How infectious was COVID-19 when it first circulated in Wuhan?

Qingyuan Zhao

23rd March, 2022
Corpus Christi College Masterclass in Mathematics

Introduction

About me

- I am a mathematician/statistician by training.
- Now am University Assistant Professor at the Statistical Laboratory (a sub-department in DPMMS) and a Fellow of Corpus.
- I grew up in Wuhan, China.

Introduction

About me

- I am a mathematician/statistician by training.
- Now am University Assistant Professor at the Statistical Laboratory (a sub-department in DPMMS) and a Fellow of Corpus.
- I grew up in Wuhan, China.

In this talk

- What could we have said about the epidemic outbreak in Wuhan in January and February, 2020?
- Why mathematics and statistics are critical to answer this correctly?

Caveats

- I will refer to the disease as COVID-19, but that name has not appeared back then.
- So there is potentially a large hindsight bias.

1–20 January, 2020: Can COVID-19 be transmitted from human to human?

Table: Total number of confirmed “cases” in Wuhan by date of press release.

Date	12-31	01-03	**01-05**	**01-11**	01-12	01-13	**01-14**	01-15	01-16	01-18	01-19
Cases	27	44	**59**	**41**	41	41	**41**	41	41	45	62
Date	**01-20**	01-21	01-23	01-24	01-25	01-26	01-27	01-28	01-29	01-30	01-31
Cases	**198**	258	425	495	572	618	698	1590	1905	2261	2639

1–20 January, 2020: Can COVID-19 be transmitted from human to human?

Table: Total number of confirmed “cases” in Wuhan by date of press release.

Date	12-31	01-03	**01-05**	**01-11**	01-12	01-13	**01-14**	01-15	01-16	01-18	01-19
Cases	27	44	**59**	**41**	41	41	**41**	41	41	45	62
Date	**01-20**	01-21	01-23	01-24	01-25	01-26	01-27	01-28	01-29	01-30	01-31
Cases	**198**	258	425	495	572	618	698	1590	1905	2261	2639

Official press releases

No clear evidence *of human-to-human transmission.* **(Jan 5, 2020)**

1–20 January, 2020: Can COVID-19 be transmitted from human to human?

Table: Total number of confirmed “cases” in Wuhan by date of press release.

Date	12-31	01-03	**01-05**	**01-11**	01-12	01-13	**01-14**	01-15	01-16	01-18	01-19
Cases	27	44	**59**	**41**	41	41	**41**	41	41	45	62
Date	**01-20**	01-21	01-23	01-24	01-25	01-26	01-27	01-28	01-29	01-30	01-31
Cases	**198**	258	425	495	572	618	698	1590	1905	2261	2639

Official press releases

No clear evidence *of human-to-human transmission.* **(Jan 5, 2020)**

[The disease] **can be prevented and contained.** **Jan 9, 2020)**

1–20 January, 2020: Can COVID-19 be transmitted from human to human?

Table: Total number of confirmed "cases" in Wuhan by date of press release.

Date	12-31	01-03	**01-05**	**01-11**	01-12	01-13	**01-14**	01-15	01-16	01-18	01-19	**01-20**	01-21	01-23	01-24	01-25	01-26	01-27	01-28	01-29	01-30	01-31
Cases	27	44	**59**	**41**	41	41	**41**	41	41	45	62	**198**	258	425	495	572	618	698	1590	1905	2261	2639

Official press releases

No clear evidence *of human-to-human transmission.* **(Jan 5, 2020)**

[The disease] **can be prevented and contained.** **Jan 9, 2020)**

No clear evidence *of human-to-human transmission . . . Although we cannot exclude the possibility of limited human-to-human transmission, the* **risk of sustained transmission is low**.

1–20 January, 2020: Can COVID-19 be transmitted from human to human?

Table: Total number of confirmed "cases" in Wuhan by date of press release.

Date	12-31	01-03	**01-05**	**01-11**	01-12	01-13	**01-14**	01-15	01-16	01-18	01-19
Cases	27	44	**59**	**41**	41	41	**41**	41	41	45	62
Date	**01-20**	01-21	01-23	01-24	01-25	01-26	01-27	01-28	01-29	01-30	01-31
Cases	**198**	258	425	495	572	618	698	1590	1905	2261	2639

Official press releases

No clear evidence *of human-to-human transmission.* **(Jan 5, 2020)**

[The disease] **can be prevented and contained**. **Jan 9, 2020)**

No clear evidence *of human-to-human transmission . . . Although we cannot exclude the possibility of limited human-to-human transmission, the* **risk of sustained transmission is low***.* **(Jan 14, 2020)**

Novel coronavirus **can certainly be transmitted** *from human to human.* **(Jan 20, 2020)**

1–20 January, 2020: Can COVID-19 be transmitted from human to human?

Table: Total number of confirmed “cases” in Wuhan by date of press release.

Date	12-31	01-03	**01-05**	**01-11**	01-12	01-13	**01-14**	01-15	01-16	01-18	01-19
Cases	27	44	**59**	**41**	41	41	**41**	41	41	45	62
Date	**01-20**	01-21	01-23	01-24	01-25	01-26	01-27	01-28	01-29	01-30	01-31
Cases	**198**	258	425	495	572	618	698	1590	1905	2261	2639

Official press releases

No clear evidence *of human-to-human transmission.* **(Jan 5, 2020)**

[The disease] **can be prevented and contained**. **Jan 9, 2020)**

No clear evidence *of human-to-human transmission . . . Although we cannot exclude the possibility of limited human-to-human transmission, the* **risk of sustained transmission is low**. **(Jan 14, 2020)**

Novel coronavirus **can certainly be transmitted** *from human to human.* **(Jan 20, 2020)**

But data can (and often will) lie

Table: Total number of confirmed “cases” in Wuhan by date of press release.

Date	12-31	01-03	**01-05**	**01-11**	01-12	01-13	**01-14**	01-15	01-16	01-18	01-19
Cases	27	44	**59**	**41**	41	41	**41**	41	41	45	62
Date	**01-20**	01-21	01-23	01-24	01-25	01-26	01-27	01-28	01-29	01-30	01-31
Cases	**198**	258	425	495	572	618	698	1590	1905	2261	2639

But data can (and often will) lie

Table: Total number of confirmed “cases” in Wuhan by date of press release.

Date	12-31	01-03	**01-05**	**01-11**	01-12	01-13	**01-14**	01-15	01-16	01-18	01-19	**01-20**	01-21	01-23	01-24	01-25	01-26	01-27	01-28	01-29	01-30	01-31
Cases	27	44	**59**	**41**	41	41	**41**	41	41	45	62	**198**	258	425	495	572	618	698	1590	1905	2261	2639

- Suspicious observations: the number dropped on 11th January and stayed flat for the next week.

But data can (and often will) lie

Table: Total number of confirmed “cases” in Wuhan by date of press release.

Date	12-31	01-03	**01-05**	**01-11**	01-12	01-13	**01-14**	01-15	01-16	01-18	01-19
Cases	27	44	**59**	**41**	41	41	**41**	41	41	45	62
Date	**01-20**	01-21	01-23	01-24	01-25	01-26	01-27	01-28	01-29	01-30	01-31
Cases	**198**	258	425	495	572	618	698	1590	1905	2261	2639

- Suspicious observations: the number dropped on 11th January and stayed flat for the next week.
- What happened? PCR tests were being developed and distributed.

Evidence elsewhere

1. COVID-19’s pathogen—SARS-CoV-2—is far from the first human coronavirus.
2. Genomic sequencing already identified a SARS-like coronavirus in late December, 2019.
3. Several hospitals in Wuhan were seeing rapidly increasing numbers of suspected cases in early and mid January, 2020.

January 20–End of February, 2020: How infectious was COVID-19 in Wuhan?

Basic reproductive number

- $R_0 \approx$ average number of infectees per infected person in the beginning of an outbreak.
- Estimated R_0 of SARS: 2 to 4; MERS: < 1; seasonal influenza: < 2.
- Difficult to estimate. Precise definition depends on the mathematical model.

Initial doubling time

This is arguably more relevant in early outbreaks.

- Better captures the urgency;
- Is much easier to estimate.

Two initial studies

1. Q. Li, et al. *New England Journal of Medicine*.
   - Published online: Jan 29, 2020
   - Estimated initial doubling time: **7.4 days (95% CI, 4.2 to 14)**;
   - Estimated R_0: **2.2 (95% CI, 1.4 to 3.9)**;
2. J. T. Wu, et al. *Lancet*.
   - Published online: Jan 31, 2020.
   - Estimated initial doubling time: **6.4 days (95% CI 5.8 to 7.1)**;
   - Estimated R_0: **2.68 (95% CI 2.47 to 2.86)**.

Two initial studies

1. Q. Li, et al. *New England Journal of Medicine*.
    - Published online: Jan 29, 2020
    - Estimated initial doubling time: **7.4 days (95% CI, 4.2 to 14)**;
    - Estimated R_0: **2.2 (95% CI, 1.4 to 3.9)**;
2. J. T. Wu, et al. *Lancet*.
    - Published online: Jan 31, 2020.
    - Estimated initial doubling time: **6.4 days (95% CI 5.8 to 7.1)**;
    - Estimated R_0: **2.68 (95% CI 2.47 to 2.86)**.

CI means **confidence interval**. The meaing of "confidence" can be made precise using mathematics.

Two initial studies

1. Q. Li, et al. *New England Journal of Medicine*.
   - Published online: Jan 29, 2020
   - Estimated initial doubling time: **7.4 days (95% CI, 4.2 to 14)**;
   - Estimated R_0: **2.2 (95% CI, 1.4 to 3.9)**;
2. J. T. Wu, et al. *Lancet*.
   - Published online: Jan 31, 2020.
   - Estimated initial doubling time: **6.4 days (95% CI 5.8 to 7.1)**;
   - Estimated R_0: **2.68 (95% CI 2.47 to 2.86)**.

CI means **confidence interval**. The meaing of "confidence" can be made precise using mathematics.

Next: Reanalyses of the same data

A useful mathematical tool for us will be the **Poisson log-linear model**.

- Let Y_t be the number of "incidences" in day t.
- This model assumes Y_t follows a Poisson distribution, and $\log(\mathbb{E}[Y_t]) = a + rt$.
- So the doubling time is $\log(2)/r$ (in days), if $r > 0$.

First paper: Author description of statistical analysis

We estimated the epidemic growth rate by analyzing data on the cases with illness onset **between December 10 and January 4** *because we expected the proportion of infections identified would increase soon after the formal announcement of the outbreak in Wuhan on December 31.*

We fitted a transmission model (formulated with the use of renewal equations) with zoonotic infections to onset dates that were not linked to the Huanan Seafood Wholesale Market, and we used this model to derive the epidemic growth rate, the epidemic doubling time, and the basic reproductive number (R_0).

- No further details about the transmission model was given besides that it was fitted using MATLAB.

First paper: Re-analysis

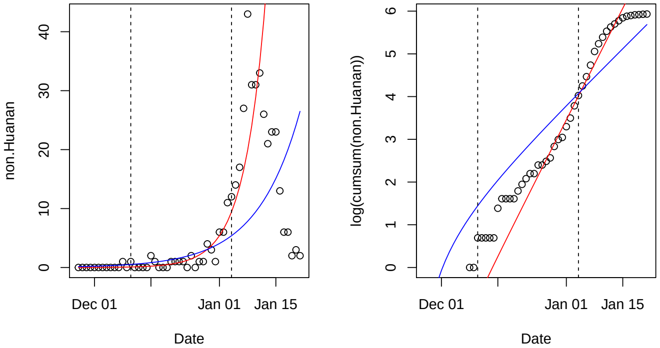


Figure: Initial epidemic curve in Wuhan and the fitted log-linear models (using the incidences between December 10 and January 4, dashed lines). The red curves correspond to an unrestricted fit; the blue curves correspond to the best fit assuming that the growth exponent correspond to a doubling time of 7.4 days.

- Estimated initial doubling time by the Poisson log-linear model: **3.7 days (2.8 to 5.1)**.

Second paper: Author description of statistical analysis

78 exported cases *from Wuhan to areas outside mainland China.*

We used the following **susceptible-exposed-infectious-recovered (SEIR) model** *to simulate the Wuhan epidemic*

We assumed that . . . international case exportation occurred according to a **non-homogeneous [Poisson] process**.

We estimated R_0 *using* **Markov Chain Monte Carlo** *methods with Gibbs sampling and non-informative flat prior.*

We estimated . . . on the basis of . . . confirmed cases . . . whose symptom onset date had been reported to fall from **Dec 25, 2019, to Jan 19, 2020**. *. . . This end date was chosen to minimise the effect of lead time bias on case confirmation.*

- Not mentioned: this criterion left them with **just 17 cases**.

Second paper: Re-analysis

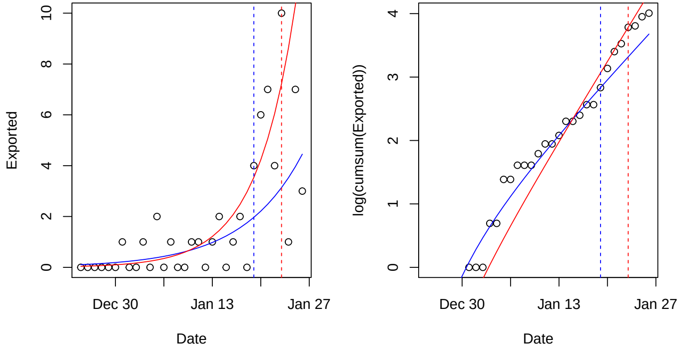


Figure: Initial epidemic curve for Wuhan-exported cases and the fitted log-linear models. The blue curves correspond to using the incidences between December 25 and January 19 (blue dashed lines). The red curves correspond to fitting the same model using data up to January 23 (red dashed lines).

- Estimated initial doubling time by a Poisson log-linear model: **5.9 days (3.4 to 15.7)**.
- Confidence interval is much wider than the original paper.
- Smaller point estimate (3.9 days) had the end date been extended.

Beyond modelling epidemic curves

Point process perspective

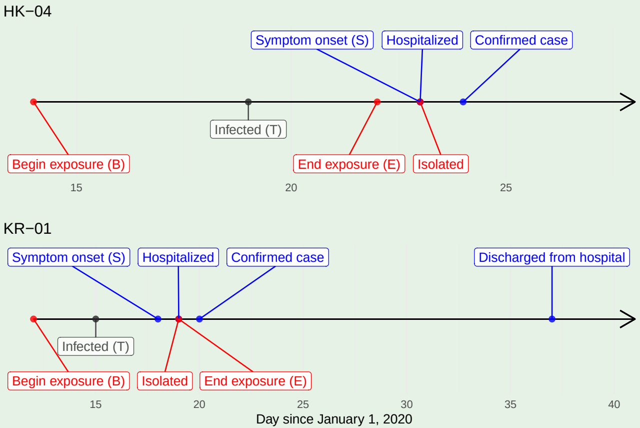


Figure: Timelines of two COVID-19 cases. The color indicates type of event.

A better statistical model for exported cases

Key events in the trajectory of one COVID-19 case

1. Begin of exposure, B;
2. End of exposure, E;
3. Onset of symptoms, S.

These events can help to infer the latent time of infection, T:

- Logical constraint: $B \leq T \leq E$;
- $S - T$ is the incubation period.

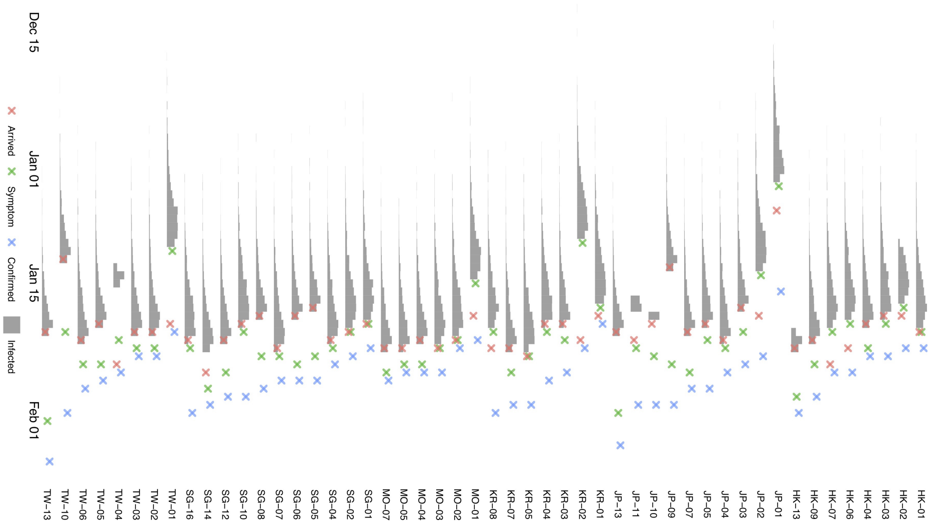
The effect of travel quarantine: **Selection bias**

Instead of a simple exponential growth model for the density of T: $f_T(t) \propto e^{rt}$, a better model is

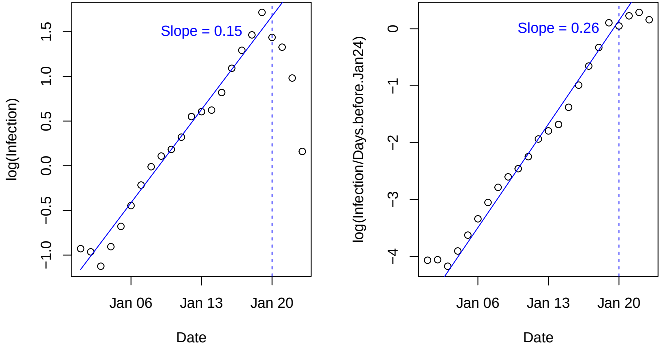
$$f_T(t) \propto e^{rt} \cdot \max(L - t, 0),$$

where L is the time of travel quarantine of Wuhan (23rd January, 2020).

Impute infection times



Fit modified growth models



- Each point corresponds to estimated $f_T(t)$ after the specified transformation.
- Left: $f_T(t) \propto e^{rt} \implies \log f_T(t) = rt + \text{constant}$.
- Right: $f_T(t) \propto e^{rt} \max(L - t, 0) \implies \log\{f_T(t)/(L - t)\} = rt + \text{constant for } t \leq L$.

Estimating R_0

Two ways to estimate R_0

1. R_0 as a derived parameter from compartmental models;
2. Through the formula $R = 1/M(-r)$:
    - $M(\cdot)$ is the **moment generating function** of the generation time;
    - r is the growth exponent.

- Both approaches often require estimating external parameters (e.g. distributions of the incubation period or generation time).
- Surprisingly (to serious statisticians), none of the early COVID-19 studies considered uncertainty in external parameters.

Large unaccounted uncertainty about R_0

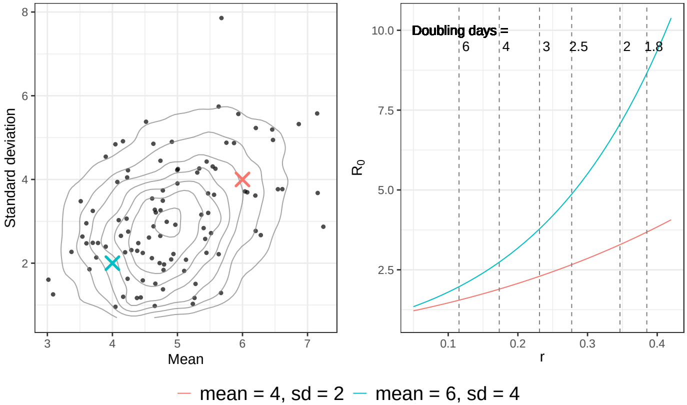


Figure: Estimated generation time and the implied basic reproductive number. The left panel shows 100 samples from the posterior distribution of the mean and standard deviation of the generation time in a replication analysis. The right panel shows the implied R_0 for different values of r and two generation time distributions.

What did we talk about?

- Data can (and often will) lie.
- We can tame the data by a good understanding of how they are generated and applying the appropriate mathematical tools.

Advanced mathematical/statistical concepts in this talk

- **Confidence interval**: *IB Statistics*
- **Poisson log-linear model**: *II Statistical Modelling*;
- **SEIR model**: *IA Differential Equations* (and many more courses).
- **Poisson process**, which is a special **point process**: *II Applied Probability*.
- **Markov Chain Monte Carlo**: *IB Markov Chains*, *II Principles of Statistics*.
- **Moment generating function**: *IA Probability*.