1 NTT Docomo and Apple mobility data compared as countermeasures against
2 COVID-19 outbreak in Japan

3

4

5 Yoshiyuki Sugishita¹⁾, Junko Kurita²⁾, Tamie Sugawara¹⁾, Yasushi Ohkusa²⁾

6 1) National Institute of Infectious Diseases, Tokyo, Japan

7 2) Department of Nursing Collage of Nursing, Tokiwa University, Ibaraki, Japan

8

9

10

11 Keywords: COVID-19, prediction, Mobility Data, Apple, NTT docomo

12

13 Corresponding author: Junko Kurita, kuritaj@tokiwa.ac.jp

14

15

16 Abstract

17 Background: In Japan, as a measure to inhibit the COVID-19 outbreak, voluntary

18 restrictions against going out (VRG) have been applied.

19 Object: Mobility information provided by Apple Inc. and NTT Docomo were assessed

20 in terms of its usefulness in predicting conditions exacerbating an outbreak.

21 Method: A polynomial function was applied to daily Apple and Docomo data to

22 calculate the observed $R(t)$.

23 Results: The correlation coefficient among Apple and Docomo data was 0.91. The

24 adjusted coefficient of determination for $R(t)$ for the whole study period was higher

25 using Docomo data than when Apple data were used. When we regressed $R(t)$ on daily

26 Apple and Docomo data simultaneously, the estimated coefficient of Docomo data was

27 not significant.

28 Discussion and Conclusion: We demonstrated that Apple mobility data might be

29 superior to Docomo data for explaining the entire course of the COVID-19 outbreak in

30 Japan.

31

32 Introduction

33 To support planning and evaluation, some methods to avoid and overcome COVID-19
34 outbreak peaks can be inferred. Those two methods are herd immunity [1] and infection
35 countermeasures [2].

36 In Japan, in preference to lockdowns such as those mandated in European and
37 North American countries, voluntary restrictions against going out (VRG) were
38 announced by national and local governments from the end of March [3]. However, the
39 VRG program intensity has changed over time. Public cooperation with VRG has also
40 been changing. Lockdowns such as those instituted by the respective governments in
41 Europe and North America might eventually engender cessation of widespread infection.
42 However, in Japan, efforts are voluntary: people can adjust their own degree of
43 cooperation independently. Therefore, the government should monitor current outbreak
44 phenomena when moderating requirements for VRG. One difficulty of such monitoring
45 efforts is that reporting the number of the newly infected patients entails some delay.
46 One delay is attributable to the incubation period from infection to onset; another delay
47 is that occurring after onset until reporting. Taken together, for approximately two
48 weeks, one is unable to observe a precise daily number of newly infected persons.
49 Therefore, because the latest information might not be timely, policies might be less
50 effective if a government waits two weeks to conduct decision-making.

51 Data for many variables monitored for decision making have been reported from
 52 several services, including those of Apple Inc. and Alphabet Inc. (hereinafter Apple and
 53 Google, respectively) worldwide, and Nippon Telegraph and Telephone (NTT) Docomo
 54 (hereinafter Docomo) and East and West Japan Railway companies (JR) in Japan. Of
 55 those companies, Apple and Docomo started to publish related daily data from January
 56 [6,7]. Docomo data were used as an index of VRG by the government. Then one
 57 question must be raised: Which dataset can explain the outbreak better? The present
 58 study compares two data sources from the perspective of their power to predict
 59 outbreaks.

60 Going out was defined as a route search on Apple map using Apple data. Therefore,
 61 it excludes going out without searching for a well known place; it also incorrectly
 62 includes searches done at home. By contrast, Docomo data are based on the mobile
 63 phone location, as measured by 500 m grid. Therefore, it might be a more appropriate
 64 definition of going out.

65

66 Methods

67 We first estimate the effective reproduction number $R(t)$ in Japan assuming an
 68 incubation period following the empirical distribution in Japan. The number of

69 symptomatic patients reported by the Ministry of Labour, Health and Welfare (MLHW)
 70 for January 14 – May 25 published [7] on May 27 were used. Some patients were
 71 excluded from data: those presumed to be persons infected abroad or infected as
 72 passengers on the Diamond Princess. Those patients were presumed not to represent
 73 community-acquired infection in Japan.

74 For the onset dates of some symptomatic patients that were unknown, we estimated
 75 their onset date from an empirical distribution with durations extending from onset to
 76 the report date among patients for whom the onset date had been reported. We estimated
 77 the onset date of patients for whom onset dates were not reported as follows: Letting
 78 $f(k)$ represent this empirical distribution and letting N_t denote the number of patients for
 79 whom onset dates were not published by date t , then the number of patients for whom
 80 the onset date was known is $t-1$. The number of patients for whom onset dates were not
 81 available was estimated as $f(1)N_t$. Similarly, the number of patients with onset date $t-2$
 82 and for whom onset dates were not available was estimated as $f(2)N_t$. Therefore, the
 83 total number of patients for whom the onset date was not available, given an onset date
 84 of s , was estimated as $\sum_{k=1}^{\infty} f(k)N_{s+k}$ for the long duration extending from s .

85 Moreover, the reporting delay for published data from MHLW might be
 86 considerable. In other words, if $s+k$ is larger than that in the current period t , then $s+k$

87 represents the future for period t . For that reason, N_{s+k} is not observable. Such a
 88 reporting delay engenders underestimation bias of the number of patients. For that
 89 reason, it must be adjusted as $\sum_{k=1}^{t-s} f(k)N_{s+k} / \sum_{k=1}^{t-s} f(k)$. Similarly, patients for whom the
 90 onset dates were available are expected to be affected by the reporting delay. Therefore,
 91 we have $M_{s|t} / \sum_{k=1}^{t-s} f(k)$, where $M_{s|t}$ represents the reported number of patients for whom
 92 onset dates were within period s , extending until the current period t .

93 We defined $R(t)$ as the number of infected patients on day t divided by the number
 94 of patients who were presumed to be infectious. The number of infected patients was
 95 calculated from the epidemic curve by the onset date using a distribution of the
 96 incubation period. The distribution of infectiousness in symptomatic and asymptomatic
 97 cases was assumed to be 30% on the onset day, 20% on the following day, and 10% for
 98 the subsequent five days [8].

99 To clarify associations among the $R(t)$ and mobility data, we regressed $R(t)$ on a
 100 polynomial function of daily Apple and Docomo data, respectively. Moreover, we
 101 regressed $R(t)$ simultaneously on a polynomial function of daily Apple and Docomo
 102 data. In any case, the order of the polynomial function was selected stepwise while all
 103 coefficients were significant.

104 The study period was February 10 through May 25. Moreover, we analyzed only

105 those data recorded after March 10. In both cases, the data used for regression were

106 those up to the end of April. Then we evaluated the predictive power in May.

107

108 Ethical consideration

109 All information used for this study has been published [6–8]. Therefore there is no
110 ethical issue related to this study.

111

112

113 Results

114 As of May 27, using data for January 14 – May 25 in Japan, 14,972

115 community-acquired cases were identified, excluding asymptomatic cases. Figure 1

116 presents an empirical distribution of the duration of onset to reporting in Japan. The

117 maximum delay was 30 days. Figure 2 depicts the empirical distribution of incubation

118 periods among 125 cases for which the exposed date and onset date were published by

119 MHLW in Japan. The mode was six days. The average was 6.6 days.

120 Figure 3 presents data of Apple and Docomo. Their correlation coefficient among

121 was 0.91. It was high, but not extremely high. Estimation results are summarized in the

122 table. The order polynomial function was determined as one except for Docomo data in

123 the whole period as three. The adjusted coefficients of determination, which were used

124 as an index of the goodness of fit, were higher for Docomo data during the entire period.

125 However, after March 10, the adjusted coefficients of determination were better in

126 Apple data.

127 When we regressed $R(t)$ on a polynomial function of daily Apple and Docomo data

128 simultaneously, the estimated coefficient of Apple data was 0.0648 ($p = 0.000$), but the

129 estimated coefficient of Docomo data was -0.01577 ($p = 0.386$) throughout the whole

130 period. After March 10, for Apple data, it was 0.0640 ($p = 0.000$), but for Docomo data,

131 it was 0.0108 ($p = 0.627$). Therefore, Apple data were significant conditional on

132 Docomo data. However, Docomo data were not significant conditional on Apple data.

133 Figure 4 depicts the observed $R(t)$ and prediction lines from Apple and Docomo

134 data since March 10. Clearly, neither dataset is able to explain the peak of $R(t)$ around

135 mid-March. Especially, Docomo data were found to have less goodness of fit than

136 Apple data. Conversely, in early April, prediction by Docomo data might overshoot.

137

138 Discussion

139 We showed that Apple data and Docomo data were similar, but Apple data were

140 more informative than Docomo data. Apple data might sufficiently reflect the situation

141 of the COVID-19 outbreak in real time.

142 The present study has some limitations. First, although we examined the
143 explanatory power for the COVID-19 outbreak nationwide, its applicability at the
144 prefecture level must be verified. Secondly, one must be reminded that Apple and
145 Docomo data indicate the proportion of users in an area who leave their residence. The
146 data do not directly indicate a number of contacts or even a rate of contact. In other
147 words, Apple and Docomo data reflect no intensity of the respective contacts. In fact,
148 such measurement of contact intensity is extremely difficult. Methods to obtain such
149 measurements stand as an objective for future research.

150

151 Conclusion

152 We demonstrated that mobility data from Apple might be better than Docomo data
153 for explaining the entire course of the outbreak in COVID-19 in Japan. Therefore,
154 monitoring Apple data might be sufficient to adjust control measures to maintain the
155 effective reproduction number as less than one. Recently, though it was beyond the
156 study duration of the present study, as of the end of May, Apple data indicate the
157 predicted $R(t)$ as about 1.6. The importance of using Apple data for real time
158 recognition of characteristics of the COVID-19 outbreak is expected to increase.

159

160

161 Acknowledgments

162 We acknowledge the great efforts of all staff at public health centers, medical
163 institutions, and other facilities who are fighting the spread and destruction associated
164 with COVID-19.

165

166 Reference

- 167 1. Advisory meeting for COVID-19. Situation awareness and recommendation for
168 counter measure against COVID-19 (May 1,2020)
169 <https://www.mhlw.go.jp/content/10900000/000627254.pdf> (in Japanese)
170 [accessed on June 10, 2020]
- 171 2. Kurita J, Sugawara T Ohkusa Y. Preliminary evaluation of voluntary event
172 cancellation as a countermeasure against the COVID-19 outbreak in Japan as of 11
173 March, 2020. <https://www.medrxiv.org/content/10.1101/2020.03.12.20035220v1>
- 174 3. Japan Times. Tokyo governor urges people to stay indoors over weekend as virus
175 cases spike
176 [https://www.japantimes.co.jp/news/2020/03/25/national/science-health/tokyo-logs-](https://www.japantimes.co.jp/news/2020/03/25/national/science-health/tokyo-logs-40-coronavirus-cases/#.Xr4TV2eP604)
177 [40-coronavirus-cases/#.Xr4TV2eP604](https://www.japantimes.co.jp/news/2020/03/25/national/science-health/tokyo-logs-40-coronavirus-cases/#.Xr4TV2eP604) [accessed on May 14, 2020]
- 178 4. Kurita J, Sugawara T Ohkusa Y. Forecast of the COVID-19 outbreak, collapse of
179 medical facilities, and lockdown effects in Tokyo, Japan.
180 <https://medrxiv.org/cgi/content/short/2020.04.02.20051490v1>
- 181 5. Ohkusa Y, Sugawara T, Taniguchi K, Okabe N. Real-time estimation and
182 prediction for pandemic A/H1N1(2009) in Japan. J Infect Chemother.
183 2011;17:468-72.
- 184 6. Apple. Mobility trend Data.<https://www.apple.com/covid19/mobility> (in Japanese)
185 [accessed on May 13]
- 186 7. Mizuno laboratory, Special site for COVID-19 : Visualization for voluntary
187 restrictions against going out. <http://research.nii.ac.jp/~mizuno/> (in
188 Japanese)[accessed on May 27]

- 189 8. Japan Ministry of Health, Labour and Welfare. Press Releases.
190 https://www.mhlw.go.jp/stf/newpage_10723.html (in Japanese) [accessed on May
191 14, 2020]
- 192 9. Kimball A, Hatfield KM, Arons M, James A, Taylor J, Spicer K, Bardossy AC,
193 Oakley LP, Tanwar S, Chisty Z, Bell JM, Methner M, Harney J, Jacobs JR, Carlson
194 CM, McLaughlin HP, Stone N, Clark S, Brostrom-Smith C, Page LC, Kay M,
195 Lewis J, Russell D, Hiatt B, Gant J, Duchin JS, Clark TA, Honein MA, Reddy SC,
196 Jernigan JA; Public Health ? Seattle & King County; CDC COVID-19
197 Investigation Team. Asymptomatic and Presymptomatic SARS-CoV-2 Infections in
198 Residents of a Long-Term Care Skilled Nursing Facility - King County,
199 Washington, March 2020. Morb Mortal Wkly Rep. 2020 Apr 3;69(13):377-381.
200 doi: 10.15585/mmwr.mm6913e1.
- 201 10. Kurita J, Sugawara T Ohkusa Y. Evaluating Apple Inc. mobility trend data related
202 to the COVID19 outbreak in Japan.
203 <https://www.medrxiv.org/content/10.1101/2020.05.01.20087155v2>
- 204 11. Zhao S, Lin Q, Ran J, Musa SS, Yang G , Wang W, Lou Y, Gao D, Yang L, He D,
205 Wang M. Preliminary Estimation of the Basic Reproduction Number of Novel
206 Coronavirus (2019-nCoV) in China, From 2019 to 2020: A Data-Driven Analysis in
207 the Early Phase of the Outbreak. Int J Infect Dis 2020[Online ahead of print]
- 208 12. Liu Y, Gayle AA, Wilder-Smith A, Rockly J. The reproductive number of
209 COVID-19 is higher compared to SARS coronavirus. J Travel Med. 2020.DOI:
210 10.1093/jtm/taaa021
- 211 13. Lai C, Shih T, Ko W, Tang H, Hsueh P. Severe Acute Respiratory Syndrome
212 Coronavirus 2 (SARS-CoV-2) and Coronavirus disease-2019 (COVID-19): The

213 Epidemic and the Challenges. Int J Antimicrob Agents.

214 DOI:10.1016/j.ijantimicag.2020.105924

215 14. Bergman N, Fishman R. Mobility Reduction and Covid-19 Transmission Rates.**doi:**

216 <https://doi.org/10.1101/2020.05.06.20093039>

217

218

1 Figure 1: Empirical distribution of duration from onset to report by MLHW, Japan.

2 (%)

3

4 Note: Bars represent the probability of duration from onset to report based on 657

5 patients for whom the onset date was available in Japan. Data were obtained from

6 MLHW, Japan.

1 Figure 2: Empirical distribution of the incubation period published by MLHW, Japan.

2 (%)

5 Notes: Bars show the distribution of incubation periods for 91 cases for which the
6 exposure date and onset date were published by MLHW, Japan. The patients for whom
7 incubation was longer than 14 days are included in the bar shown for day 14.

8

9

1 Figure 3: Proportion of going out in Apple and Docomo data.

2

3

4 Note: The red line represents the proportion of going out in Apple data comparison with

5 normal level as January 13. The green line represents proportion of going out in

6 Docomo data defined by 500 m mesh.

7

1 Figure 4: Observed $R(t)$ and the fitted lines from Apple or Docomo data since March
2 10.

1 Table: Estimation results of estimate R(t) on the Apple or Docomo data

	Apple				Docomo			
	Whole period		After March		Whole period		After March 10	
	Est.	<i>p</i>	Est.	<i>p</i>	Est.	<i>p</i>	Est.	<i>p</i>
	coeff.		coeff.		coeff.		coeff.	
Linear term	0.000		0.072	0.00	-1.098	0.003	0.08247	0.000
	0.0521			0				
Quadratic					0.018	0.001	-	
term					81			
Cubic term					-0.967	0.000		
					0×10^{-4}			
Constant	-2.570	0.000	-3.969	0.00	20.47	0.009	-3.085	0.000
				0				
Adjusted	0.4191		0.6422		0.4508		0.5516	
determinant								
coefficient								
No.	81		52		81		52	

observations

- 2 Note: The dependent variable was $R(t)$ and explanatory variables were polynomial
- 3 function of Apple or Docomo data. The order was selected stepwisely while all
- 4 coefficients were significant.