medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069724; this version posted May 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

ESTIMATION OF LENGTH OF IN-HOSPITAL STAY USING DEMOGRAPHIC DATA OF THE FIRST 1000 COVID-19 PATIENTS IN SINGAPORE

A PREPRINT

Jaya Sreevalsan-Nair* **Reddy Rani Vangimalla** **Pritesh Rajesh Ghogale**
Graphics-Visualization-Computing Lab *and* E-Health Research Center
International Institute of Information Technology Bangalore, Karnataka 560100, India
http://www.iiitb.ac.in/gvcl http://ehrc.iiitb.ac.in

May 26, 2020

ABSTRACT

COVID-19 has been declared as a global pandemic by the World Health Organization (WHO) on March 11, 2020. In this paper, we investigate various aspects of the clinical recovery of the first 1000 COVID-19 patients in Singapore, spanning from January 23 to April 01, 2020. This data consists of 245 clinically recovered patients. The first part of the paper studies the descriptive statistics and the influence of demographic parameters, namely age and gender, in the clinical recovery-period of COVID-19 patients. The second part of the paper is on identifying the distribution of the length of the in-hospital stay (LOS) for the patients. Based on the hospital protocols followed in Singapore, this is equivalent to the virus shedding period. We determine a periodization based on critical changes in the transmission model. As expected, the overall recovery rate has reduced drastically during the exponential increase of incidences. However, our work on identifying parameters for a regression model shows that there is a shift in the age-group of incidences to the younger population, and the recovery-period of the younger population is considerably lower. Here, we have estimated the most likely length of in-hospital stay to be 9 days, as opposed to 15 or 20 days. Overall, the prognosis of COVID-19 indicates a reduction in length of in-hospital stay owing to the government-mandated practices of restricted mobility of the older population and aggressive contact tracing.

***K*eywords** COVID-19, Length of in-hospital stay, Regression

1 Introduction

The viral contagion, named COVID-19, has been declared as a global pandemic by the World Health Organization (WHO) on March 11, 2020[2]. The pandemic, characterized by atypical pneumonia, is caused by a virus from the coronavirus family, namely SARS-CoV2 (Severe Acute Respiratory Syndrome Coronavirus-2), which is a positive-sense single-stranded RNA virus. As of April 2, 2020, there are 827,419 positively confirmed cases, and 40,777 deaths, spread across 206 countries [3]. The total number of recovered patients in an unofficial count is 193,989 out of 935,197 positively confirmed patients, which implies that the ratio of the recovered to the infected patients, r_{ri} is $\sim$0.2, as of April 1, 2020 [4].

In this paper, we analyze the statistics of hospital recovery of patients tested positive for COVID-19 infection in Singapore [1]. The case study of Singapore has been carefully chosen owing to the reliable, accessible, and available data from the official press releases of the Ministry of Health (MoH), Government of Singapore. The healthcare

*jnair@iiitb.ac.in

[2]https://www.who.int/dg/speeches

[3]https://www.who.int/emergencies/diseases/novel-coronavirus-2019

[4]As retrieved on April 02, 2020, from https://www.worldometers.info/coronavirus/

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

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069724; this version posted May 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

system of Singapore has been unique in its handling of the widespread contagion in terms of imposing strict lockdown, quarantine, and isolation, and aggressive, large-scale contact tracing and testing. 1000 individuals have been confirmed positive for COVID-19 during January 23-April 01, 2020, of which 245 have been discharged after clinical recovery. This gives an overall r_{ri} higher than that of the world, as r_{ri}=0.245. The positive confirmation has been made using the real-time reverse transcription polymerase chain reaction (RT-PCR) tests on respiratory samples (sputum or nasal/throat/nasopharyngeal swabs), based on experiential learning from the outbreak of SARS in 2003 [5]. Similarly, the protocol for clinical recovery or hospital discharge has been based on the results on the RT-PCR tests of two consecutive samples being negative over two days [2]. The other clinical outcome, apart from recovery, is death. The first two deaths in Singapore owing to the COVID-19 contagion occurred on March 21, 2020, and increased to three by April 01.

Since the SARS outbreak in 2003, Singapore has systematically strengthened the system of managing the spread of infectious diseases [2]. The measures include opening dedicated facilities (the National Center for Infectious Diseases (NCID), National Public Health Laboratory, and more biosafety level-3 laboratories), increasing capacity in the public healthcare system (e.g., negative pressure isolation beds, personal protective equipment, trained health professionals), and deploying formal (digital) platforms for inter-governmental agency cooperation. For containing the spread of contagions, systems have been in place for upscale, quick-responsive, and aggressive contact tracing at entry points of the country (airports) and through local healthcare providers. There has been a holistic improvement, supported by increased economic investment, in building expertise in infectious disease management. This organized system has thus facilitated a controlled management of the pandemic COVID-19 in Singapore with patient-wise reporting to the public. A demographic analysis of the discharged patients in Singapore enables a systematic understanding of the influence of government policies on the clinical recovery from COVID-19 in the country. We study the clinical recovery from the data pertaining to the length of in-hospital stay (LOS). This article also emphasizes the *reliability* of such a study owing to the credibility of the open data made available by the government.

The following timeline of nationwide measures mandated by the Government of Singapore is relevant for our study:

1. January 23, 2020: Contact tracing had commenced. Singapore had expanded travel advisory that recommends Singaporeans to avoid travel to Hubei Province, China, in addition to the travel restrictions that China has imposed on Huanggang and Ezhou. Hospital surveillance of local cases had also begun.
2. February 04, 2020: The first four cases of local transmission (cases 19, 20, 21, and 24) were identified. The early identification has been a result of enhanced surveillance at the hospitals. Additional measures to contain the risk of spread beyond such local clusters to the broader community had also been enhanced. Aggressive contact tracing, and expanded surveillance efforts to conduct testing on all pneumonia cases in the hospitals were intensified.

   Precautions had been implemented to reduce spread in vulnerable groups, such as,

   - Measures in schools to lower the risk of transmission among students and staff , including staggered recesses and suspension of school assemblies,
   - Suspension of large group and communal activities in preschools,
   - Suspension of external excursions and large scale gatherings in eldercare facilities.
3. March 3, 2020: The border started banning travelers from countries which were hotspots. A complete ban on international travel into Singapore was implemented on March 22.
4. March 16-19, 2020: The Ministry of Education and the Ministry of Foreign Affairs, with support from the Singapore Airlines, began repatriation of Singaporeans working and studying outside of Singapore. This activity continued until the end of April.

One of the significant benefits of aggressive contact tracing has been hospital isolation within 5 days from the onset of symptoms [3]. That said, 12.6% of transmission has been found to be presymptomatic [4], as analyzed for seven clusters found in Singapore. Thus, there is an innate uncertainty in the onset of symptoms in the symptomatic cases to a positive confirmation. On the other hand, the hospital stay from positive confirmation to discharge upon negative results exhibits more cohesive statistics, as observed in similar studies [5].

South Korea, similar to Singapore, implemented all measures, but did not focus on early containment of local transmission. South Korea crossed the first 1000 positive confirmations[6] during January 20-February 26, 2020, i.e., within 38 days, compared to 70 days in Singapore. Thus, the slow transmission of COVID-19 in Singapore indicates that its early measures have been effective in reducing local transmission. We hypothesize that these measures would have indirectly reduced LOS for COVID-19. Hence, in this work, we perform statistical analysis of the length of

[5] https://www.cdc.gov/mmwr/preview/mmwrhtml/mm5218a1.htm

[6] https://www.cdc.go.kr/board/board.es?mid=a30402000000&bid=0030&tag=&act=view&list_no=366352

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069724; this version posted May 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

in-hospital stay of the COVID-19 patients. An additional motivation behind studying the clinical recovery of patients is to assess the overall load on the healthcare system in terms of patient occupancy as the hospital stays determines the load. The demographic analysis of the recovered patients gives insight to shifts in gender and age-groups in the recovery of patients. This analysis further complements the studies with observation in the transmission rate is higher in older males with comorbidities [6]. Fitting the data of hospitalization period to a statistical distribution is essential for estimating Δt_{LOS}.

Since the contagion has time-varying reproduction number (R_t) with characteristic trends in specific time-periods [7], we *periodize* the time-period of January 23-April 01, to perform a piecewise analysis of the timeline [7, 8]. We perform two analyses on the periodized timeline. Firstly, we study the age-gender distribution of the patients who have been confirmed positive of COVID-19 and those who have clinically recovered. Secondly, we extract the distribution of the LOS and fit regression models. In both analyses, we discuss the observable period-wise shifts in trends and their influencing factors, thus estimating Δt_{LOS}. The novel contribution of our work is in estimation of the length of in-hospital stay in Singapore by determining regression model parameters.

2 Methods

The data for our work has been collated from the public press releases made by the MoH, the Government of Singapore [7]. This dataset includes the case-ID's, age, gender, positive confirmation date, discharge date, and date of onset of symptoms[8]. The data has been cross-verified with dashboard [9] for case details. We have analyzed this patient-wise data pertaining to age, gender, and timeline of the disease progression.

We define the *length of in-hospital stay* Δt_{LOS} as the time elapsed between positive COVID-19 confirmation using RT-PCR test, and the discharge date from hospital after two consecutive negative results, using RT-PCR tests. Owing to the strict protocols followed in the Singapore healthcare system, the recovery-period can be considered equivalent to the *virus shedding period*. In the existing literature, the virus shedding period is estimated to be 15 days [9] or 20 days [10]. Also, we consider Δt_{LOS} as an *observed* count variable.

Age-gender distribution: There has been early evidence of the influence of both age and gender in susceptibility of COVID-19 infection [6, 11]. Hence, we look at the influence of age and gender in the clinical recovery of patients, with a specific focus on the length of the recovery-period.

There are 1000 patients (576M, 424F) in the population confirmed COVID-19 positive, of which 245 patients (137M,108F) have clinically recovered (§Figure 1(i)). The age distribution of the positively confirmed patients is (13 in (0-9), 29 in (10-19), 273 in (20-29), 202 in (30-39), 155 in (40-49), 164 in (50-59), 110 in (60-69), 38 in (70-79), 15 in (80-89), 1 in (90+)) patients in age groups given in years in parentheses. Similarly the age distribution of the discharged patients is (4 in (0-9), 4 in (10-19), 36 in (20-29), 58 in (30-39), 44 in (40-49), 55 in (50-59), 34 in (60-69), 8 in (70-79), 2 in (80-89)). The preliminary counts indicate that the restrictions used for slowing the contagion down in the susceptible group of the population, namely the older males [11], have led to the shifts in distribution of contagion with respect to age and gender. There is a conspicuous shift to the younger population, namely, in the age groups [20-29], and [30-39], and more evenly across both genders.

Periodization: As the pandemic progresses, the evolution needs to be studied piecewise in different time periods [7, 8]. From the timeline of disease progression in Singapore, we have identified the following significant dates:

- On January 23, the first patient was confirmed COVID-19 positive, which is an imported case from Wuhan, China.
- On February 4, the local transmission began in Singapore, and coincidentally, the first clinically recovered patient was discharged from the hospital.
- On March 17, the total/cumulative number of positively confirmed patients has a spike with 44 new cases on a single day. This is the start of community transmission.
- On March 21, the first two deaths owing to complications from COVID-19 were recorded.
- On April 1, the number of confirmed patients reached 1000, with 245 discharged and 3 deceased.

[7] `https://www.moh.gov.sg/news-highlights/`

[8] The press releases did not carry patient-wise information after March 17, 2020, owing to which the date of onset of symptoms is not available beyond that date.

[9] `https://experience.arcgis.com/experience/7e30edc490a5441a874f9efe67bd8b89`

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069724; this version posted May 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

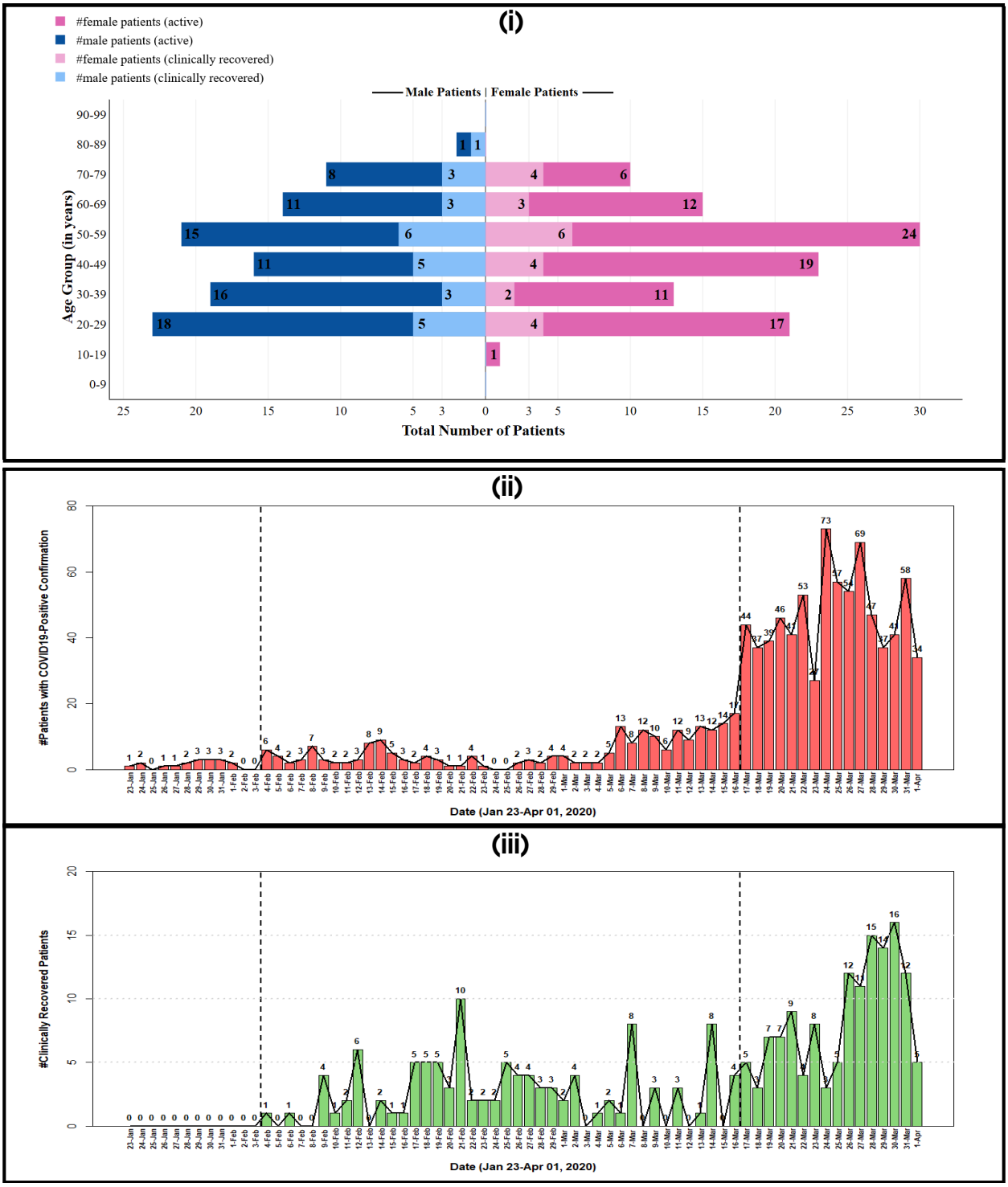


Figure 1: (i) Stacked population pyramid of age-gender distribution of population of the 245 discharged and 754 active (i.e., COVID-19 positive but not discharged, including the deceased) during the same period (1 active patient who is a 102 years old female and confirmed positive on April 01, 2020, has been excluded in this population pyramid). Daily profile of count of patients in Singapore during January 23-April 01, 2020, (ii) who were confirmed positive for COVID-19, with a total of 1000, and (iii) who got clinically recovered and were discharged from the hospitals, with a total of 245. The dashed lines indicate the three periods we have introduced in this work.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069724; this version posted May 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

These trends can be observed in the daily profile of patient counts (§Figure 1 (ii),(iii)), i.e., the cumulative number of positive confirmed cases, N_i, and the cumulative number of discharged or clinically recovered cases, N_r. We consider the following three periods:

1. Period P_1 during January 23-February 3, which is the *period of imported cases*, and also the *period with no clinically recovered cases*. $N_i = 18$, $N_r = 0$.
2. Period P_2 during February 4-March 16, which is the *period of local transmission*. $N_i = 243$, $N_r = 109$,
3. Period P_3 during March 17-April 1, which is the *period of community transmission*, where there is a sudden spike in daily incidences. $N_i = 1000$, $N_r = 245$, and there are three deceased.

Our periodization corresponds to the stages 2-to-4 of the five stages of transmission model of COVID-19 for a country, as published by the WHO[10]. The five stages are, namely, "no cases, sporadic cases, clusters of cases, community transmission, and country-wide transmission."

The absolute numbers indicate that r_{ri}, i.e., $\frac{N_r}{N_i}$, has increased from 0.00 at the end of P_1 to 0.45 $\left(=\frac{109}{243}\right)$ at the end of P_2, and again dipped to 0.245 $\left(=\frac{245}{1000}\right)$ at the end of P_3. The dip is unfavorable for our hypothesis of decreasing Δt_{LOS}, given that the number of positive COVID-19 confirmations has increased exponentially, starting the beginning of P_3. However, the absolute counts N_i and N_r do not explicitly show the shifts in the age-gender distribution of the infected population (§Figure 2(i)), when using age bins of 10 years. Thus, we propose computing the time-varying Δt_{LOS} using the shifts in the age-gender distribution in the periodized timeline.

LOS analysis: Our goal is to study the period-wise changes in Δt_{LOS} owing to the shift in the demographic structure. We perform descriptive statistical analysis using median and interquartile range (IQR), followed by fitting appropriate regression models. We consider two types of regression models. Firstly, we use the time-series of Δt_{LOS} and fit a line using the loess model. Loess is a non-parametric local regression model for smoothening empirical time-series data [12] and scatterplots [13]. Secondly, we use the number of patients recovering for a specific Δt_{LOS} as a count variable and fit multivariate linear regression models considering age, gender, and period as independent variables, and Δt_{LOS} as the dependent variable. We use generalized linear models (GLM) for regression, which is semi-parametric. LOS is known to have a *naturally skewed* distribution, for which GLM's such as the Poisson regression model (PRM) and negative binomial regression model (NBM) have been used [5]. Hence, we propose the use of PRM and NBM for modeling Δt_{LOS}. For the period-wise analysis of Δt_{LOS}, we group the clinically recovered patients using two strategies.

1. Grouping based on recovery date, $\mathbb{G}_{\mathbf{-cfrmDt}}$: The two groups of clinically recovered patients can be obtained as per the period in which their discharge/recovery date falls, namely, P_2 and P_3 (§Figure 2 (ii),(a)). $\mathbb{G}_{\mathbf{-cfrmDt}}$ corresponds to a group of 109 patients who were discharged in P_2, and 136 patients in P_3.
2. Grouping based on positive confirmation date, $\mathbb{G}_{\mathbf{+cfrmDt}}$: However, we can perform a finer-grain analysis of the groups of patients based on the period in which their date of positive confirmation/hospital admission falls (§Figure 2 (ii),(b)). This gives us groups of patients who got tested positive in a period and got clinically recovered during the entire period of our study here. This gives us 18 patients in P_1, 161 in P_2, and 66 in P_3.

Symptom-onset period Δt_{SO} analysis: We additionally have data of the date of onset of symptoms for 227 of the 245 clinically recovered patients, who were confirmed positive during P_1 and P_2. We define the *symptom-onset period*, Δt_{SO}, as the number of days between the onset of symptoms and positive confirmation of COVID-19. We use the time-series of Δt_{SO} to fit a loess model, similar to Δt_{LOS}.

3 Results

Table 1 gives the percentages of the count data presented in Figure 2.

Descriptive statistical analysis: We present the descriptive statistics either as a Δt_{LOS} five-number summary[11] and as "median [IQR]" of the observed count variable, i.e., Δt_{LOS}.

Δt_{LOS} observed during the entire period, January 23-April 01, has the following median and IQR values:

- Overall: 11 [7] days.
- Gender-wise: 11 [6.25] days for females and 11 [8] days for males (§Figure 3(i)).

[10] `https://www.who.int/publications-detail/responding-to-community-spread-of-covid-19`

[11] A five-number summary is (minimum, first quartile, median, third quartile, maximum) values of a (count) variable.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069724; this version posted May 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

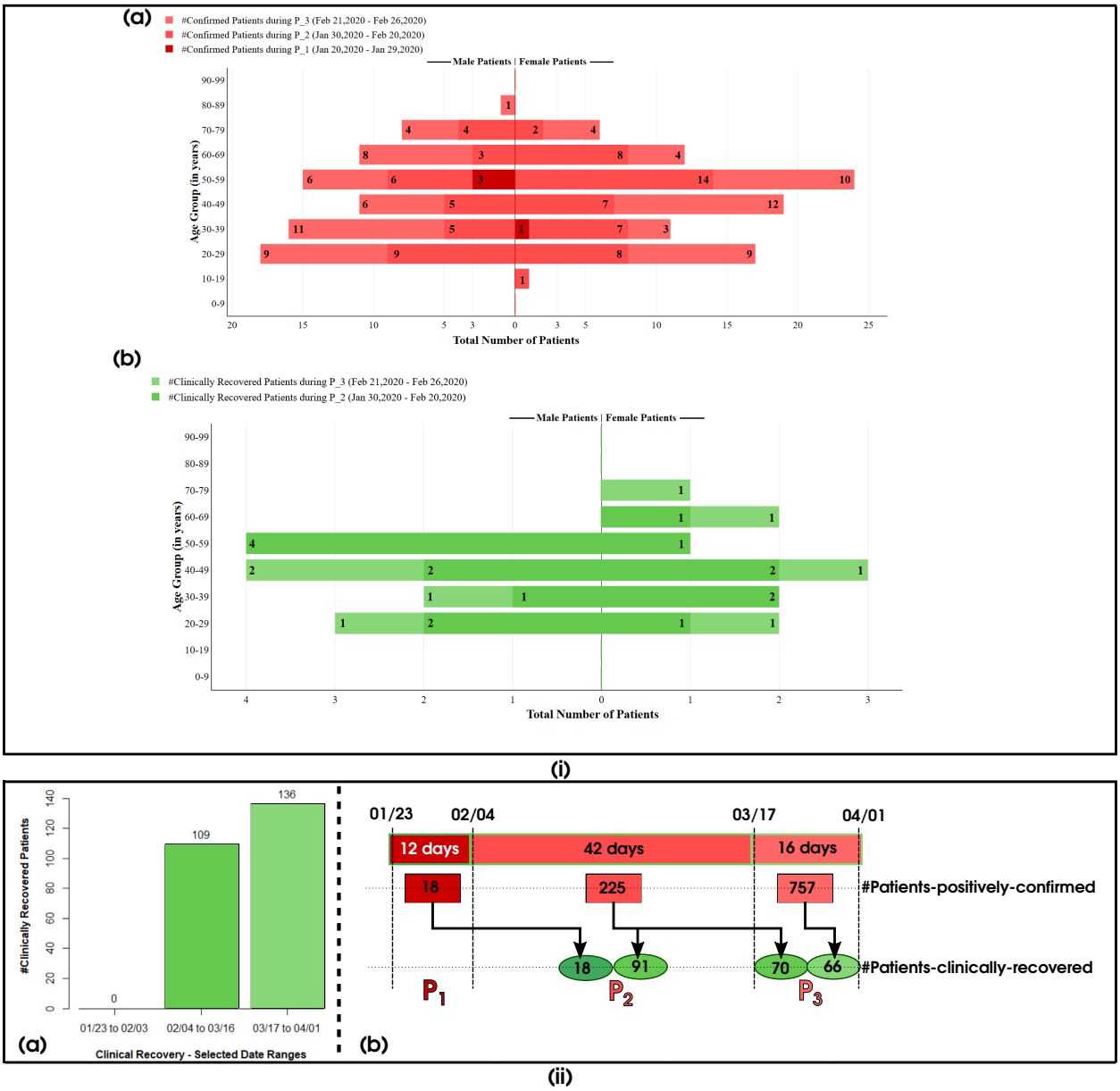


Figure 2: We periodize the entire period of our study into January 23-February 03 [P_1], February 04-March 16 [P_2], and March 17-April 01 [P_3]. (i) Period-wise stacked population pyramids to show the age-gender distribution of patients: ((i),(a)) 999 who were confirmed positive (excluding one 102-year old female patient who tested positive on April 01), and ((i),(b)) 245 who were discharged. (ii) Period-wise counts of discharged patients: ((ii),(a)) a barplot of the counts, and ((ii),(b)) a schematic diagram to show the counts of patients grouped based on periods when they were confirmed positive *and* discharged/clinically recovered.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069724; this version posted May 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Age-group (in years) → Gender ↓	(0-9)	(10-19)	(20-29)	(30-39)	(40-49)	(50-59)	(60-69)	(70-79)	(80-89)	(90+)	Total
Positive COVID-19 Confirmation (in %-age of 1000 patients)											
Male	0.6	1.2	15.4	12.4	9.8	9.3	6.1	2.2	6	0.0	57.6
Female	0.7	1.7	11.9	7.8	5.7	7.1	4.9	1.6	0.9	0.1	42.4
Total	1.3	2.9	**27.3**	**20.2**	15.5	16.4	11	3.8	1.5	0.1	100
Clinically Recovery (in %-age of 245 patients)											
Male	1.2	1.2	9.7	13.8	11	11	6	0.8	0.4	0	55.9
Female	0.4	0.4	4.9	9.8	6.9	11	7.7	2.4	0.4	0	44.1
Total	1.6	1.6	**14.6**	**23.7**	17.9	22.4	13.9	3.3	0.8	0	100

Table 1: Percentage values of the age-gender structure of population confirmed positive with COVID-19 during January 23-April 01, 2020 in Singapore (§Figures 1(i), and 2(i)).

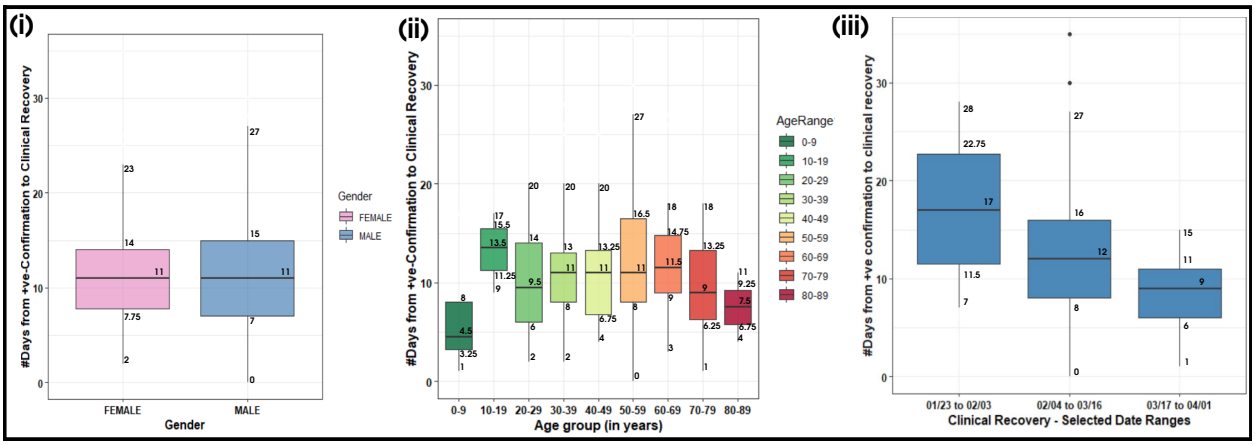


Figure 3: Box and whisker plots of 245 discharged patients out of 1000 positively confirmed COVID-19 patients during January 23-April 01, 2020, in Singapore, based on (i) gender, (ii) age, and (iii) period, based on $\mathbb{G}_{\mathbf{+cfrmDt}}$ grouping.

- Age-wise: 4.5 [4.75] days for (0-9), 13.5 [4.25] days for (10-19), 9.5 [8] days for (20-29), 11 [5] days for (30-39), 11 [6.5] days for (40-49), 11 [8.5] days for (50-59), 11.5 [5.75] days for (60-69), 9 [7] days for (70-79), and 7.5 [3.5] days for (80-89), for age groups in parentheses given in years (§Figure 3(ii)).
- Period-wise, based on $\mathbb{G}_{\mathbf{+cfrmDt}}$ grouping: 17 [11.25] days for P_1, 12 [8] days for P_2, and 9 [5] days for P_3 (§Figure 3(iii)).

We observe that the Δt_{LOS} is overall lesser in this dataset than reported in early analyses of COVID-19 patients, i.e., 15 days [9] and 20 days [10]. The overall median of Δt_{LOS} is 11 days, which is the same as the gender-weighted median, and the age-weighted and period-weighted medians are 10.7 and 11.6 days, respectively . Thus, the overall descriptive statistical analysis gives a conservative estimate of Δt_{LOS} to be $\sim$11 days.

There is a strong influence of age as well as period on Δt_{LOS}, unlike gender, where the median values are similar for both genders. Δt_{LOS} is relatively lower for the age group of (20-29) years, specifically. This age group comprises of 14.6% of the discharged patients (§Table 1). This result is significant as the age group of (20-29) years contributes the highest (27.3%) to the infected population. The infections have increased in this age group over time, as, 88.6% of the patients in this age group have been confirmed positive in P_3. Thus, our key conclusion is that since the most susceptible group of people has lower Δt_{LOS}, the overall Δt_{LOS} is bound to progressively decrease in the period after April 01, even further from our estimated value.

Δt_{LOS} monotonously decreases with progression in period. The period-wise five-number summaries for Δt_{LOS}, using $\mathbb{G}_{\mathbf{+cfrmDt}}$ grouping, are:

- Overall: (7, 11.5, 17, 22.75, 28) for P_1, (0, 8, 12, 16, 27) for P_2, and (1, 6, 9, 11, 15) for P_3.
- For females: (10, 11, 13, 22, 26) for P_1, (0, 7, 12, 15, 26) for P_2, and (3, 6.25, 9, 11, 14) for P_3.
- For males: (11, 17, 20, 23, 29) for P_1, (0, 8, 12, 16, 27) for P_2, and (1, 5.75, 8, 11, 15) for P_3.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069724; this version posted May 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Age-group	F	M	F	M
	Jan/23-Apr/01 §Figure 4(a)		P_1 [$\mathbb{G}_{\mathbf{+cfrmDt}}$] §Figure 4((b),(i))	
(0-9 years)	(4, 4, 4, 4, 4)	(1, 3, 5, 11, 17)	-	-
(10-19 years)	(12, 12, 12, 12, 12)	(9, 12, 15, 16, 17)	-	-
(20-29 years)	(3, 8.25, **10**, 14, 18)	(2, 6, **8.5**, 14.25, 20)	-	-
(30-39 years)	(3, 9, 11, 13, 18)	(2, 8, 12, 15, 25)	(11, 12, 13, 19.5, 26)	(17, 17, 17, 20, 20)
(40-49 years)	(4, 6, 11, 14, 25)	(4, 7, 11, 13, 18)	(10, 11.75, 13.5, 15.25, 17)	-
(50-59 years)	(3, 8.5, 13, 16.5, 25)	(0, 7.75, 10, 16.5, 27)	(13, 17.5, 22, 23.5, 25)	(11, 15.5, 20, 21.5, 23)
(60-69 years)	(3, 7, 10, 14, 18)	(8, 10, 12, 15, 18)	-	(27, 27, 27, 27, 27)
(70-79 years)	(4, 7.25, 9, 12.25, 14)	(1, 5.25, 9.5, 13.75, 18)	(10, 10, 10, 10, 10)	-
(80-89 years)	(4, 4, 4, 4, 4)	(11, 11, 11, 11, 11)	-	-
	P_2 [$\mathbb{G}_{\mathbf{+cfrmDt}}$] §Figure 4((b),(ii))		P_3 [$\mathbb{G}_{\mathbf{+cfrmDt}}$] §Figure 4((b),(iii))	
(0-9 years)	(4, 4, 4, 4, 4)	(1, 3, 5, 11, 17)	-	-
(10-19 years)	-	(9, 11, 13, 15, 17)	(12, 12, 12, 12, 12)	(15, 15, 15, 15, 15)
(20-29 years)	(6, 12, 18, 18, 18)	(2, 7, 13, 16, 20)	(3, 9, 10, 13, 14)	(4, 6, 8, 9, 13)
(30-39 years)	(3, 9, 11, 13, 18)	(2, 8, 12, 13, 18)	(3, 3.75, 7, 10.25, 11)	(4, 7, 9.5, 12, 15)
(40-49 years)	(11, 11, 13, 14, 14)	(5, 7.5, 11, 14, 18)	(5, 6, 7.5, 9.75, 11)	(4, 4.75, 9, 11.5, 13)
(50-59 years)	(3, 7.5, 12, 16.5, 23)	(0, 9.5, 16, 18.5, 30)	(8, 9, 10, 13, 13)	(4, 6.25, 8, 9.75, 11)
(60-69 years)	(3, 6, 10, 15, 26)	(8, 9.5, 12, 15, 18)	(8, 9, 10, 11, 12)	(10, 11, 12, 12.5, 13)
(70-79 years)	(4, 8.5, 13, 13.5, 14)	(18, 18, 18, 18, 18)	(7, 7.25, 7.5, 7.75, 8)	(1, 1, 1, 1, 1)
(80-89 years)	-	(11, 11, 11, 11, 11)	(4, 4, 4, 4, 4)	-
	P_2 [$\mathbb{G}_{\mathbf{-cfrmDt}}$] §Figure 4((c),(ii))		P_3 [$\mathbb{G}_{\mathbf{-cfrmDt}}$] §Figure 4((c),(iii))	
(0-9 years)	(4, 4, 4, 4, 4)	(1, 3, 5, 11, 17)	-	-
(10-19 years)	-	(9, 11, 13, 15, 17)	(12, 12, 12, 12, 12)	(15, 15, 15, 15, 15)
(20-29 years)	(6, 9, 12, 15, 18)	(2, 7, 15, 16, 20)	(3, 9, 10, 13.75, 18)	(4, 6, 8, 12, 14)
(30-39 years)	(9, 9, 11, 12.5, 13)	(2, 5.5, 10, 16.5, 28)	(3, 9, 11, 13, 18)	(8, 10.5, 12, 13, 15)
(40-49 years)	(4, 10, 12, 14, 17)	(5, 7, 9, 14, 16)	(5, 6, 9.5, 11.5, 13)	(4, 7.5, 11, 13, 18)
(50-59 years)	(3, 7.75, 11.5, 17, 25)	(0, 9.25, 11.5, 22.25, 30)	(4, 9.5, 13, 14.5, 17)	(4, 7.25, 9.5, 14.75, 18)
(60-69 years)	(3, 3, 4, 5, 6)	(8, 8.75, 10.5, 15.75, 16)	(8, 10, 12.5, 15.75, 18)	(9, 10.5, 13, 15, 18)
(70-79 years)	(10, 11.5, 13, 13.5, 14)	-	(4, 5.5, 7, 7.5, 8)	(1, 5.25, 9.5, 13.75, 18)
(80-89 years)	-	-	(4, 4, 4, 4, 4)	(11, 11, 11, 11, 11)

Table 2: The five-number summaries of age-gender-period based box and whisker plots in Figure 3 and 4.

The influence of gender reduces as the periods progress. We observe similar trends when considering $\mathbb{G}_{\mathbf{-cfrmDt}}$ grouping. The range, the IQR, and the median of Δt_{LOS} decrease from P_1 to P_3, with or without the information of gender. This indicates that irrespective of gender, the *measure of the spread* of Δt_{LOS} decreases with time, similar to the trend in the value of Δt_{LOS}. The age-wise minima reduce sharply from P_1 to P_2, and increase slightly further to P_3, indicating an overall trend of decrease in minima. This supports the overall decrease in Δt_{LOS} value from P_1 to P_3.

Table 2 shows the fine-grained five-number summaries of the age-gender-period based box and whisker plots in Figure 4. We observe that there is a strong influence of age and period on Δt_{LOS}. While the overall median values for females and males are similar, we observe that there are variations across different age groups. In particular, for the age group of (20-29) years, the median [IQR] value of males of 8.5 [8.25] days is lower than the median value of females of 10 [5.75] days. This also shows that there is a higher spread (IQR) of Δt_{LOS} values in males. We further observe that the gender difference in IQR occurs in P_3 in the $\mathbb{G}_{\mathbf{+cfrmDt}}$ grouping, equivalent to that of P_2 in the $\mathbb{G}_{\mathbf{-cfrmDt}}$ grouping. The minima and median values are overall lower in P_3 in the $\mathbb{G}_{\mathbf{+cfrmDt}}$ than the $\mathbb{G}_{\mathbf{-cfrmDt}}$ grouping. Since the protocols followed in the hospital are similar for patients with closer hospital admission dates, the Δt_{LOS} has more cohesive descriptive statistics in the $\mathbb{G}_{\mathbf{+cfrmDt}}$ grouping than the $\mathbb{G}_{\mathbf{-cfrmDt}}$ one. Hence, we use the $\mathbb{G}_{\mathbf{+cfrmDt}}$ grouping exclusively in the regression analysis.

Since the symptom-onset dates have been studied [7, 8], we report the five-number summaries of Δt_{SO}, for which data is available for P_1 and P_2 only.

- Overall: (0, 3, 6, 9, 16) for P_1 and (1, 2, 5, 8, 17) for P_2.
- For females: (0, 2, 5, 7, 12) for P_1, and (1, 3, 5, 8, 15) for P_2.
- For males: (1, 4.75, 8, 9, 15) for P_1, and (1, 2, 5, 8, 17) for P_2.

We observe that Δt_{SO} has similar IQR and median across P_1 and P_2, irrespective of gender. We attribute this to the continuous monitoring of susceptible cases in Singapore leading to lesser delays in positive confirmations, thus showing a lower measure of spread and overall lower values for Δt_{SO}. Overall, we do not emphasize on further analyzing Δt_{SO}, apart from the loess model, owing to the clinical uncertainties involved [4].

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069724; this version posted May 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

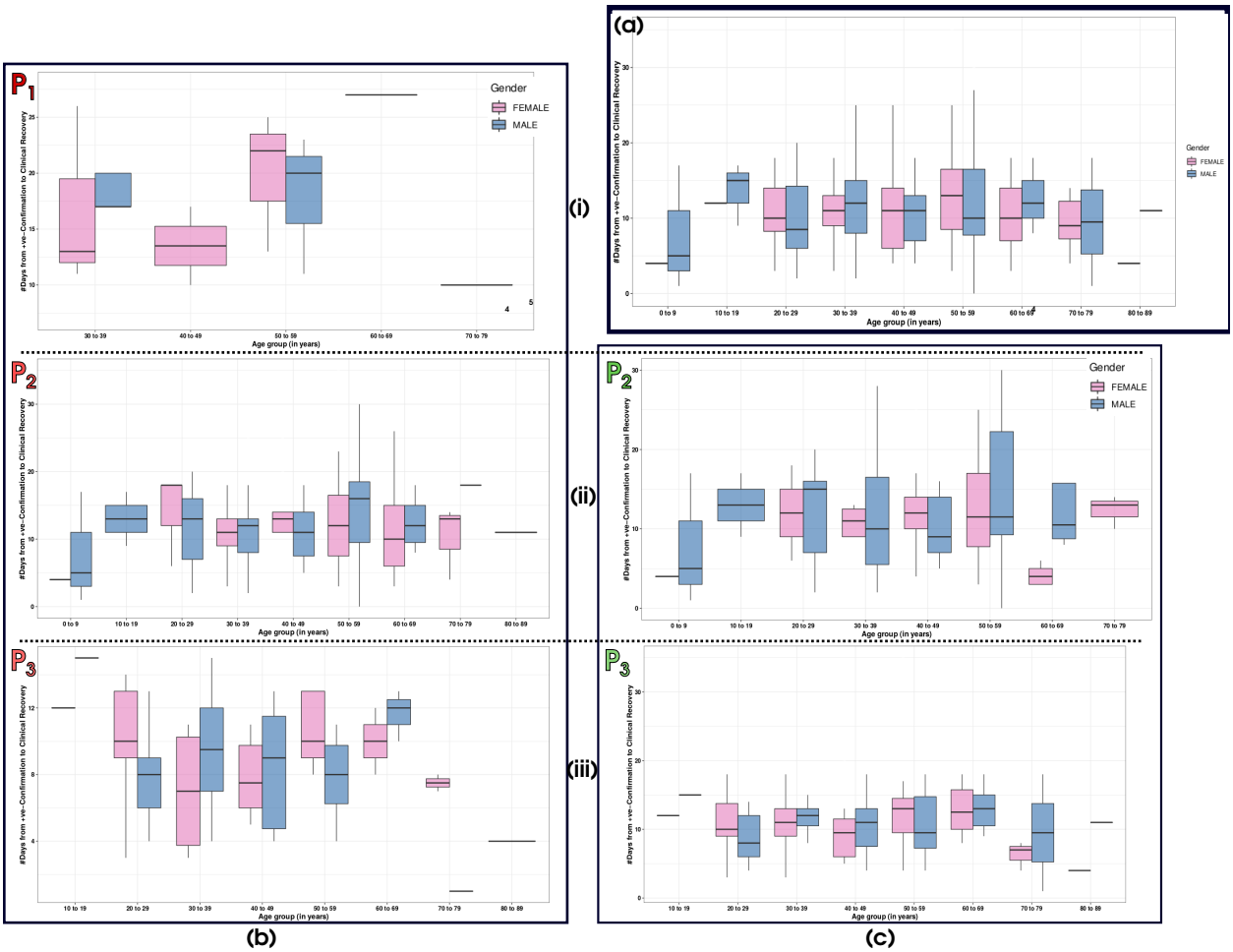


Figure 4: Gender- and age-wise box and whisker plots of 245 clinically recovered patients grouped in periods during (i) January 23-February 03 [P_1] (ii) February 04-March 16 [P_2], and (iii) March 17-April 01 [P_3]. Counterclockwise from top-right: (a) during January 23-April 01, 2020 [entire period], (b) period-wise grouping of patients based on COVID-19 positive confirmation date ($\mathbb{G}_{\mathbf{+cfrmDt}}$), and (c) period-wise grouping of patients based on clinical recovery/hospital discharge dates ($\mathbb{G}_{\mathbf{-cfrmDt}}$). The number of patients considered is (a) 245, ((b),(i)) 18, ((b),(ii)) 161, ((b),(iii)) 66, ((c),(ii)) 109, ((c),(iii)) 136.

Loess model: We now use the loess model to confirm the trend in the change in Δt_{LOS}. The loess model has been estimated on the time-series of the Δt_{LOS} and Δt_{SO} values, represented as scatter plots (§Figure 5). The loess model has been implemented using the stats.loess[12] in R [14]. We have considered the scatter plots based on the positive COVID-19 confirmation date, akin to $\mathbb{G}_{\mathbf{+cfrmDt}}$ grouping. The loess model for Δt_{LOS}, considered during January 23-April 01, 2020, uses 245 observations, with 5.37 degrees of freedom (DoF) and residual standard error (RSE) of 145.4 (§Figure 5(i)). The loess models on period-wise data (§Figure 5(ii)) have the following details: for P_1 uses 18 observations, with 4.66 DoF and RES of 5.817; for P_2 uses 161 observations, with 5.23 DoF and RSE of 61.74; and for P_3 uses 66 observations, with 4.91 DoF and RSE of 109.8. The loess model for Δt_{SO} uses 227 observations, with 5.03 DoF and RSE of 70 (§Figure 5(iii)).

The degrees of freedom roughly corresponds to the degree of the polynomial used to generate the fitting curve. Thus, both Δt_{LOS} and Δt_{SO} can be modeled using a 5th degree polynomial. Higher degree polynomial implies less bias but larger variance. Δt_{LOS} has a slope of -25° in P_3 when using the loess model for the entire time period (§Figure 5(i)) and that of -6° when using loess model for P_3 (§Figure 5(ii)). Thus, the key conclusion from the local regression model on the time-series is the negative slope, i.e., a *downward* trend in Δt_{LOS} in P_3.

[12]The loess model is a default local regression model used for a sample with less than 1000 observations in stats package in R.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069724; this version posted May 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

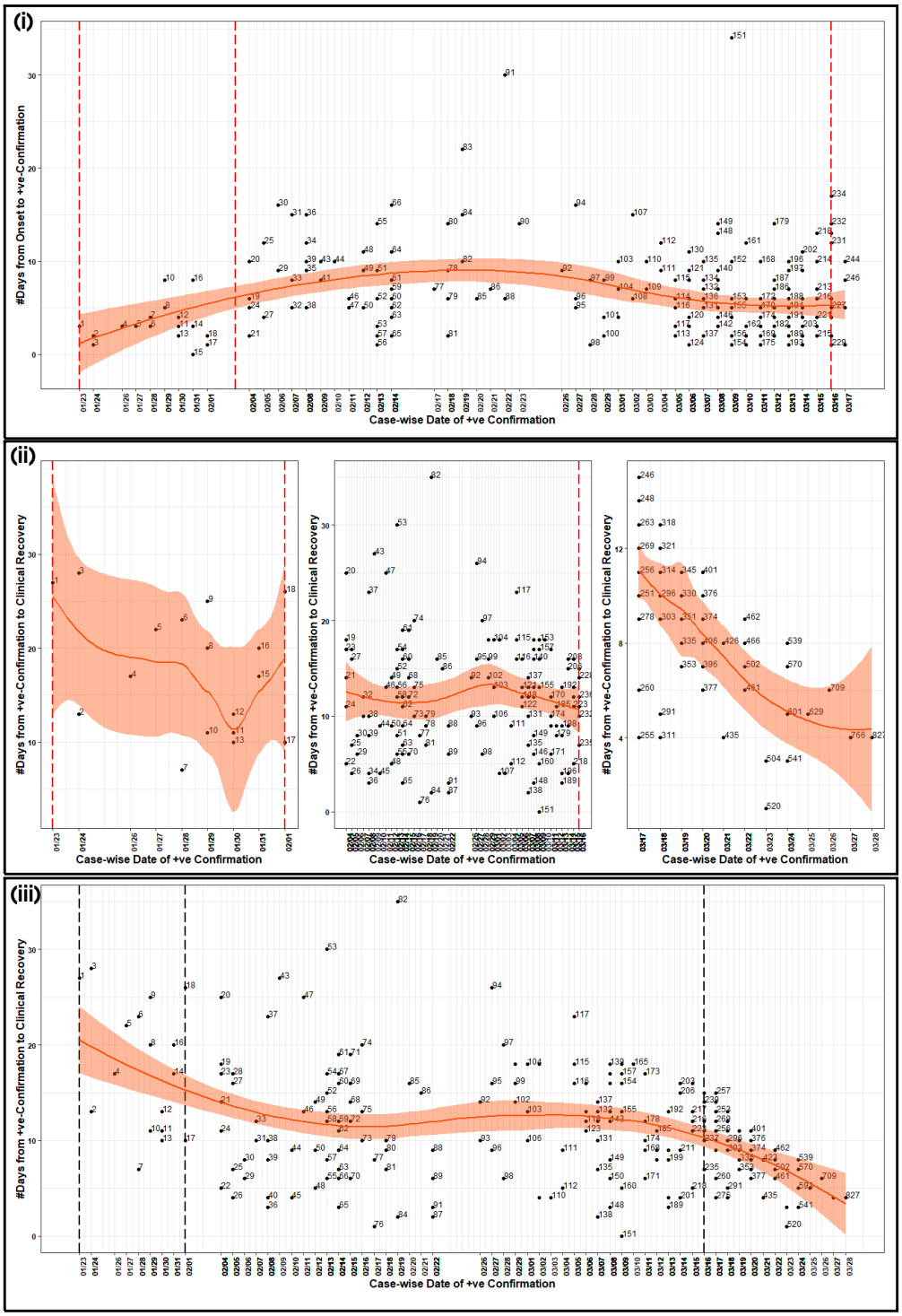


Figure 5: Scatter plot of the number of days with the timeline of positive confirmation of COVID-19 in Singapore during January 23-April 01, 2020, estimated using a Loess model. The scatter points using patient (case) ID, fitted loess curve and the error ribbon are shown for 245 discharged patients of (i) LOS, Δt_{LOS}, for the entire period, (ii) LOS, Δt_{LOS}, for each period, and for 227 patients of (ii) duration from the onset of symptoms to positive confirmation of COVID-19, Δt_{SO} for P_1 and P_2.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069724; this version posted May 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

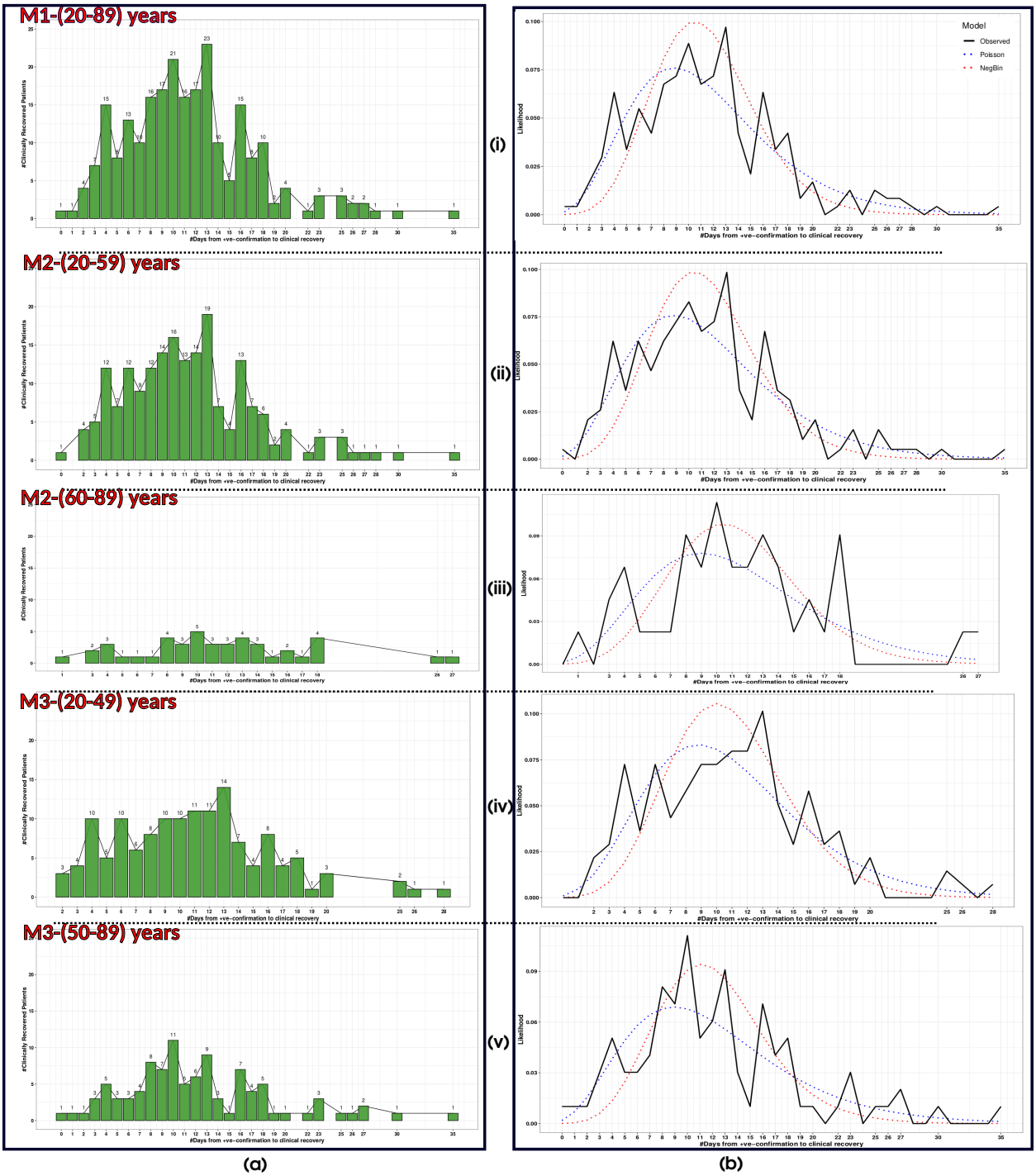


Figure 6: Multivariate linear regression model of the recovery-period Δt_{LOS} for all 237 discharged patients of (20-89) years of age amongst 245 COVID-19 discharged cases in Singapore during January 23-April 01, for (i) age (20-89) years [M_1], (ii) age (20-59) years [M_2], (iii) age (60-89) years [M_2], (iv) age (20-49) years [M_3] and (v) age (50-89) years [M_3]; uses (a) patient-count distribution and (b) fitting Poisson and negative binomial distributions. The number of patients considered is (i) 237, (ii) 193, (iii) 44, (iv) 138, and (v) 99, using $\mathbb{G}_{\mathbf{+cfrmDt}}$ grouping.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069724; this version posted May 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Multivariate linear regression model: Now that we have observed and concluded from both the descriptive statistical analysis and loess model that Δt_{LOS} is decreasing during the period of January 23-April 01, our next step is to predict the value of Δt_{LOS}. We experiment with the generalized linear model (GLM) for a multivariate linear regression model for Δt_{LOS} using Poisson (PRM) and negative binomial (NBM) distributions. Our choice of model and distributions are commonly used for count data [5], and LOS is commonly over-dispersed data [15].

We observe that the median Δt_{LOS} monotonously increases from age bins (20-29) to (60-69), and then monotonously decreases (§Figure 3 (ii)). We exclude age bins (0-9) and (10-19) for the modeling, owing to the anomaly in monotonous behavior. We also observe that gender does not influence Δt_{LOS} (§Figure 3 (i)) and Δt_{LOS} monotonously decreases with the period (§Figure 3 (iii)). Hence, we consider a GLM with piecewise linear models with respect to age. We consider three modeling scenarios, where we use age, gender, and period as categorical variables:

- M_1 which is a single GLM for age groups between (20-29) and (80-89);
- M_2 with two GLM's for age groups between (20-29) and (50-59), and for those between (60-69) and (80-89);
- M_3 with two GLM's for age groups between (20-29) and (40-49), and for those between (50-59) and (80-89).

We use the Akaike Information Criterion (AIC) and its corrected version for a small sample size (AIC_c) for determining the *goodness of fit* of our proposed models. We have implemented these models using the stats.glm in R [14]. For GLM with Poisson and binomial families, the dispersion is fixed at 1.0, and the number of parameters (k) is the same as the number of coefficients in the regression model [14]. The negative binomial distribution has an additional parameter to model over-dispersion in the data. For the number of samples (n) in the data, AIC is used if $\left(\frac{n}{k} > 40\right)$, and AIC_c is used otherwise [16]. Thus, we use AIC and AIC_c appropriately for scenarios based on the $\frac{n}{k}$ value (§Table 3).

Model Specs. \Scenario	M_1 (age:(20-89))	M_2 (age:(20-59))	M_2(age:(60-89))	M_3 (age:(20-49))	M_3 (age:(50-89))
#Samples (n)	237	193	44	138	99
#Degree of freedom	236	192	43	137	98
Poisson Regression Model (PRM)					
#Parameters (k); $\frac{n}{k}$	4; 59	4; 48	4; 10.75	4; 34.25	4; 24.50
Coefficients*					
(Intercept)	**0.000** [2.830, 3.311]	**0.000** [2.274, 3.283]	**0.000** [2.408, 4.741]	**0.000** [2.802, 3.558]	**0.000** [2.970, 4.070]
Age	0.164 [-0.001, 0.005]	**0.012** [0,001, 0.009]	0.245 [-0.027, 0.006]	0.734 [-0.008, 0.005]	0.188 [-0.014, 0.002]
Gender	0.533 [-0.053, 0.102]	0.636 [-0.106, 0.065]	**0.021** [0.031, 0.385]	0.658 [-0.127, 0.081]	0.125 [-0.025, 0.203]
Period	**0.000** [-0.424, -0.287]	**0.000** [-0.417, -0.270]	**0.000** [-0.563, -0.175]	**0.000** [-0.422, -0.239]	**0.000** [-0.499, -0.289]
#F. S. Iter.†	4	4	4	4	4
AIC	1572.2	1280	-	-	-
AIC_c	-	-	289.108	864.395	707.401
Deviance residuals**	0.025 [2.136, 10.300]	**0.046 [2.126, 9.985]**	**0.017 [2.159, 7.174]**	0.157 [2.058, 6.998]	-0.039 [2.070, 10.545]
MSE◇	0.300	0.284	0.338	0.279	0.317
RMSE◇	0.548	0.533	0.581	0.528	0.563
Max. likelihood value	9.92% at $\Delta t_{\text{LOS}} = 11$	**9.8% at $\Delta t_{\text{LOS}} = 10$**	**9.8% at $\Delta t_{\text{LOS}} = 10$**	10.5% at $\Delta t_{\text{LOS}} = 10$	9.4% at $\Delta t_{\text{LOS}} = 11$
Negative Binomial Regression Model (NBM)					
#Parameters (k); $\frac{n}{k}$	5; 47.20	5; 38.40	5; 8.60	5; 27.40	5; 19.60
Coefficients*					
(Intercept)	**0.000** [2.670, 3.470]	**0.000** [2.590, 3.474]	**0.000** [1.903, 5.320]	**0.000** [2.632, 3.723]	**0.000** [2.605, 4.500]
Age	0.436 [-0.002, 0.006]	0.151 [-0.002, 0.010]	0.373 [-0.035, 0.013]	0.802 [-0.011, 0.008]	0.392 [-0.020, 0.008]
Gender	0.798 [-0.105, 0.137]	0.685 [-0.162, 0.106]	0.146 [-0.069, 0.466]	0.741 [-0.178, 0.126]	0.428 [-0.116, 0.276]
Period	**0.000** [-0.459, -0.243]	**0.000** [-0.457, -0.225]	**0.018** [-0.636, -0.066]	**0.000** [-0.456, -0.194]	**0.000** [-0.572, -0.211]
#F. S. Iter.†	1	1	1	1	1
AIC	1448.4	-	-	-	-
AIC_c	-	1180.284	275.193	821.457	632.502
Deviance residuals**	0.021 [1.396, 6.779]	**0.035** [1.347, 6.583]	**-0.011** [1.437, 5.038]	0.106 [1.483, 4.879]	-0.039 [1.224, 6.514]
MSE◇	0.300	0.284	0.338	0.279	0.563
RMSE◇	0.548	0.533	0.581	0.528	0.563
Max. likelihood value	7.6% at $\Delta t_{\text{LOS}} = 9$	**7.5% at $\Delta t_{\text{LOS}} = 9$**	**7.8% at $\Delta t_{\text{LOS}} = 9$**	8.3% at $\Delta t_{\text{LOS}} = 9$	6.9% at $\Delta t_{\text{LOS}} = 9$

* *p-value [95% confidence interval (CI)]* ** *median [IQR, range]* † *Fisher Scoring Iterations* ◇ *Error with respect to observed data*

Table 3: Results of the generalized linear models using Poisson distribution and negative binomial distribution of recovery-period Δt_{LOS} using $\mathbb{G}_{\textbf{+cfrmDt}}$ grouping.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069724; this version posted May 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Table 3 gives the results of our models. We infer the following:

- The "period" variable is significant for both models and all scenarios, whereas the "age" and "gender" variables are significant only in the PRM for one scenario each, when considering a p-value for the coefficients corresponding to the variable being less than 5%.
- The NBM shows lower values for the median and the variance (observable from range and IQR) of deviance residuals, and AIC/AIC$_c$ than the PRM. Thus, we conclude that NBM is a better fit than PRM.
- For both PRM and NBM, the piecewise models in scenario M_2 are a better fit than the single model, M_1, and the piecewise models of M_3, owing to the lower error measures (AIC/AIC$_c$ and deviance residuals) as well as the consistency in the maximum likelihood value.

Choosing NBM and M_2, we get that the maximum likely value of Δt_{LOS} to be 9 days. These observations may be influenced by the relatively small sample size for P_1 and P_3. Since we do not have a large number of variables to discard, we retain the "gender" and "age" variables in the model despite its relative insignificance, compared to "period."

The key conclusion from the multivariate regression analysis is that a GLM with NBM for M_2 is the best model to estimate Δt_{LOS}. This helps us to estimate the value of Δt_{LOS} to be 9 days, with the maximum likelihood of $\sim$7.5-9.8%, from both PRM and NBM. Hence, overall, we conclude that Δt_{LOS} can be estimated to have decreased to 9 days during January 23-April 01, 2020, in Singapore.

4 Discussion

The reduced Δt_{LOS}, as per our estimate, is an outcome of the implementation of government-mandated protocols in Singapore for containment of the contagion. The containment strategies have ensured delays in the spread of the disease. A side benefit of the containment of the spread is the shift in the age-group of the population getting infected. The infection in the vulnerable groups, such as the children and the older population, has reduced with time. This temporal shift in the demographic structure has led to the overall decrease in Δt_{LOS}, e.g., we observe that "age" is a significant variable for the multivariate regression model with PRM for the population of the age-group (20-59) years in M_2.

Our work has the following limitation that we have modeled recovery isolated from the infection. Since the number of infected persons in Singapore has increased exponentially since March 17, 2020, a change in the protocol followed in the hospitals is expected for treating positive confirmed patients. Nevertheless, our estimated Δt_{LOS} is still applicable for assessing the hospital load in terms of required resources for in-hospital stay.

In summary, we have considered the demographic data and timeline of the first 1000 COVID-19 patients in Singapore during January 23-April 01, 2020. We have closely analyzed the data on the positive confirmation and discharge/clinical recovery dates of 245 patients who recovered during this time period. We have used regression analysis, subsequent to a descriptive statistical analysis, to get an estimate of recovery-period Δt_{LOS} (i.e., hospital length of stay (LOS)). We have found that the Δt_{LOS} is time-varying, after performing periodization to find three significant periods based on the transmission model, namely P_1 (January 23-February 03), P_2 (February 04-March 16), and P_3 (March 17-April 01). The median value of Δt_{LOS} varied from $\sim$17 days in P_1 to $\sim$10 days in P_2 to $\sim$9 days in P_3. We have used the loess model for time-series data to demonstrate the negative slope of the regression curve of Δt_{LOS} in P_3, in particular. We then estimated age-wise Δt_{LOS} using generalized linear models for multivariate (linear) regression with Poisson and negative binomial distributions for count data, using age, gender, and period as independent variables. Our estimated value of $\Delta t_{\text{LOS}} = 9$ days shows an improvement in the values published for Δt_{LOS}, i.e., 20 days [10] and 15 days [9].

Acknowledgements

The authors would like to thank IIIT Bangalore, and particularly, the Graphics-Visualization-Computing Lab and the E-Health Research Center, for supporting this work.

References

[1] Rachael Pung, Calvin J. Chiew, Barnaby E. Young, Sarah Chin, Mark I-C Chen, Hannah E. Clapham, Alex R. Cook, Sebastian Maurer-Stroh, Matthias P. H. S. Toh, Cuiqin Poh, Mabel Low, Joshua Lum, Valerie T. J. Koh, Tze M. Mak, Lin Cui, Raymond V. T. P. Lin, Derrick Heng, Yee-Sin Leo, David C. Lye, and Vernon J. M. Lee. Investigation of three clusters of COVID-19 in Singapore: implications for surveillance and response measures. *The Lancet*, 395:1039–46, 2020.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069724; this version posted May 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

[2] John E. L. Wong, Yee Sin Leo, and Chorh Chuan Tan. COVID-19 in Singapore—Current Experience: Critical Global Issues That Require Attention and Action. *JAMA*, 02 2020.

[3] Yixiang Ng, Zongbin Li, Yi Xian Chua, Wei Liang Chaw, Zheng Zhao, Benjamin Er, Rachael Pung, Calvin J. Chiew, David C. Lye, Derrick Heng, and Vernon J. Lee. Evaluation of the Effectiveness of Surveillance and Containment Measures for the First 100 Patients with COVID-19 in Singapore — January 2–February 29, 2020. *Morbidity and Mortality Weekly Reports (MMWR)*, 69(11):307–311, 2020.

[4] Wycliffe E. Wei, Zongbin Li, Calvin J. Chiew Chiew, Sarah E. Yong, Matthias P. Toh, and Vernon J. Lee Lee. EPresymptomatic Transmission of SARS-CoV-2 — Singapore, January 23–March 16, 2020. *Morbidity and Mortality Weekly Reports (MMWR)*, 69(14):411–415, 2020.

[5] Evelene M Carter and Henry WW Potts. Predicting length of stay from an electronic patient record system: a primary total knee replacement example. *BMC medical informatics and decision making*, 14(1):26, 2014.

[6] Nanshan Chen, Min Zhou, Xuan Dong, Jieming Qu, Fengyun Gong, Yang Han, Yang Qiu, Jingli Wang, Ying Liu, Yuan Wei, and et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The Lancet*, 395(10223):507–513, 2020.

[7] Adam J Kucharski, Timothy W. Russell, Charlie Diamond, Yang Liu, John Edmunds, Sebastian Funk, and Rosalind M. Eggo. Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The Lancet Infectious Diseases*, 2020.

[8] Qun Li, Xuhua Guan, Peng Wu, Xiaoye Wang, Lei Zhou, Yeqing Tong, Ruiqi Ren, Kathy S.M. Leung, Eric H.Y. Lau, Jessica Y. Wong, Xuesen Xing, Nijuan Xiang, Yang Wu, Chao Li, Qi Chen, Dan Li, Tian Liu, Jing Zhao, Man Liu, Wenxiao Tu, Chuding Chen, Lianmei Jin, Rui Yang, Qi Wang, Suhua Zhou, Rui Wang, Hui Liu, Yinbo Luo, Yuan Liu, Ge Shao, Huan Li, Zhongfa Tao, Yang Yang, Zhiqiang Deng, Boxi Liu, Zhitao Ma, Yanping Zhang, Guoqing Shi, Tommy T.Y. Lam, Joseph T. Wu, George F. Gao, Benjamin J. Cowling, Bo Yang, Gabriel M. Leung, and Zijian Feng. Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia. *New England Journal of Medicine*, 382(13):1199–1207, 2020.

[9] Cleo Anastassopoulou, Lucia Russo, Athanasios Tsakris, and Constantinos Siettos. Data-based analysis, modelling and forecasting of the COVID-19 outbreak. *PloS one*, 15(3):e0230405, 2020.

[10] Fei Zhou, Ting Yu, Ronghui Du, Guohui Fan, Ying Liu, Zhibo Liu, Jie Xiang, Yeming Wang, Bin Song, Xiaoying Gu, Lulu Guan, Yuan Wei, Hui Li, Xudong Wu, Jiuyang Xu, Shengjin Tu, Yi Zhang, Hua Chen, and Bin Cao. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *The Lancet*, 395(10229):1054–62, 2020.

[11] Xiaobo Yang, Yuan Yu, Jiqian Xu, Huaqing Shu, Jia'an Xia, Hong Liu, Yongran Wu, Lu Zhang, Zhui Yu, Minghao Fang, Ting Yu, Yaxin Wang, Shangwen Pan, Xiaojing Zou, Shiying Yuan, and You Shang. Clinical course and outcomes of critically ill patients with SARS-CoV-2 pneumonia in Wuhan, China: a single-centered, retrospective, observational study. *The Lancet Respiratory Medicine*, 2020.

[12] William S Cleveland and Susan J Devlin. Locally weighted regression: an approach to regression analysis by local fitting. *Journal of the American statistical association*, 83(403):596–610, 1988.

[13] William G Jacoby. Loess:: a nonparametric, graphical tool for depicting relationships between variables. *Electoral Studies*, 19(4):577–613, 2000.

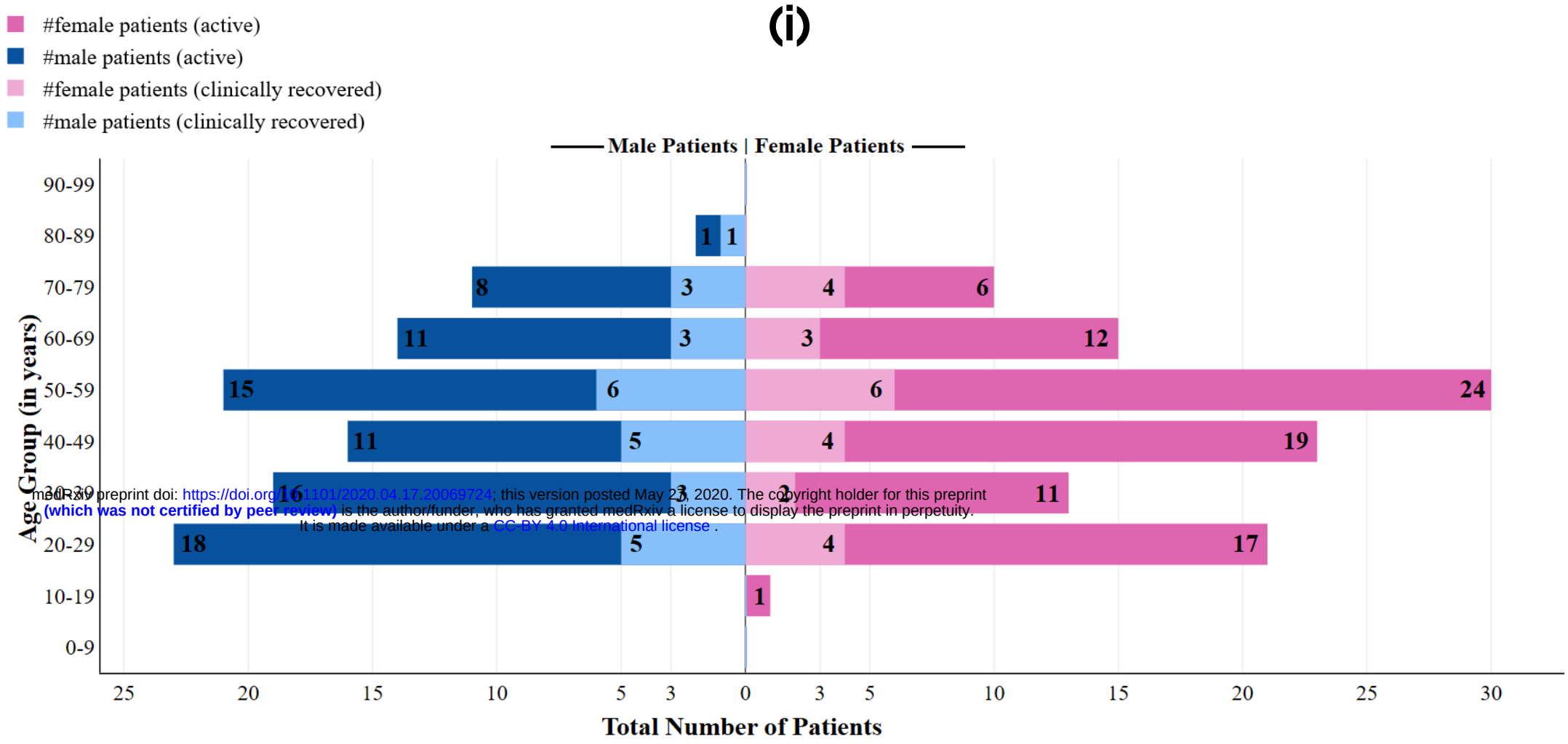
[14] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2013. ISBN 3-900051-07-0.

[15] Liming Xiang, Andy H Lee, Kelvin KW Yau, and Geoffrey J McLachlan. A score test for overdispersion in zero-inflated poisson mixed regression model. *Statistics in medicine*, 26(7):1608–1622, 2007.

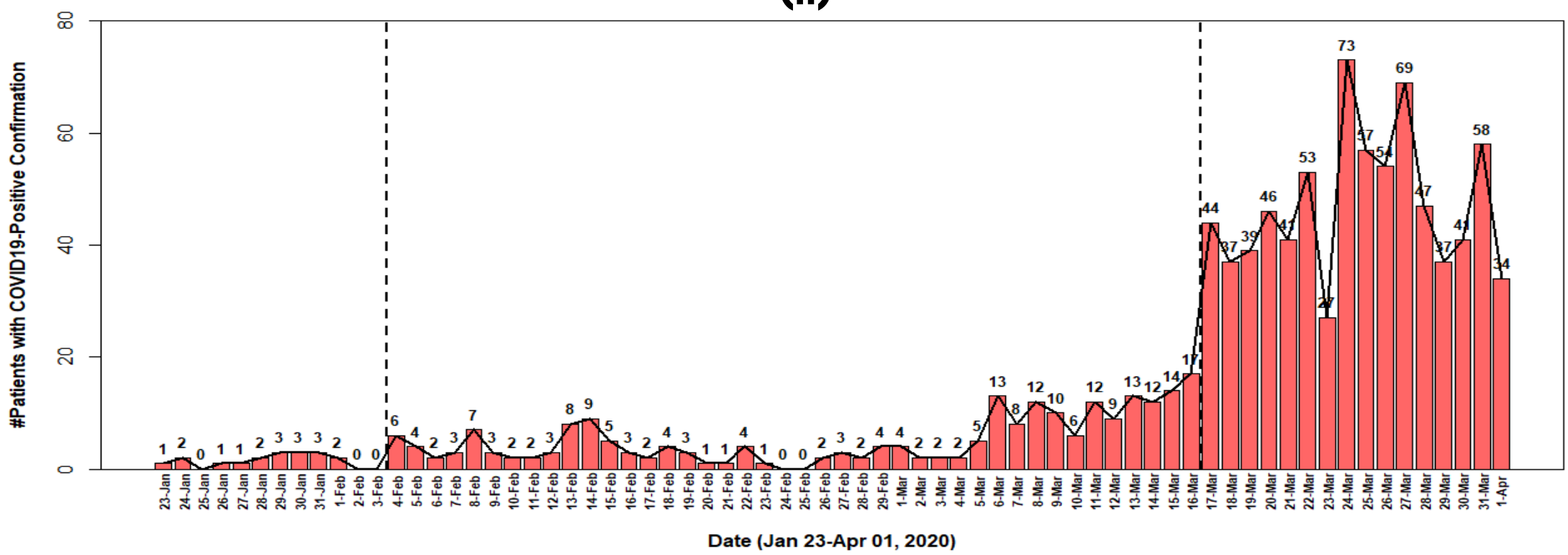
[16] K. P. Burnham and D. R. Anderson. *Model selection and inference: a practical information-theoretic approach*. Springer Verlag, 1998.

medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069724; this version posted May 23, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-4.0 International license .

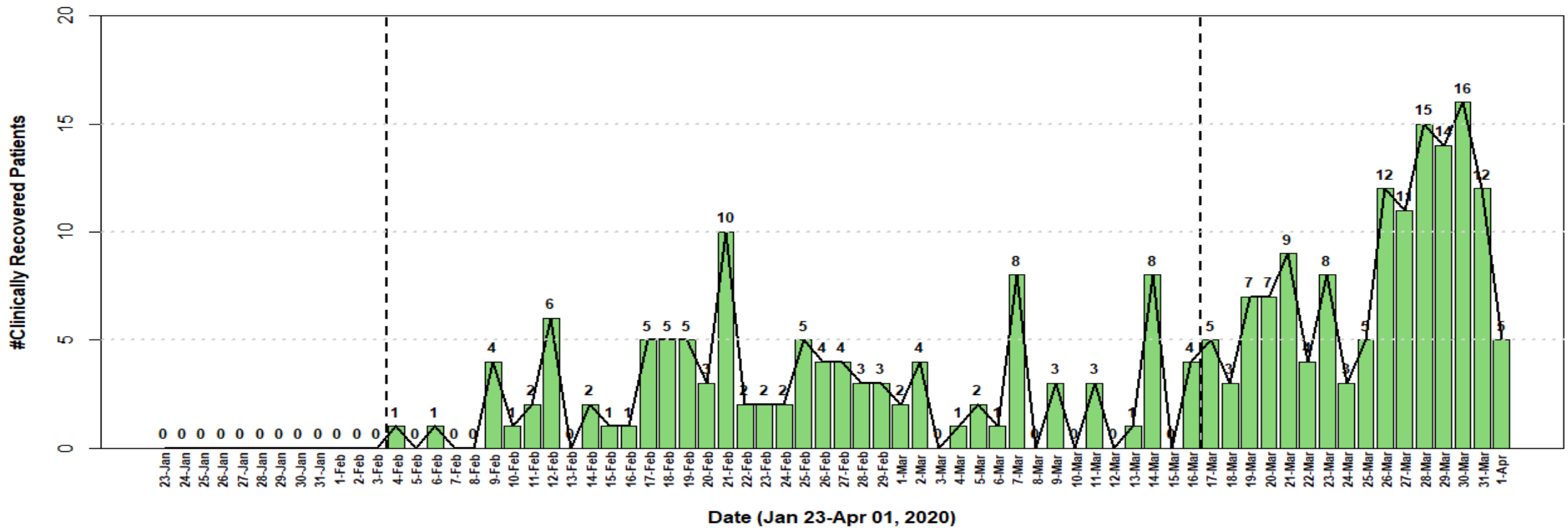
(i)



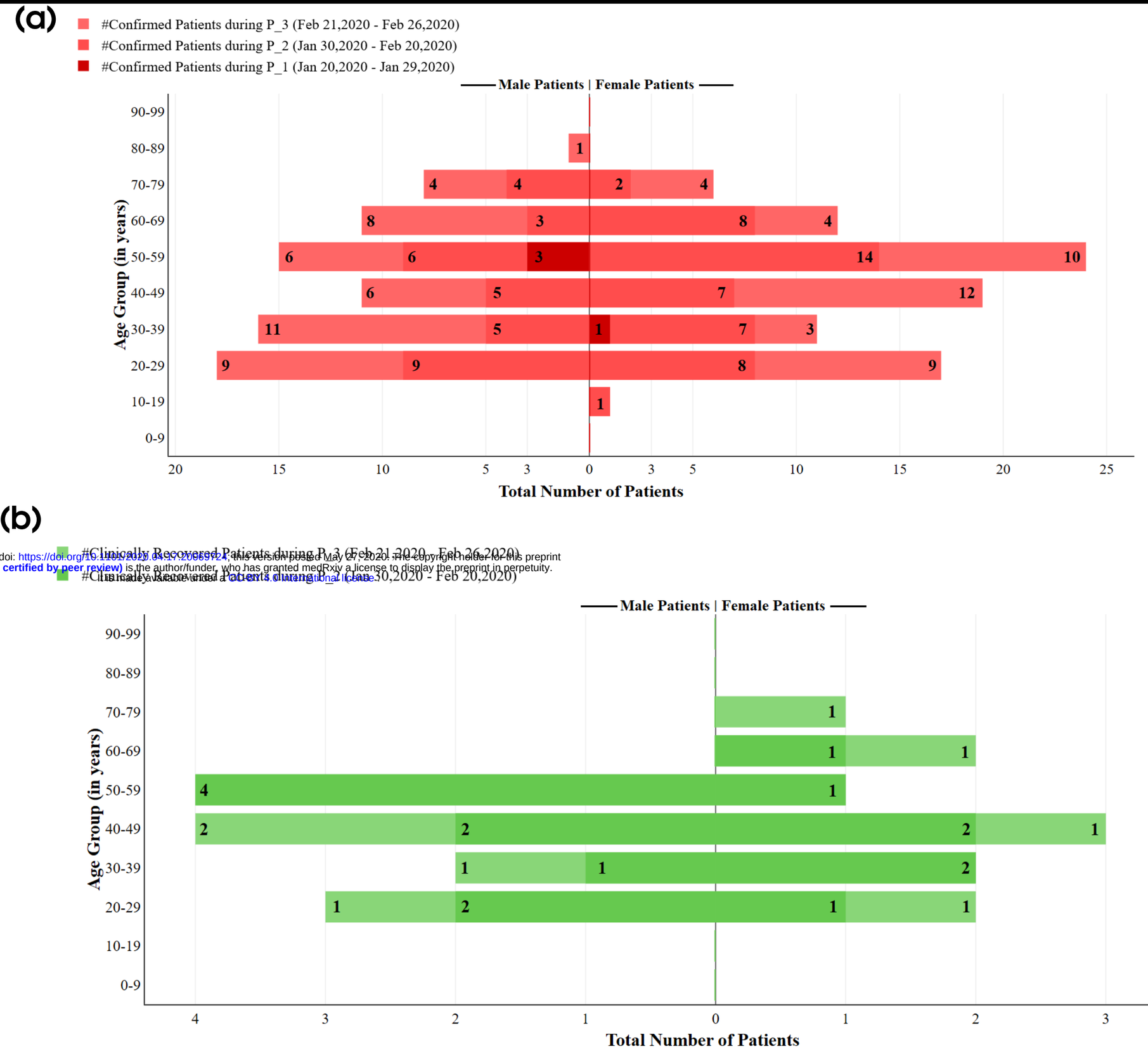
(ii)



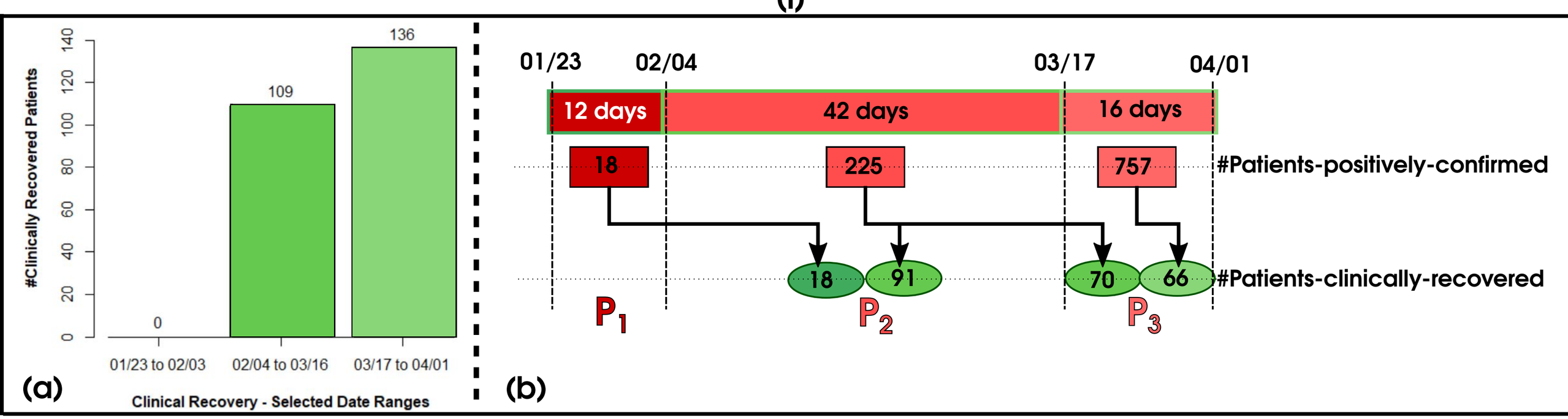
(iii)



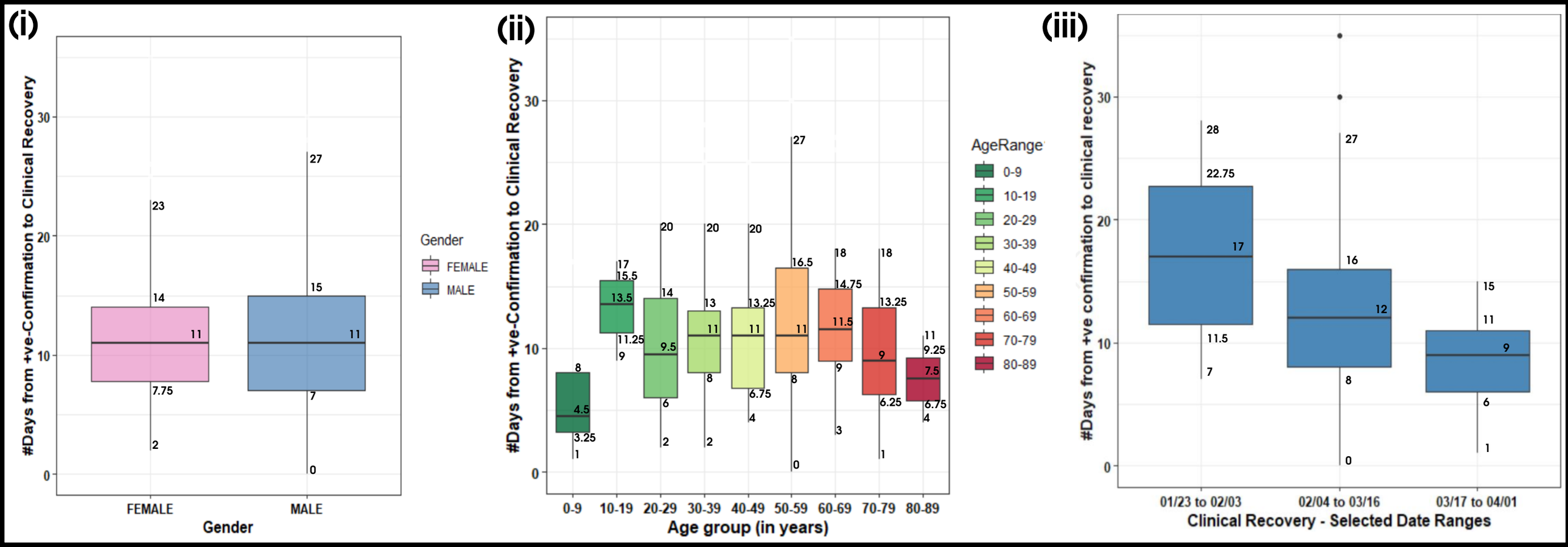
medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20065924, this version posted May 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-NC-ND 4.0 International license.

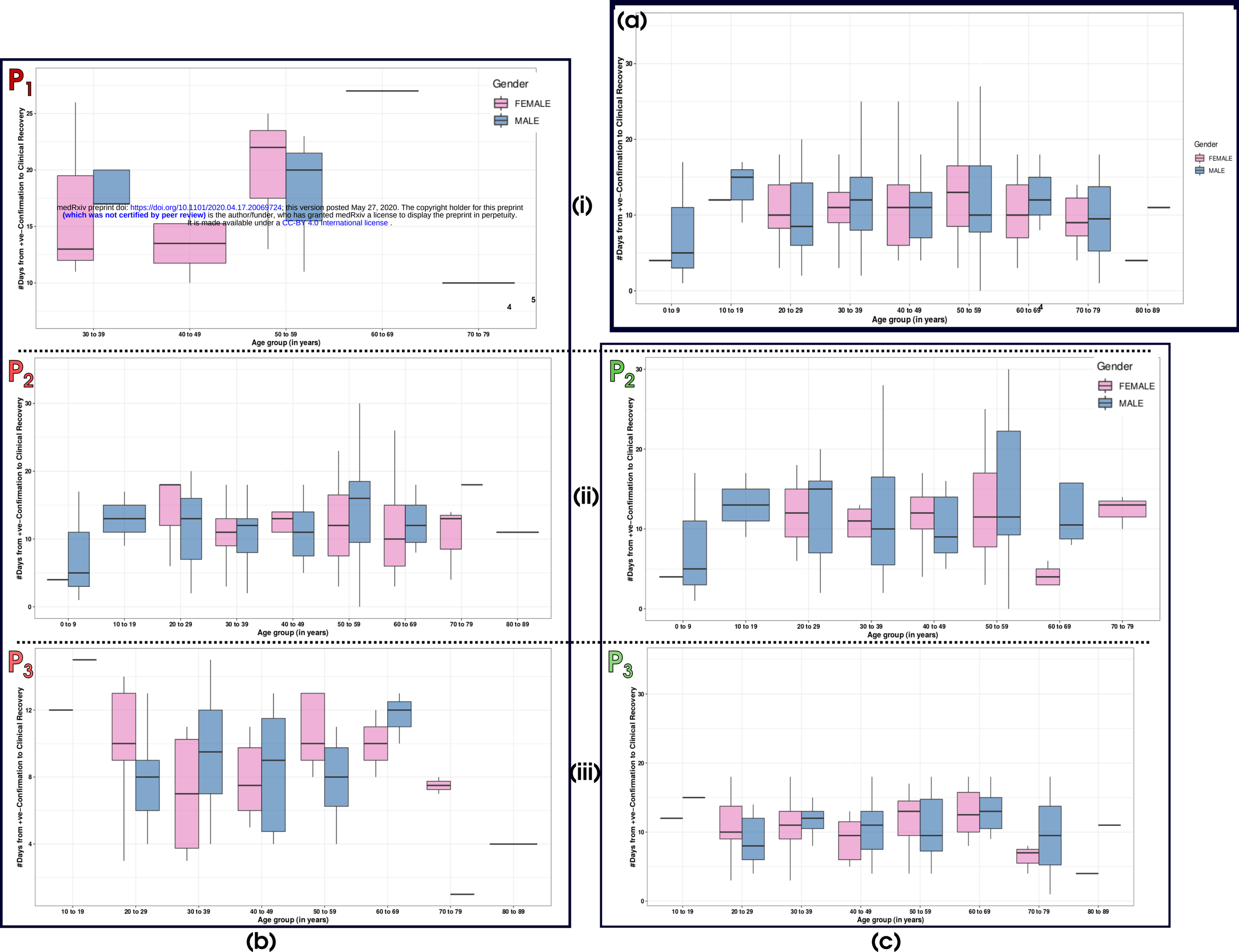


(i)

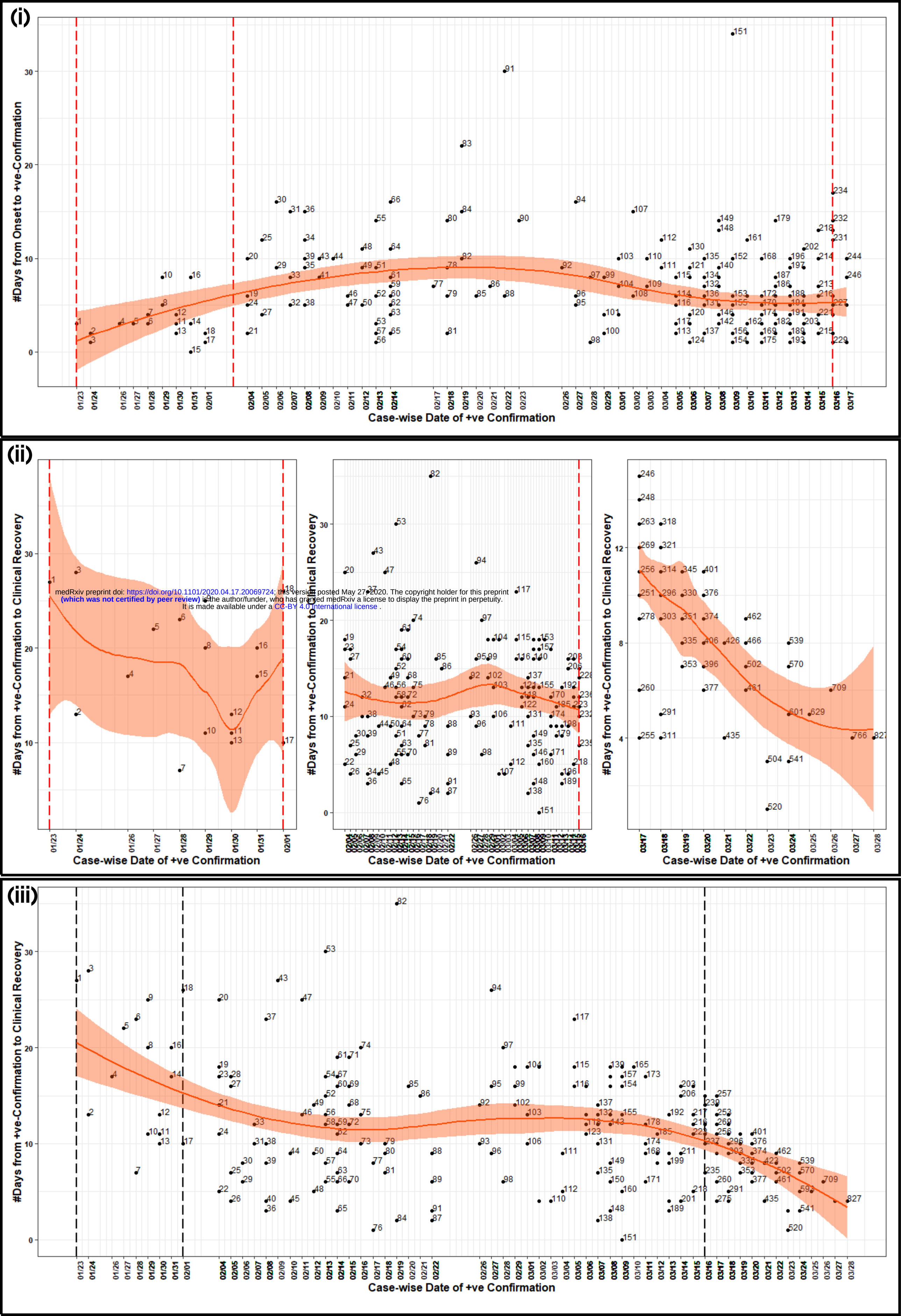


(ii)

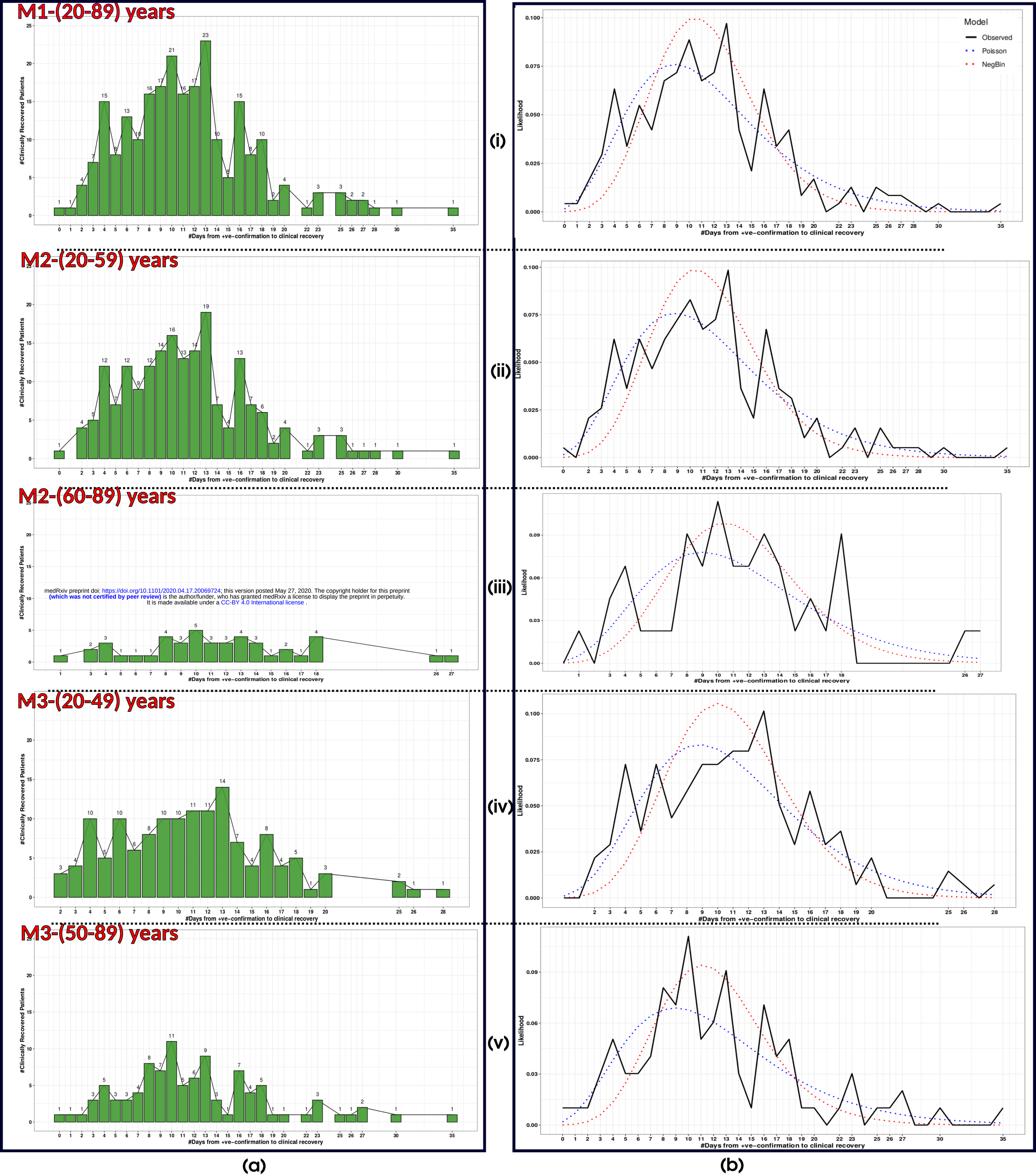

(i)
#Days from +ve-Confirmation to Clinical Recovery
Gender
FEMALE
MALE
(ii)
#Days from +ve-Confirmation to Clinical Recovery
Age group (in years)
0-9
10-19
20-29
30-39
40-49
50-59
60-69
70-79
80-89
AgeRange
(iii)
#Days from +ve confirmation to clinical recovery
01/23 to 02/03
02/04 to 03/16
03/17 to 04/01
Clinical Recovery - Selected Date Ranges



medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069724; this version posted May 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .



medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069724; this version posted May 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .



medRxiv preprint doi: https://doi.org/10.1101/2020.04.17.20069724; this version posted May 27, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .