



Article

# Estimation of the Actual Incidence of Coronavirus Disease (COVID-19) in Emergent Hotspots: The Example of Hokkaido, Japan during February–March 2020

Andrei R. Akhmetzhanov <sup>1,2</sup> , Kenji Mizumoto <sup>3,4</sup> , Sung-Mok Jung <sup>1,5</sup> , Natalie M. Linton <sup>1,5</sup> , Ryosuke Omori <sup>6</sup> and Hiroshi Nishiura <sup>1,5,7,\*</sup>

- <sup>1</sup> Graduate School of Medicine, Hokkaido University, Kita 15 Jo Nishi 7 Chome, Kita-ku, Sapporo-shi, Hokkaido 060-8638, Japan; akhmetzhanov@gmail.com (A.R.A.); seductmd@gmail.com (S.-M.J.); nlinton@gmail.com (N.M.L.)
  - <sup>2</sup> Global Health Program, Institute of Epidemiology and Preventive Medicine, College of Public Health, National Taiwan University, 17 Xu-Zhou Road, Taipei 10055, Taiwan
  - <sup>3</sup> Graduate School of Advanced Integrated Studies in Human Survivability, Kyoto University, Yoshida-Nakaadachi-cho, Sakyo-ku, Kyoto 606-8501, Japan; mizumotokenji@gmail.com
  - <sup>4</sup> Hakubi Center for Advanced Research, Kyoto University, Yoshidahonmachi, Sakyo-ku, Kyoto 606-8306, Japan
  - <sup>5</sup> School of Public Health, Kyoto University, Yoshidakonoe cho, Sakyo ku, Kyoto 606-8501, Japan
  - <sup>6</sup> Research Center for Zoonosis Control, Hokkaido University, Kita 19 Jo Nishi 10 Chome, Kita-ku, Sapporo-shi, Hokkaido 001-0019, Japan; r.omori12@gmail.com
  - <sup>7</sup> Core Research for Evolutional Science and Technology (CREST), Japan Science and Technology Agency, Honcho 4-1-8, Kawaguchi, Saitama 332-0012, Japan
- \* Correspondence: nishiura.hiroshi.5r@kyoto-u.ac.jp; Tel.: +81-75-753-4490



**Citation:** Akhmetzhanov, A.R.; Mizumoto, K.; Jung, S.-M.; Linton, N.M.; Omori, R.; Nishiura, H. Estimation of the Actual Incidence of Coronavirus Disease (COVID-19) in Emergent Hotspots: The Example of Hokkaido, Japan during February–March 2020. *J. Clin. Med.* **2021**, *10*, 2392. <https://doi.org/10.3390/jcm10112392>

Academic Editor: Emmanuel Andr es

Received: 16 March 2021

Accepted: 24 May 2021

Published: 28 May 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Abstract:** Following the first report of the coronavirus disease 2019 (COVID-19) in Sapporo city, Hokkaido Prefecture, Japan, on 14 February 2020, a surge of cases was observed in Hokkaido during February and March. As of 6 March, 90 cases were diagnosed in Hokkaido. Unfortunately, many infected persons may not have been recognized due to having mild or no symptoms during the initial months of the outbreak. We therefore aimed to predict the actual number of COVID-19 cases in (i) Hokkaido Prefecture and (ii) Sapporo city using data on cases diagnosed outside these areas. Two statistical frameworks involving a balance equation and an extrapolated linear regression model with a negative binomial link were used for deriving both estimates, respectively. The estimated cumulative incidence in Hokkaido as of 27 February was 2,297 cases (95% confidence interval (CI): 382–7091) based on data on travelers outbound from Hokkaido. The cumulative incidence in Sapporo city as of 28 February was estimated at 2233 cases (95% CI: 0–4893) based on the count of confirmed cases within Hokkaido. Both approaches resulted in similar estimates, indicating a higher incidence of infections in Hokkaido than were detected by the surveillance system. This quantification of the gap between detected and estimated cases helped to inform the public health response at the beginning of the pandemic and provided insight into the possible scope of undetected transmission for future assessments.

**Keywords:** epidemiology; travel medicine; COVID-19; emerging infectious diseases



**Copyright:**   2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

In December 2019, a cluster of 41 patients with atypical pneumonia of unknown etiology was reported in the city of Wuhan, China [1,2]. The number of atypical pneumonia cases in Wuhan rapidly increased in early January and cases began appearing across China and in other countries [3,4]. The cause of the pneumonia was recognized as a newly emerged coronavirus called severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), and the disease it causes was given the name coronavirus disease 2019 (COVID-19). Cases of COVID-19 have now been reported in more than 120 countries worldwide [5],

and the WHO declared it a pandemic on 11 March 2020. By the end of 2020, the disease had spread to nearly every country in the world. In the present paper, we examine the emergence of a COVID-19 hotspot in Hokkaido Prefecture, Japan, during the first wave of the pandemic. Because of the limited testing capacity in the first months of the outbreak [6] and the specific contact-based, centered surveillance of SARS-CoV-2 infections in Japan [7], a substantial fraction of infected individuals may have remained undetected during the first months of the pandemic.

Control of COVID-19 spread is complicated by various factors, including nonspecific clinical symptoms, especially during the early phase of infection [8–10]; a relatively high proportion of pre-symptomatic and asymptomatic infections [11–15]; multiple possible routes of transmission—including aerosol, direct contact, and fecal–oral transmission [10,16,17]—as well as the moderate to high potential of superspreading events generated by a small proportion of cases [18,19]. Additionally, the incubation period can vary widely [20,21], and re-occurrence of the disease after primary infection is possible [22]. Nonetheless, the transmission dynamics of COVID-19 in the community can be roughly grasped by exploring epidemiological indicators such as the number of confirmed infections in a given location, or the number of exported cases.

Clinical manifestation of COVID-19 begins with nonspecific symptoms (similar to seasonal influenza), which remain mild for most persons. Some infections may remain entirely asymptomatic. For a small fraction of symptomatic cases, the disease may resolve into a life-threatening onset of acute respiratory distress [19,23]. Despite efforts towards mitigation, the number of confirmed cases worldwide represents only a fraction of all infections, and the actual number of cases including mild and asymptomatic infections may be as large as four- or ten-fold [15,24,25]. Hence, the derivation of reliable estimates of the actual number of cases is needed to conduct risk assessments of the spread of COVID-19 and inform policies related to containment and mitigation [26]. This was especially important for newly emerging hotspots in Japan and some other countries at the beginning of 2020 [4,27].

The increased number of cases in and exportation of cases from Hokkaido in February–March 2020 signaled an urgent need to increase containment efforts within the prefecture. To help guide the policymaking process and better understand the risk of further spread within and exportation from Hokkaido, we derived real-time estimates of the actual incidence by accounting for under-ascertained cases. The methods can be further expended to other areas or countries experiencing a surge of new cases of COVID-19, any other emergent disease, or their new variants.

## 2. Methods

### 2.1. Epidemiological Data

Three sources of data were leveraged to estimate the actual incidence of COVID-19 in Hokkaido. Firstly, cases exported domestically and internationally from Hokkaido were identified based on government reports. The domestically exported cases were Kumamoto (1 case), Nagano (1 case), and Chiba (1 case) prefecture residents [28–30]. The internationally exported cases were reported from Malaysia (1 case) and Thailand (2 cases) [31,32]. Considering that the two Thailand cases were husband and wife and one case was asymptomatic, one infection may have been the result of household transmission, as similarly described in [33], rather than exported from Japan. Secondly, the total volume of air passengers from all airports in Hokkaido to Thailand and Malaysia over the two-month period of January and February 2019 was retrieved from the International Air Transport Association (IATA) and included 9349 passengers flown directly to Malaysia and 45,137 to Thailand. Lastly, the population size of Hokkaido was obtained from the Hokkaido Prefectural Government website [34].

The calculation of COVID-19 incidence in Sapporo was based on cases of non-Sapporo residents who (i) became infected in Sapporo and were diagnosed outside of Sapporo or (ii) cases diagnosed outside of Sapporo who had no information on their source of infection

(i.e., unknown link). Cases who were infected outside of Sapporo by Sapporo residents were excluded from the analysis, as they did not represent true exportation of infection from Sapporo by a local case. For example, in one instance, a Sapporo resident travelled to Kitami city in Okhotsk subprefecture and caused secondary infections among Kitami residents. These Kitami residents were excluded from our analysis, as they were infected outside of Sapporo. However, if a case from Kitami were to travel to Sapporo, have likely been infected there, and then be subsequently diagnosed in Kitami, they would be eligible for our analysis. Data on cases reported through the end of February 2020, including dates of illness onset, subprefecture of diagnosis, history of travel to Sapporo, and epidemiologic linkage to other confirmed cases, were retrieved from the websites of government entities in Hokkaido.

The estimate of the fraction of commuters between subprefectures was obtained with reference to a survey of the daytime and nighttime population sizes in Sapporo [35]. By subtracting the daytime population size from the nighttime population size and dividing the difference by the nighttime population, we obtained a value of 3.7% [35], reflecting the change in the population size of Sapporo due to commuting. However, this estimate accounts for the movement of individuals both within Ishikari subprefecture (including Sapporo and several other cities) and between subprefectures. In contrast, our model requires the average estimate between all subprefectures of Hokkaido. Therefore, we set this parameter at 1% and performed a sensitivity analysis wherein we varied the fraction of commuters between 0.5% and 4%.

## 2.2. Modeling Commuter Movement between Subprefectures of Hokkaido

Two single-parameter models of human movement were employed to predict the rate of commuting between Sapporo and the various subprefectures of Hokkaido—a radiation model and a model of uniform selection. Both emphasize the attractiveness of large population centers [36–39]. The uniform selection model [40] predicts the commuting volume using the following formula:

$$M_{ij} = M_i \frac{P_i}{N - Q_j}, \quad (1)$$

where  $P_i$  is the population size at the origin  $i$ ,  $Q_j$  is the population size at the destination  $j$ , and  $N$  is the total population of Hokkaido. The only parameter to be estimated is  $M_i$ , which defines the scaling of the ongoing number of commuters from the source subprefecture and is equal to the proportion of commuters multiplied on the population size  $P_i$  [39]. In contrast, the radiation model [41] incorporates the distance metric between the origin and the destination and defines a commuting volume with the following formula:

$$M_{ij} = M_i \frac{P_i Q_j}{(P_i + R_{ij})(P_i + Q_j + R_{ij})} \quad (2)$$

where  $R_{ij}$  denotes the total population within a radius around two centers,  $P_i$  and  $Q_j$ . Both matrices  $M_{ij}$  (1)–(2) were made symmetric in our simulations. Either model (1) or (2) was selected based on the values of a widely applicable information criterium (WAIC), with a lower value regarded as better.

An aggregated dataset of population sizes and central point coordinates for each subprefecture was used to estimate the commuting rates using the R package “movement” [38]. The package required a fraction of the commuting individuals to be a single input parameter for the function. Due to the uncertainty of this value, we performed a sensitivity analysis with reference to the daytime and nighttime populations of Sapporo [35].

## 2.3. Estimating Actual Incidence of COVID-19 Cases in Hokkaido Prefecture and Sapporo City

The actual incidence was estimated in (i) Hokkaido Prefecture using the number of detected cases among international and domestic travelers outbound from Hokkaido, and

in (ii) Sapporo city using the number of confirmed cases within different subprefectures of Hokkaido. Both estimates were thought to be close to each other, because Sapporo was the epicenter for COVID-19 spread within Hokkaido.

The framework was based on two key assumptions. First, the outbreak was considered to be in its initial stage during the time period of February to March 2020, when the first cases in Sapporo emerged and had just expanded to other areas in Hokkaido. In such an instance, the disentanglement of locally acquired infections from the imported cases was still possible in each subprefecture. Second, the travel volume between subprefectures in Hokkaido was assumed to be well-approximated by the models of commuting movement, which were described above.

- (i) Estimating incidence in Hokkaido by using imported cases among travelers outbound from Hokkaido

A balance equation was implemented to predict incidence across all Hokkaido using the case count of air passengers from Hokkaido with international or domestic destinations [3,42–44]. Given the observed cumulative count of exported cases  $C$ , we estimated the fraction of infected individuals in Hokkaido  $\hat{p}$  by the following equation:

$$\hat{p} = C \frac{365}{mT}, \quad (3)$$

where  $m$  is the total volume of passengers to the corresponding destination and  $T$  is the infectious period, approximated from the observed average virus shedding period of 5.0 days [45]. Accordingly, the incidence in Hokkaido was defined by  $\hat{p}n$ , where  $n$  is the catchment population size for the international airport in Hokkaido. To account for stochasticity, we used a binomial sampling process, maximum likelihood estimation, and 95% confidence intervals derived from the profile likelihood.

- (ii) Estimating incidence in Sapporo using cases imported within Hokkaido

Following a previously developed framework [46], the number of infections  $C_j$  in subprefecture  $j$  of Hokkaido Prefecture was sampled from a negative binomial distribution with the mean  $\lambda_j$  linearly proportional to the predicted travel volume between Sapporo and subprefecture  $j$ :  $\lambda_j = \beta \cdot M_{oj}$ , and the dispersion parameter  $k$ :

$$C_j \sim \text{Negative Binomial}(\text{mean} = \lambda_j, \text{overdispersion} = k), \quad (4)$$

$$\lambda_j = \beta \cdot M_{oj}, \quad j = 1 \dots 14,$$

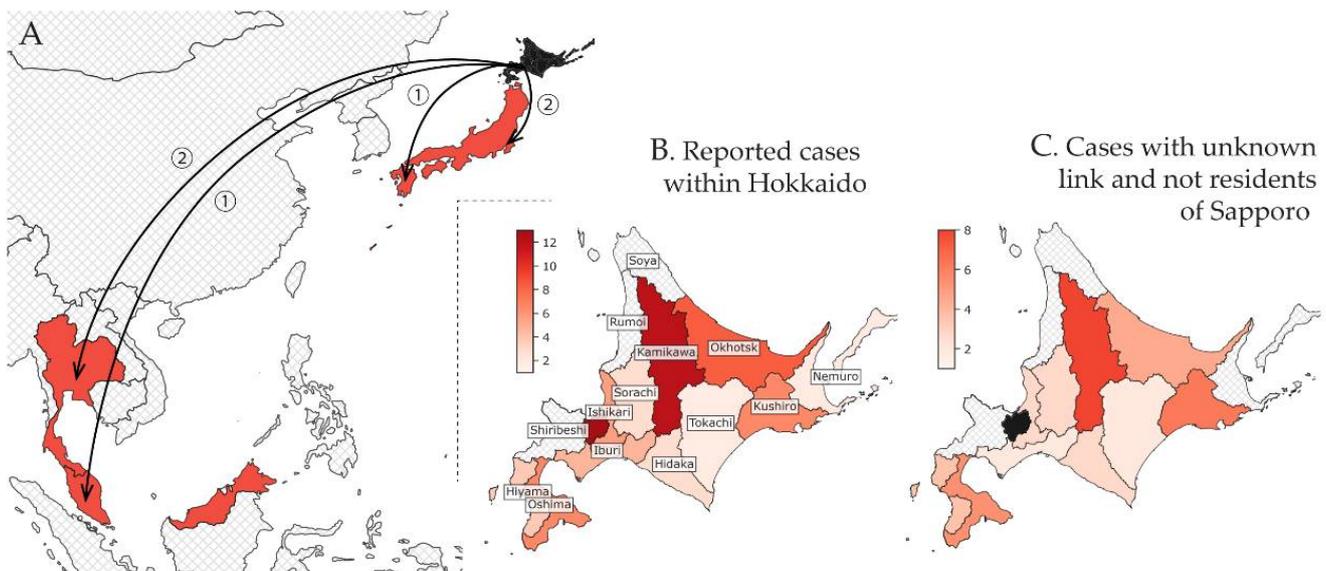
where the regression coefficient  $\beta$  is considered to be independent of a subprefecture and to be fitted to the data.

Because Sapporo is a part of Ishikari subprefecture, for our study, we defined Ishikari subprefecture as all cities within Ishikari subprefecture, excluding Sapporo (Figure 1B). The flow  $M_{oj}$  was predicted using the uniform selection and radiation models of human movement [40,41], as described in Equations (1) and (2). The cumulative incidence in Sapporo was then estimated by extending Equation (4) to the entire population of Sapporo:

$$C_o \sim \text{NegativeBinomial}(\beta \cdot N_o, k), \quad (5)$$

where  $N_o$  is the population size of Sapporo.

The derived estimates of the linear regression model (3) were also used to validate our assumption of linearity between the number of reported COVID-19 infections in each subprefecture and the daily commuting rate [47]. Using the fit of the predicted travel volume and observed case counts, the best-fit model between (1) and (2) was selected based on the difference in the WAIC values.



**Figure 1.** Geographical distribution of confirmed cases linked to Hokkaido by subprefecture of diagnosis as of 29 February 2020. (A) International and domestic cases (excluding Hokkaido). Hokkaido is shaded in black, whereas affected countries are depicted in red. The circled numbers indicate the total count of cases for each import destination. (B) All cases confirmed in Hokkaido. The labels indicate the names of the subprefectures of Hokkaido. (C) Only cases included in our analysis: non-residents of Sapporo with unknown links either with or without history of travel to Sapporo. Ishikari subprefecture was separated into two subregions: Sapporo city and outside Sapporo city. Hatched areas in grey indicate subprefectures with zero counts.

#### 2.4. Reporting Delay between Illness Onset and Confirmation

Because the dataset for this study was collected at a near-real-time setting, the assessment of the under-ascertainment of reported cases was made by evaluating the reporting delay distribution. For this purpose, the time interval from illness onset day  $S_i$  to report of case confirmation day  $R_i$  for each confirmed case  $i$  was extracted from the data. The doubly interval-censored likelihood function  $\mathcal{L}$  [20,47] was implemented for estimation:

$$\mathcal{L}(\theta | D) = \prod_i \int_{S_i}^{S_i+1} \int_{R_i}^{R_i+1} h(s) \cdot f(r - s | \theta) dr ds. \tag{6}$$

Here,  $h(\cdot)$  is the probability distribution function (PDF) of illness onset time following a uniform distribution, and  $f(\cdot | \theta)$  is the PDF of the reporting delay independent of  $h(\cdot)$ .  $D$  represents a dataset among all confirmed cases  $i$ , where both the time of illness onset  $s$  and the time of case confirmation  $r$  are defined with precision to one day, i.e., they belong to the time intervals  $(S_i, S_i + 1)$  and  $(R_i, R_i + 1)$ , respectively. The distribution to be defined is  $f(\cdot | \theta)$ , and it was fitted to gamma, lognormal, and Weibull distributions, each with a set of parameters  $\theta$ . The best-fit distribution was then selected by comparing the WAIC values and selecting the one with a lower value.

We note that an alternative way to implement a right-truncated likelihood is by using the following formula:

$$\mathcal{L}(\theta | D) = \prod_i \int_{S_i}^{S_i+1} \int_{R_i}^{R_i+1} h(s) \cdot \frac{f(r - s | \theta)}{F(t^* - s | \theta)} dr ds, \tag{7}$$

where  $t^*$  is the cut-off time at noon on 29 February 2020, and  $F(\cdot | \theta)$  is the cumulative distribution function of  $f(\cdot | \theta)$ . However, our analysis has shown that the mean delay was longer than two weeks, and this was inconsistent with real observations. The reason is that the right truncation does not account for the effect of control measures implemented in late February (see Discussion).

### 2.5. Simulation Platform

The data were processed using R version 3.6.2 and Python version 3.6.10. Markov chain Monte Carlo (MCMC) simulations were performed in Stan (cmdStan version 2.22.1 [48]) for estimation of the delay distribution, and in PyMC3 version 3.8 [49] for all other estimates. The code is available online [50].

## 3. Results

### 3.1. Epidemiological Situation

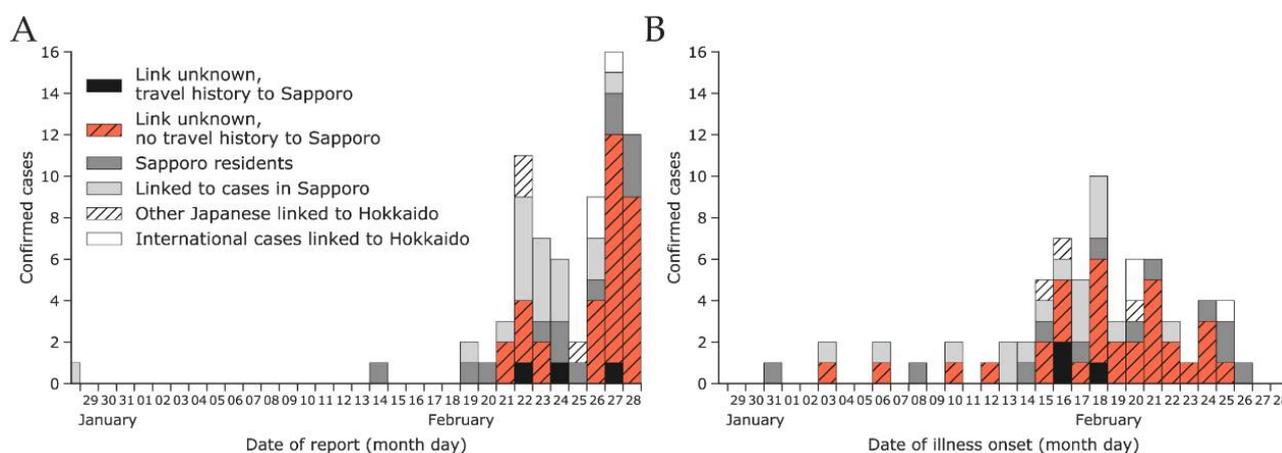
The first case was reported in Hokkaido on 28 January 2020. One month later, on 28 February 2020, the total count of confirmed cases reached 65, with 13 cases reported by Sapporo and another 51 cases reported by 11 of the 14 subprefectures of Hokkaido (and one elsewhere). The place of diagnosis for the first case was unspecified. Initially, cases were predominantly linked to the Sapporo Snow Festival, but overall, the geographic distribution of COVID-19 cases was widespread. By 28 February, Japan had also been notified of three cases diagnosed in Thailand and Malaysia who were believed to have been infected in Hokkaido, as well as three domestic cases likely exposed in Hokkaido that were reported by other prefectures (Table 1, Figure 1A).

**Table 1.** Exportation events and estimated incidence in Hokkaido by date of report.

Case Number	Diagnosis Location	Date of Illness Onset	Date of Report	Cumulative Count	Estimated Incidence in Hokkaido (95% CI)
1	Kumamoto	15 February 2020	22 February 2020	2	24 (4–74)
2	Chiba	16 February 2020	25 February 2020	3	36 (9–93)
3	Nagano	20 February 2020	26 February 2020	6	3446 (857–8931) #
4–5	Thailand	20 February 2020 NA	26 February 2020	6	2298 (382–7091) ##
6	Malaysia	25 February 2020	27 February 2020		

CI, confidence interval (the 95% CI was derived from profile likelihood); NA, not available. The estimated incidence is updated by the function of the date of report. The estimated incidence for the latest available reporting date (27 February 2020) depends on whether both cases 4 and 5 acquired their infection in Hokkaido (#) or whether one was infected by the other (##).

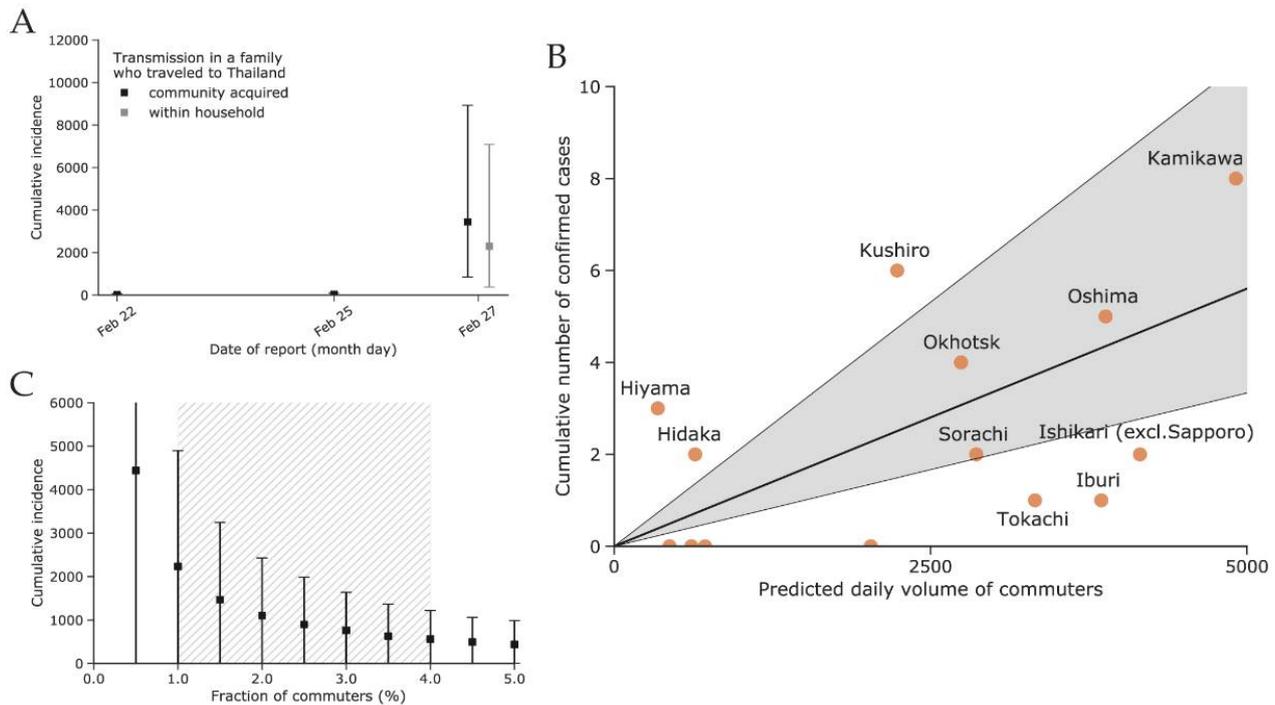
A surge of cases was observed at the end of February (Figure 2A). In the first half of February, peaks of reported cases by date of illness onset were separated by 3–4 days on average, consistent with estimates of the serial interval [51]. The epidemic curve peaked around 18 February 2020, but the subsequent decline in cases can be explained by the delay in reporting. The delay distribution fitted with the gamma distribution had a mean of 7.9 days (95% CI: 6.9–9.0) and a standard deviation of 4.2 days (95% CI: 3.3–5.2). The 95th percentile was 15.6 days (95% CI: 13.4–18.8), implying that cases with illness onset in the last two weeks of February were likely to be under-ascertained.



**Figure 2.** Epidemic curves by date of report (A) and date of illness onset (B) as of 29 February 2020 for confirmed cases among Japanese nationals that were linked to Hokkaido. The total number of cases is characterized according to the legend shown in (A).

### 3.2. Estimated Incidence in Hokkaido Using Confirmed Cases Diagnosed Outside Hokkaido (Method (i))

The first three cases diagnosed in early February outside of Hokkaido indicated a low incidence of COVID-19 in Hokkaido with an estimated upper bound (95th percentile) of fewer than 100 cases. Our estimate of the cumulative incidence in Hokkaido as of 25 February 2020 is dependent on whether the transmission within the infected couple who travelled to Thailand was acquired in the community or within the household, and we estimated the incidence in Hokkaido to be higher in the former scenario, at 3446 cases (95% CI: 857–8931), compared to the latter, with 2297 cases (95% CI: 382–7091); see Figure 3A.



**Figure 3.** (A) Estimated cumulative incidence of coronavirus disease 2019 (COVID-2019) in Hokkaido using the count of imported cases either internationally or domestically outbound from Hokkaido (method (i)). (B) Modeled daily volume of commuters and observed incidence of COVID-19 by subprefecture of diagnosis. The four subprefectures with zero case counts are Rumoi, Soya, Nemuro, and Shiribeshi (from left to right). (C) Estimated cumulative incidence in Sapporo as of 29 February with a varying fraction of commuters (method (ii)). The plausible range between 0.5% and 4% is indicated by the hatched grey area. The square points in (A,C) indicate the estimated mean, whereas the error bars indicate the 95% confidence intervals.

### 3.3. Estimated Incidence in Sapporo Using Confirmed Cases Diagnosed within Hokkaido (Method (ii))

A total of 34 Hokkaido cases were included in our model (Figure 1B). We fitted the radiation and uniform selection models to the population data for each subprefecture based on their centroids (Figure S1). The resulting fit to the observed incidence of COVID-19 cases showed that the uniform selection model performed better than the radiation model (WAIC values: 60.2 vs. 68.3, respectively). Figure 3B shows the fit using the model with uniform selection. If we assume the fraction of commuters to be at 1%, then the mean estimated incidence in Sapporo was 2233 cases (95% CI: 0–4893) as of 28 February 2020. Variation in the fraction of commuters between 0.5% and 4% results in an estimated incidence between 4440 (95% CI: 0–9687) and 563 (95% CI: 0–1221) for 0.5% and 4%, respectively (Figure 3C).

## 4. Discussion

Our estimate of the incidence in Sapporo in February–March 2020 is in the range of 1000–10,000 cases and resembles early estimates of COVID-19 incidence in Wuhan city, China that used data on the first cases among international travelers [3,52]. Assuming, however, that our estimates are correct, a substantial proportion of these infected persons

likely had only a mild or asymptomatic course of the disease [11,21], did not seek medical care, and were not detected by the surveillance system. The median estimated infection-to-confirmed-case ratio in Hokkaido was 26:1 (95% CI: 4.2–78.8), which includes, but is expectedly higher than, the 7:1 (95% CI: 5.5–10.0) estimate for the first wave of infections in the United States [25]. However, in the early stage of the pandemic, the case-finding strategy in Japan was based on extensive contact tracing [7], and differences in adopted case definitions between the two countries may have contributed to the observed difference in estimates. Our inference method was also less specific and did not rely on a compartmental Susceptible–Infected–Recovered (SIR) model as in [25].

Use of data on the travel volume of commuters and tourists within Hokkaido could be beneficial for fitting the travel volume to the observed incidence rather than adopting a model of human movement (Figure 3B). However, publicly available data on train, bus, and private car travel in Hokkaido were inconsistent and scarce, and we were unable to use them for this study. Nevertheless, our results demonstrate that the use of mathematical models of human movement at the scale of estimating the travel volume between administrative subunits of a given region is a promising alternative to the use of observed transportation data. Previously, this was also demonstrated by other researchers in assessing the risk of spread of yellow fever in Angola [36] or during epidemics in resource-poor settings [37].

In addition, surveillance among travelers at international borders may detect infections not found through standard surveillance systems due to the implementation of additional detection methods [53,54]. As shown by other researchers [46,55], the linear regression fit of the travel volume among international travelers to observed disease incidence represents an efficient way to distinguish areas that are able to detect most new infections from those that are likely to miss a larger fraction of cases. One limitation of our approach is that we were unable to distinguish international travelers from Hokkaido to Malaysia and Thailand who actually stayed in those countries from those who transited to another international destination. We applied the same approach to compare the performance of the various subprefectures of Hokkaido, as shown in Figure 3B. Subprefectures that fall within the 95% CI of the linear regression line are considered to have surveillance systems that succeed in detecting COVID-19 infections within their jurisdiction. Subprefectures that fall below the 95% CI of the linear regression line may not have surveillance systems that are sensitive enough to detect all imported cases [46]. The four subprefectures with no reported cases by the end of March were Rumoi, Soya, Nemuro, and Shiribeshi. The first three are distant and less connected to Sapporo compared to, for example, Kamikawa or Oshima, which contain the second and third largest cities of Hokkaido, Asahikawa and Hakodate (Figure 1B). However, Shiribeshi is located next to Ishikari subprefecture/Sapporo and includes the relatively large city of Otaru. The lack of reports in Shiribeshi may be explained by our use of place of diagnosis over place of residence for determining the subprefecture assigned to cases. This was done because the specific place of residence was not always reported. We suspect that some cases residing in Shiribeshi subprefecture were diagnosed in Ishikari subprefecture.

Our analysis of the reporting delay did not implement right truncation of the likelihood because the low case count (by date of illness onset) seen in late February through the beginning of March 2020 was highly likely due to the effect of control measures in Hokkaido rather than delays in reporting. In support of this assumption, there were fewer cases reported between 28 February and 6 March 2020 compared to the week of 21–27 February 2020 (cf. Figure 2 and Figure S2). Right truncation should, therefore, be used with caution when it is unclear whether recent incidence is increasing or decreasing.

In conclusion, our study demonstrates that combining data sources of imported COVID-19 cases at different (sub-prefectural, inter-prefectural, or international) levels represents a valuable way for estimating the actual incidence at the origin. Based on this approach, we showed that the under-ascertainment rate of COVID-19 cases in Hokkaido Prefecture at the beginning of the pandemic in February–March 2020 was much higher

than the analogous estimate from the United States in the corresponding period [25]. This could be explained by the different case-finding strategy used in Japan in early 2020, when mass testing of the public was rarely implemented and only the contacts of confirmed cases were investigated by the authorities [6,7]. Preliminary analysis of the offspring distribution among infected individuals in the detected clusters around Japan has shown that approximately 80% of cases do not produce secondary infections (i.e., their reproduction number is equal to zero) [56]. Because most of the clusters have been linked to the known cases and could be traced back in time, we argue that there is still a window of opportunity for containment of the disease, and only some small fraction of the infections may drive the epidemic [18]. We agree with other studies, e.g., [57,58], that a successful strategy for the control of COVID-19 lies in strict movement restriction, avoidance of social gatherings, and an intense investigation effort in contact tracing with the possible help of new information technologies such as digital assistance for contact tracing [59,60].

**Supplementary Materials:** The following are available online at <https://www.mdpi.com/article/10.3390/jcm10112392/s1>, Figure S1: Reconstructed connectivity matrix between subprefectures of Hokkaido, Figure S2: Epidemic curves by date of confirmation (A) and date of illness onset (B) as of 6 March 2020 for confirmed cases among Japanese nationals linked to Hokkaido.

**Author Contributions:** A.R.A., K.M. and H.N. conceived the study and participated in the study design. S.-M.J., K.M. and N.M.L. assisted in data collection. A.R.A. and K.M. analyzed the data. A.R.A., N.M.L. and K.M. drafted the manuscript. All authors interpreted the results, edited the manuscript and approved the final version. All authors have read and agreed to the published version of the manuscript.

**Funding:** This publication was supported by the German Federal Ministry of Health (BMG) COVID-19 Research and Development funding to the WHO. H.N. received funding from the Japan Agency for Medical Research and Development (AMED) (grant number: JP18fk0108050); the Japan Society for the Promotion of Science (JSPS) KAKENHI (grant numbers: 17H04701, 17H05808, 18H04895, and 19H01074); the Inamori Foundation, GAP fund program by Kyoto University; the Japan Science and Technology Agency (JST) CREST program (grant number: JPMJCR1413); and the SICORP (e-ASIA) program (JPMJSC20U3). K.M. acknowledges support from the JSPS KAKENHI (grant no. 15K20936); the Program for Advancing Strategic International Networks to Accelerate the Circulation of Talented Researchers (grant no. G2801); and the Leading Initiative for Excellent Young Researchers from the Ministry of Education, Culture, Sport, Science, and Technology of Japan. S.-M.J. and N.M.L. received graduate study scholarships from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Publicly available datasets were analyzed in this study. This data can be found here [50].

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Huang, C.; Wang, Y.; Li, X.; Zhao, J.; Hu, Y.; Zhang, L.; Fan, G.; Xu, J.; Gu, X.; Cheng, Z.; et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *Lancet* **2020**, *395*, 497–506. [CrossRef]
2. Hui, D.S.I.; Azhar, E.; Madani, T.A.; Ntoumi, F.; Kock, R.; Dar, O.; Ippolito, G.; McHugh, T.D.; Memish, Z.A.; Drosten, C.; et al. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health—The latest 2019 novel coronavirus outbreak in Wuhan, China. *Int. J. Infect. Dis.* **2020**, *91*, 264–266. [CrossRef]
3. Nishiura, H.; Jung, S.; Linton, N.M.; Ryo, K.; Yichi, Y.; Katsuma, H.; Tetsuro, K.; Baoyin, Y.; Andrei, R.A. The extent of transmission of novel coronavirus in Wuhan, China, 2020. *J. Clin. Med.* **2020**, *9*, 330. [CrossRef]
4. Chinazzi, M.; Davis, J.T.; Ajelli, M. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* **2020**, 395–400. [CrossRef] [PubMed]
5. WHO Coronavirus disease 2019 (COVID-19) Situation Report 53. 2020. Available online: <https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200313-sitrep-53-covid-19.pdf> (accessed on 13 May 2020).

6. Omori, R.; Mizumoto, K.; Chowell, G. Changes in testing rates could mask the novel coronavirus disease (COVID-19) growth rate. *Int. J. Infect. Dis.* **2020**, *94*, 116–118. [[CrossRef](#)] [[PubMed](#)]
7. Oshitani, H. The Expert Members of The National COVID-19 Cluster Taskforce at The Ministry of Health, Labour and Welfare, Japan. Cluster-Based Approach to Coronavirus Disease 2019 (COVID-19) Response in Japan, from February to April 2020. *Jpn. J. Infect. Dis.* **2020**, *73*, 491–493. [[CrossRef](#)] [[PubMed](#)]
8. Rothe, C.; Schunk, M.; Sothmann, P. Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany. *N. Engl. J. Med.* **2020**, *382*, 970–971. [[CrossRef](#)] [[PubMed](#)]
9. Tian, S.; Hu, N.; Lou, J. Characteristics of COVID-19 infection in Beijing. *J. Infect.* **2020**, *80*, 401–406. [[CrossRef](#)]
10. WHO Report of the WHO-China Joint Mission on Coronavirus Disease 2019 (COVID-19). WHO Report 16–24 February 2020. Available online: <https://www.who.int/docs/default-source/coronaviruse/who-china-joint-mission-on-covid-19-final-report.pdf> (accessed on 5 March 2020).
11. He, X.; Lau, E.H.Y.; Wu, P.; Deng, X.; Wang, J.; Hao, X.; Lau, Y.C.; Wong, Y.J.; Guan, Y.; Tan, X. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nat. Med.* **2020**, *26*, 672–675. [[CrossRef](#)]
12. Cheng, H.-Y.; Jian, S.-W.; Liu, D.-P.; Ng, T.-C.; Huang, W.-T.; Lin, H.-H.; for the Taiwan COVID-19 Outbreak Investigation Team. Contact tracing assessment of COVID-19 transmission dynamics in Taiwan and risk at different exposure periods before and after symptom onset. *JAMA Intern. Med.* **2020**, *180*, 1156–1163. [[CrossRef](#)]
13. Nishiura, H.; Kobayashi, T.; Yang, Y. The rate of underascertainment of novel coronavirus (2019-nCoV) infection: Estimation using Japanese passengers data on evacuation flights. *J. Clin. Med.* **2020**, *9*, 419. [[CrossRef](#)] [[PubMed](#)]
14. Nishiura, H. Backcalculating the incidence of infection with COVID-19 on the Diamond Princess. *J. Clin. Med.* **2020**, *9*, 657. [[CrossRef](#)] [[PubMed](#)]
15. Mizumoto, K.; Kagaya, K.; Zarebski, A.; Chowell, G. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Eurosurveillance* **2020**, *25*, 10. [[CrossRef](#)] [[PubMed](#)]
16. Ong, S.W.X.; Tan, Y.K.; Chia, P.Y.; Lee, H.T.; Oon, T.N.; Wong, Y.S.M.; Marimuthu, K. Air, surface environmental, and personal protective equipment contamination by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) from a symptomatic patient. *JAMA* **2020**, *323*, 1610–1612. [[CrossRef](#)]
17. Xu, Y.; Li, X.; Zhu, B. Characteristics of pediatric SARS-CoV-2 infection and potential evidence for persistent fecal viral shedding. *Nat. Med.* **2020**, *26*, 502–505. [[CrossRef](#)]
18. Grantz, K.; Metcalf, C.; Lessler, J. Dispersion vs. Control. 2020. Available online: <https://hopkinsidd.github.io/nCoV-Sandbox/DispersionExploration.html> (accessed on 5 March 2020).
19. Endo, A.; Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group; Abbott, S. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open. Res.* **2020**, *5*, 672. [[CrossRef](#)]
20. Linton, N.M.; Kobayashi, T.; Yang, Y. Incubation period and other Epidemiological Characteristics of 2019 Novel Coronavirus Infections with Right Truncation: A Statistical Analysis of Publicly Available Case Data. *J. Clin. Med.* **2020**, *9*, 538. [[CrossRef](#)]
21. Bi, Q.; Wu, Y.; Mei, S.; Ye, C.; Zou, X.; Zhang, Z.; Liu, X.; Wei, L.; Truelove, S.A.; Zhang, T.; et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: A retrospective cohort study. *Lancet Infect. Dis.* **2020**, *20*, 911–919. [[CrossRef](#)]
22. Lee, J.-S.; Kim, S.Y.; Kim, T.S.; Hong, K.H.; Ryoo, N.-H.; Lee, J. Evidence of Severe Acute Respiratory Syndrome Coronavirus 2 reinfection after recovery from mild Coronavirus Disease 2019. *Clin. Infect. Dis.* **2020**, *ciaa1421*. [[CrossRef](#)]
23. Guan, W.; Ni, Z.; Hu, Y. Clinical characteristics of coronavirus disease 2019 in China. *N. Engl. J. Med.* **2020**, *382*, 1708–1720. [[CrossRef](#)]
24. Hauser, A.; Counotte, M.J.; Margossian, C.C.; Konstantinoudis, G.; Low, N.; Althaus, C.; Riou, J. Estimation of SARS-CoV-2 mortality during the early stages of an epidemic: A modeling study in Hubei, China, and six regions in Europe. *PLoS Med.* **2020**, *17*, e1003189. [[CrossRef](#)]
25. Li, R.; Pei, S.; Chen, B.; Song, Y.; Zhang, T.; Yang, W.; Shaman, J. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **2020**, *368*, 489–493. [[CrossRef](#)]
26. Anderson, R.M.; Heesterbeek, H.; Klinkenberg, D.; Hollingsworth, T. How will country-based mitigation measures influence the course of the COVID-19 epidemic? *Lancet* **2020**, *395*, 931–934. [[CrossRef](#)]
27. Tuite, A.; Ng, V.; Rees, E.; Fisman, D. Estimation of COVID-19 outbreak size in Italy based on international case exportations. *Lancet Infect. Dis.* **2020**, *20*, 520. [[CrossRef](#)]
28. Kumamoto Prefectural Government. Patients Associated with the New Coronavirus 2020. Available online: <https://www.mhlw.go.jp/content/10906000/000599322.pdf> (accessed on 5 March 2020). (In Japanese)
29. Nagano Prefectural Government. Patients Associated with the New Coronavirus 2020. Available online: <https://www.mhlw.go.jp/content/10906000/000599974.pdf> (accessed on 5 March 2020). (In Japanese)
30. Chiba Prefectural Government. Patients Associated with the New Coronavirus 2020. Available online: <https://www.mhlw.go.jp/content/10906000/000599329.pdf> (accessed on 5 March 2020). (In Japanese)
31. Kyodo News. Patients Associated with the New Coronavirus 2020. Available online: <https://english.kyodonews.net/news/2020/02/2622cef60938-malaysian-woman-tests-positive-for-coronavirus-after-japan-trip.html> (accessed on 5 March 2020). (In Japanese).

32. WHO Thailand. Coronavirus Disease 2019 (COVID-19) Situation Report—26 February 2020. Available online: <https://www.who.int/docs/default-source/searo/thailand/20200226-tha-sitrep-09-covid-19-final.pdf> (accessed on 5 March 2020). (In Japanese).
33. Pan, X.; Chen, D.; Xia, Y. Asymptomatic cases in a family cluster with SARS-CoV-2 infection. *Lancet Infect. Dis.* **2020**, *20*, 410–411. [[CrossRef](#)]
34. Hokkaido Prefecture, Japan. Basic Resident Register Information 2018. Available online: <http://www.pref.hokkaido.lg.jp/ss/tuk/900brr/index2.htm> (accessed on 2 March 2020).
35. Sapporo City. National Census: Population and Employment Status by Employment Location and School Attendance and Population Moving 2015. Available online: <https://www.city.sapporo.jp/toukei/tokusyu/kokuseityosa.html> (accessed on 2 March 2020). (In Japanese).
36. Kraemer, M.U.G.; Faria, N.R.; Reiner, R.C.; Golding, N.; Nikolay, B.; Stasse, S. Spread of yellow fever virus outbreak in Angola and the Democratic Republic of the Congo 2015–16: A modelling study. *Lancet Infect. Dis.* **2017**, *17*, 330–338. [[CrossRef](#)]
37. Kraemer, M.U.G.; Golding, N.; Bisanzio, D. Utilizing general human movement models to predict the spread of emerging infectious diseases in resource poor settings. *Sci. Rep.* **2019**, *9*, 5151. [[CrossRef](#)]
38. Golding, N.; Wuensch, K.; Schofield, A. SEEG-Oxford/movement: Modelling and Analysing Human Movement Data. R package version 0.4. Available online: <https://github.com/SEEG-Oxford/movement/> (accessed on 5 March 2020).
39. Spadon, G.; de Carvalho, A.C.P.L.F.; Rodrigues, J.F., Jr.; Alves, L.G.A. Reconstructing commuters network using machine learning and urban indicators. *Sci. Rep.* **2019**, *9*, 11801. [[CrossRef](#)]
40. Simini, F.; Maritan, A.; Néda, Z. Human mobility in a continuum approach. *PLoS ONE* **2013**, *8*, e60069. [[CrossRef](#)]
41. Simini, F.; González, M.C.; Maritan, A.; Barabási, A.-L. A universal model for mobility and migration patterns. *Nature* **2012**, *484*, 96–100. [[CrossRef](#)] [[PubMed](#)]
42. Dorigatti, I.; Hamlet, A.; Aguas, R. International risk of yellow fever spread from the ongoing outbreak in Brazil, December 2016 to May 2017. *Eurosurveillance* **2017**, *22*, 3057. [[CrossRef](#)] [[PubMed](#)]
43. Tsuzuki, S.; Lee, H.; Miura, F. Dynamics of the pneumonic plague epidemic in Madagascar, August to October 2017. *Eurosurveillance* **2017**, *22*, 17-00710. [[CrossRef](#)] [[PubMed](#)]
44. Shimizu, K.; Nishiura, H.; Imamura, A. Investigation of the proportion of diagnosed people living with HIV/AIDS among foreign residents in Japan. *J. Clin. Med.* **2019**, *8*, 804. [[CrossRef](#)] [[PubMed](#)]
45. Zou, L.; Ruan, F.; Huang, M. SARS-CoV-2 viral load in upper respiratory specimens of infected patients. *N. Engl. J. Med.* **2020**, *382*, 1177–1179. [[CrossRef](#)]
46. De Salazar, P.M.; Niehus, R.; Taylor, A.; Buckee, C.; Lipsitch, M. Identifying locations with possible undetected imported severe acute respiratory syndrome coronavirus 2 cases by using importation predictions. *Emerg. Infect. Dis.* **2020**, *26*, 1465–1469. [[CrossRef](#)]
47. Reich, N.G.; Lessler, J.; Cummings, D.A.T.; Brookmeyer, R. Estimating incubation period distributions with coarse data. *Stat. Med.* **2009**, *28*, 2769–2784. [[CrossRef](#)]
48. Carpenter, B.; Gelman, A.; Hoffman, M.D. Stan: A probabilistic programming language. *J. Stat. Softw.* **2017**, *76*, 1–32. [[CrossRef](#)]
49. Salvatier, J.; Wiecki, T.V.; Fonnesbeck, C. Probabilistic programming in Python using PyMC3. *PeerJ Comput. Sci.* **2016**, *2*, e55. [[CrossRef](#)]
50. Available online: <http://github.com/aakhmetz/Covid19IncidenceHokkaidoFeb2020> (accessed on 27 May 2021).
51. Nishiura, H.; Linton, N.M.; Akhmetzhanov, A.R. Serial interval of novel coronavirus (COVID-19) infections. *Int. J. Infect. Dis.* **2020**, *93*, 284–286. [[CrossRef](#)]
52. Imai, N.; Dorigatti, I.; Cori, A. *Report 2: Estimating the Potential Total Number of Novel Coronavirus Cases in Wuhan City, China*; Imperial College London: London, UK, 2020.
53. Anzai, A.; Kobayashi, T.; Linton, N.M. Assessing the impact of reduced travel on exportation dynamics of novel coronavirus infection (COVID-19). *J. Clin. Med.* **2020**, *9*, 601. [[CrossRef](#)]
54. Quilty, B.J.; Clifford, S.; CMMID nCoV Working Group. Effectiveness of airport screening at detecting travellers infected with novel coronavirus (2019-nCoV). *Eurosurveillance* **2020**, *25*. [[CrossRef](#)]
55. Niehus, R.; De Salazar, P.M.; Taylor, A.R.; Lipsitch, M. Using observational data to quantify bias of traveller-derived COVID-19 prevalence estimates in Wuhan, China. *Lancet Infect. Dis.* **2020**, *20*, 803–808. [[CrossRef](#)]
56. Nishiura, H.; Oshitani, H.; Kobayashi, T. Closed environments facilitate secondary transmission of coronavirus disease 2019 (COVID-19). *medRxiv* **2020**. [[CrossRef](#)]
57. Wang, C.J.; Ng, C.Y.; Brook, R.H. Response to COVID-19 in Taiwan: Big data analytics, new technology, and proactive testing. *JAMA* **2020**, *323*, 1341. [[CrossRef](#)] [[PubMed](#)]
58. Zhang, J.; Litvinova, M.; Wang, W. Evolving epidemiology and transmission dynamics of coronavirus disease 2019 outside Hubei province, China: A descriptive and modelling study. *Lancet Infect. Dis.* **2020**, *20*, 793–802. [[CrossRef](#)]
59. Ferretti, L.; Wymant, C.; Kendall, M.; Zhao, L.; Nurtay, A.; Abeler-Dorner, L.; Parker, M.; Bonsall, D.; Fraser, C. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. *Science* **2020**, *368*, eabb6936. [[CrossRef](#)] [[PubMed](#)]
60. Jian, S.-H.; Cheng, H.-Y.; Huang, X.-T.; Liu, D.-P. Contact tracing with digital assistance in Taiwan’s COVID-19 outbreak response. *Intern. J. Infect. Dis.* **2020**, *101*, 348–352. [[CrossRef](#)] [[PubMed](#)]