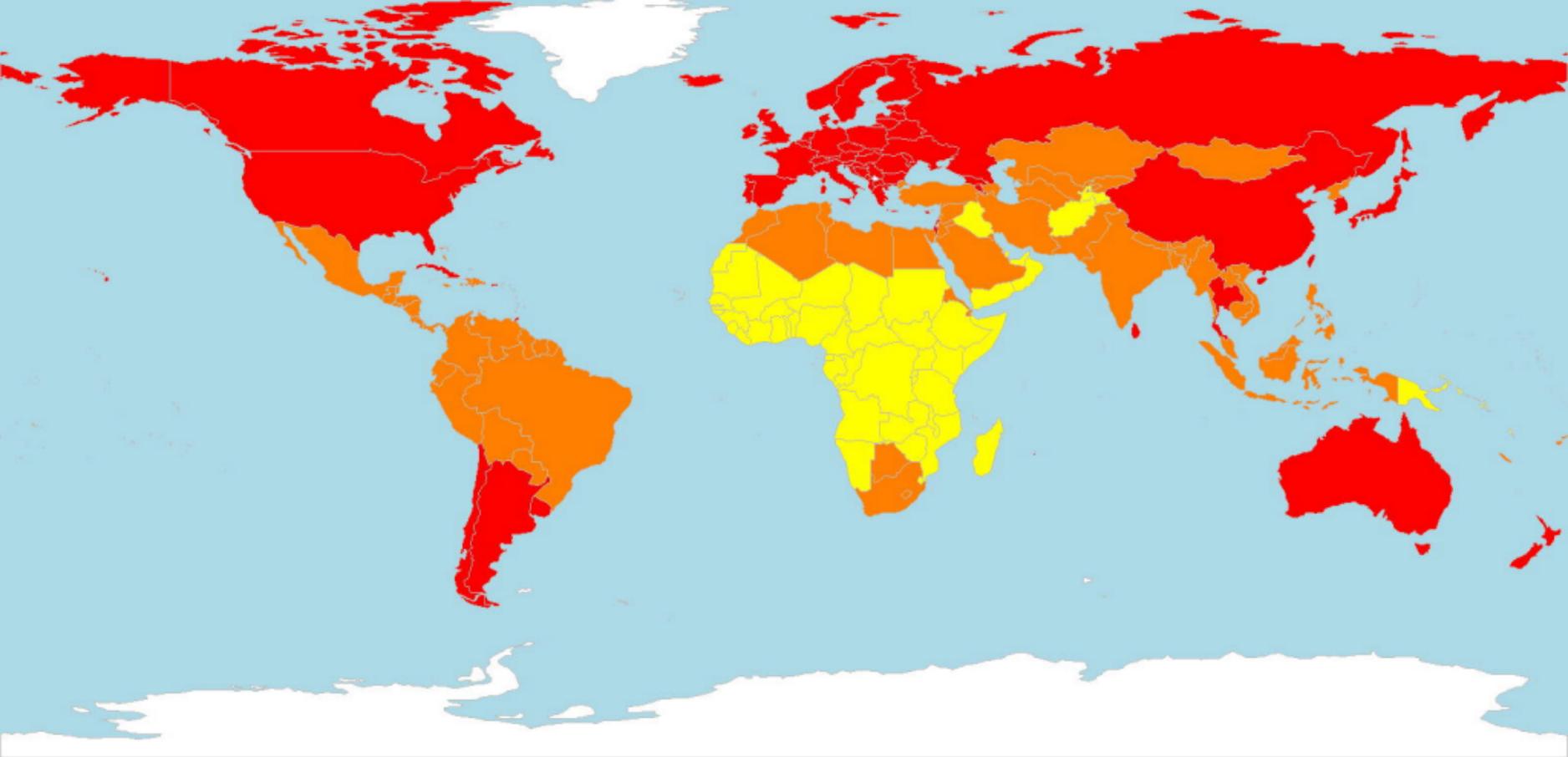
309

310 **S3 Figure: Criteria used to determine optimal number of clusters for 50 day**  
311 **trajectories.** Ccc criteria for 2-15 clusters (left) and within groups sum of squares for 1-15  
312 clusters (right).

313

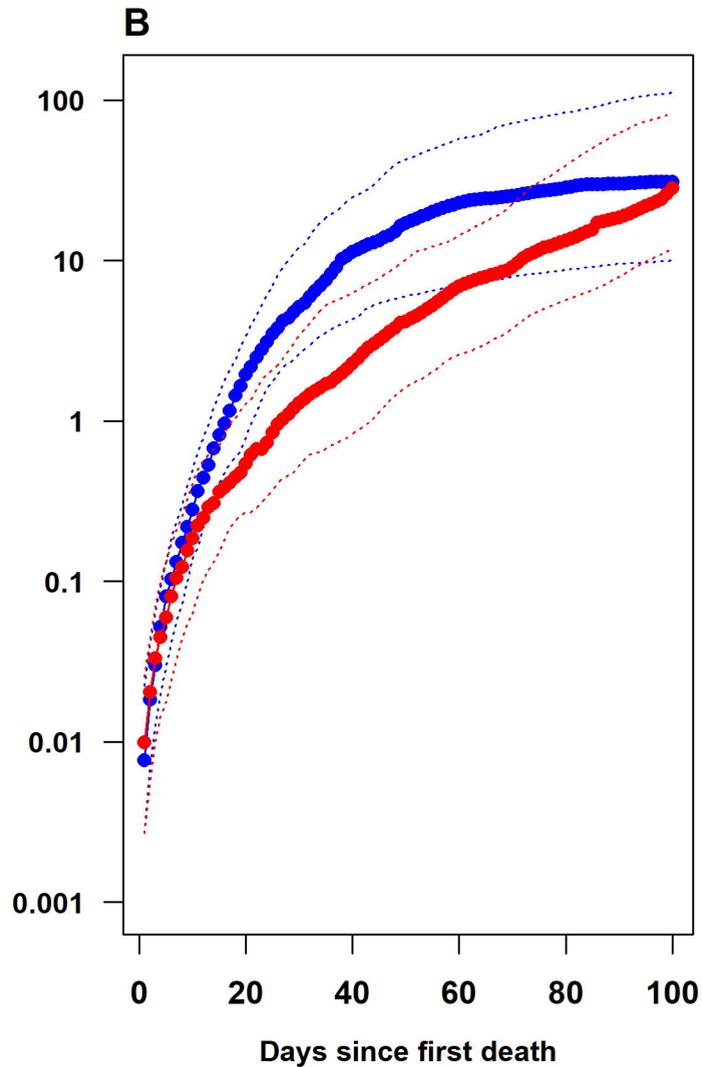
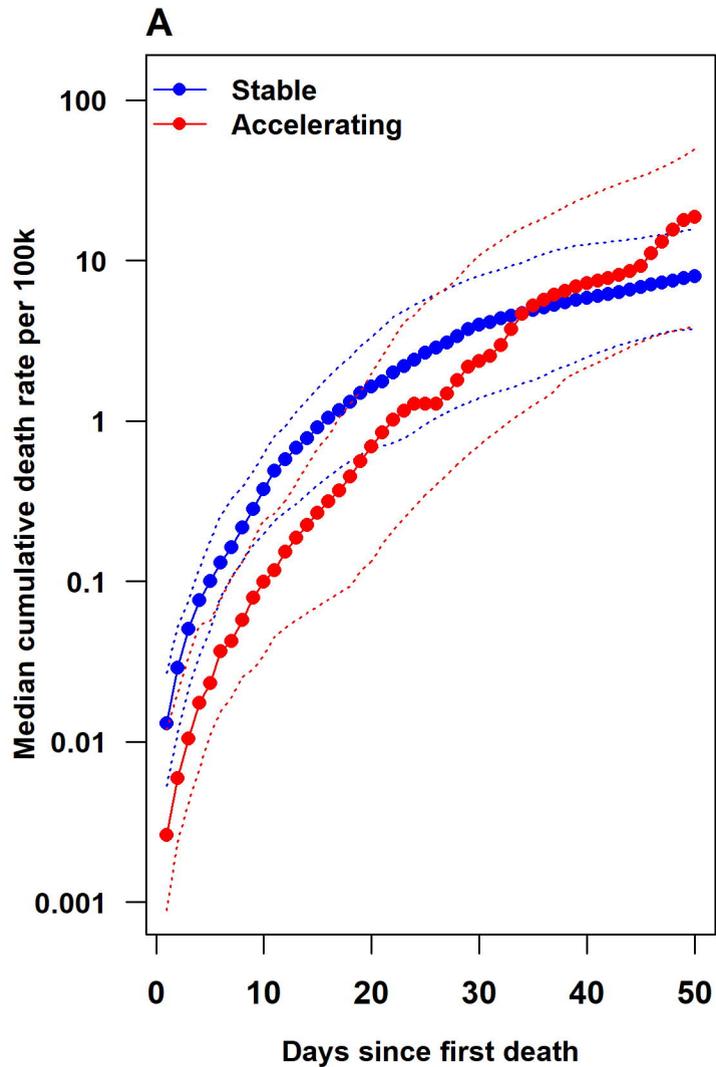
314 **S4 Figure: Criteria used to determine optimal number of clusters for 100 day**  
315 **trajectories.** Ccc criteria for 2-15 clusters (left) and within groups sum of squares for 1-15  
316 clusters (right).

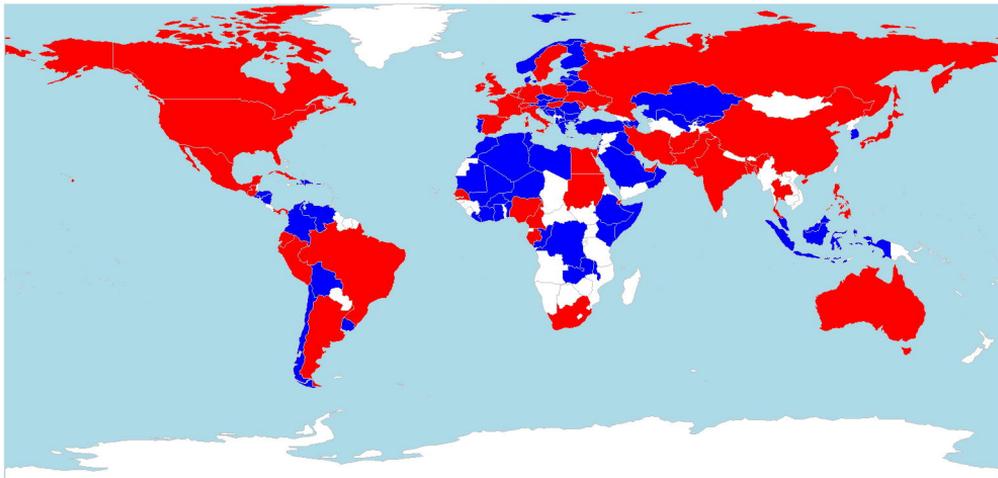
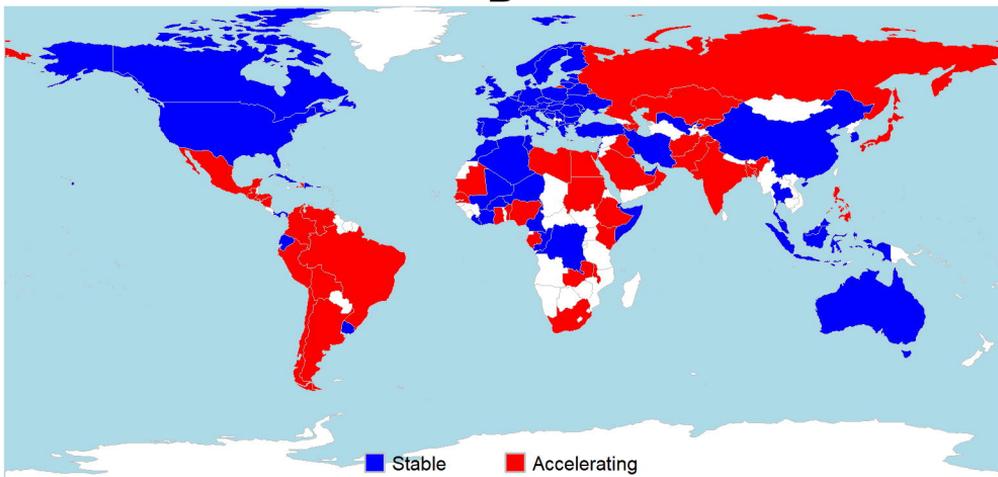


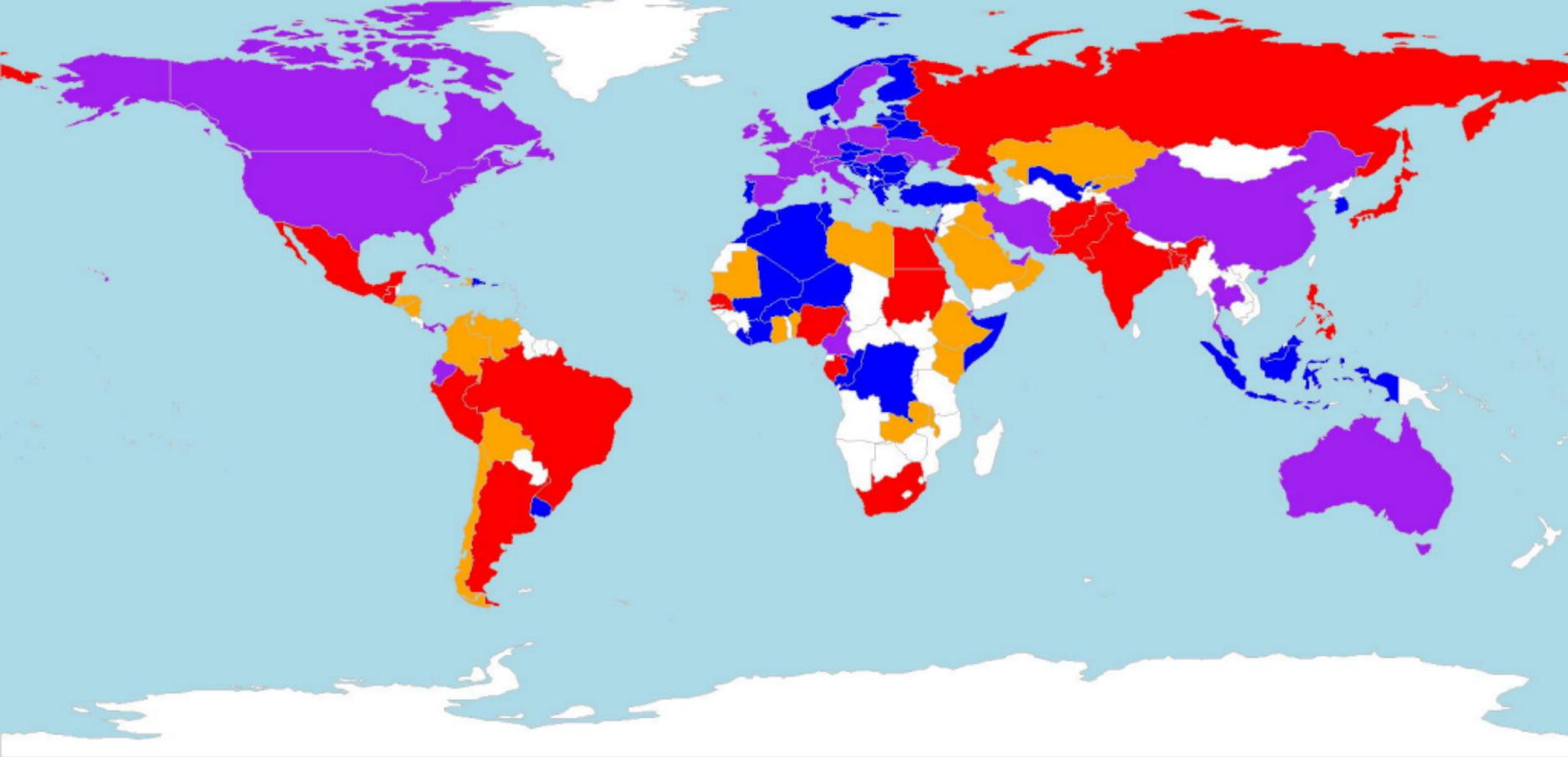


IFR (%) (95% CI)

0.20 (0.18-0.23) 0.38 (0.30-0.49) 0.93 (0.78-1.03)



**A****B**



■ Stable

■ Improved

■ Worse

■ Accelerating