Uncovering the number of undetected infections

Lucy Li, PhD

Data Scientist Chan Zuckerberg Biohub 2 @lucylmq | @czbiohub

Why are the undetected infections important?

Undetected infections can happen because of:

- Asymptomatic or subclinical infections
- Limited testing capacity
- Restricted criteria for testing

Consequence of large number of undetected infections:

- Can spread through the population without being noticed
- Potential to spread across large geographic regions unnoticed
- Takes longer to be certain of disease elimination

- Mass testing
- Case count data
- Viral genomic data

Mass testing

- Case count data



Seattle & King County

Helping Public Health Leaders Track the Spread of Coronavirus

To slow the spread of coronavirus (COVID-19), we need to learn more about it. That's where you come in. SCAN is testing both healthy and sick people to understand how the virus is spreading in King County. The findings will help our partners, including Public Health -Seattle & King County, keep people informed and make the best possible, data-driven decisions to protect our community.

RAPID COMMUNICATION

Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020

Kenji Mizumoto^{1,2,3} , Katsushi Kagaya^{2,4} , Alexander Zarebski⁵ , Gerardo Chowell³

- 1. Graduate School of Advanced Integrated Studies in Human Survivability, Kyoto University Yoshida-Nakaadachi-cho, Sakyo-ku, Kyoto, Japan
- 2. Hakubi Center for Advanced Research, Kyoto University, Yoshidahonmachi, Sakyo-ku, Kyoto, Japan
- 3. Department of Population Health Sciences, School of Public Health, Georgia State University, Atlanta, Georgia, United States
- 4. Seto Marine Biological Laboratory, Field Science, Education and Reseach Center, Kyoto University, Shirahama-cho, Nishimurogun, Wakayama, Japan
- 5. Department of Zoology, University of Oxford, Oxford, United Kingdom

Correspondence: Kenji Mizumoto (mizumoto.kenji.5a@kyoto-u.ac.jp)

Citation style for this article:

Mizumoto Kenji, Kagaya Katsushi, Zarebski Alexander, Chowell Gerardo. Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. Euro Surveill. 2020;25(10):pii=2000180. https://doi.org/10.2807/1560-7917. ES.2020.25.10.2000180



- Mass testing
- Case count data
- Viral genomic data

Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV2)

Ruiyun Li^{1,*}, Sen Pei^{2,*,†}, Bin Chen^{3,*}, Yimeng Song⁴, Tao Zhang⁵, Wan Yang⁶, Jeffrey Shaman^{2,†} + See all authors and affiliations

Science 16 Mar 2020: eabb3221 DOI: 10.1126/science.abb3221 The Rate of Underascertainment of Novel Coronavirus (2019-nCoV) Infection: Estimation Using Japanese Passengers Data on Evacuation Flights

- by ♥ Hiroshi Nishiura ^{1,2,*} ♥ ♥, ♥ Tetsuro Kobayashi ¹ ♥, ♥ Yichi Yang ¹ ♥, ♥ Katsuma Hayashi ¹ ♥, ♥ Takeshi Miyama ³ ♥, ♥ Ryo Kinoshita ¹ ♥ ♥, ♥ Natalie M. Linton ¹ ♥ ♥, ♥ Sung-mok Jung ¹ ♥ ♥, ♥ Baoyin Yuan ¹ ♥, ♥ Ayako Suzuki ¹ ♥ and ♥ Andrei R. Akhmetzhanov ¹♥
- ¹ Graduate School of Medicine, Hokkaido University, Kita 15 Jo Nishi 7 Chome, Kita-ku, Sapporo-shi, Hokkaido 060-8638, Japan
- ² Core Research for Evolutionary Science and Technology, Japan Science and Technology Agency, Honcho 4-1-8, Kawaguchi, Saitama 332-0012, Japan
- ³ Osaka Institute of Public Health, Nakamichi 1-3-69, Higashinari, Osaka 537-0025, Japan
- * Author to whom correspondence should be addressed.

J. Clin. Med. 2020, 9(2), 419; https://doi.org/10.3390/jcm9020419

- Mass testing
- Case count data
- Viral genomic data

Cryptic transmission of novel coronavirus revealed by genomic epidemiology

2 Mar 2020 by Trevor Bedford

- Mass testing
- Case count data
- Viral genomic data



Just for illustrative purposes! -> SARS-CoV2 does not mutate this quickly

What did I find by analyzing viral genomes?



- Branching process model assuming the average time for an infected individual to pass on their infection is 5 days (SD: 1.9 days).
- Data up to 31 January 2020



What did I find by analyzing viral genomes?

- 87% (59%-98%) of infections were undetected based on early pandemic data from Hubei
- 95% (88%-98%) were undetected when accounting for infections globally
- % detected = % of total infections that were reported in the JHU time series (sources include WHO, China CDC, US CDC and other similar organizations)



Infected but Feeling Fine: The Unwitting Coronavirus Spreaders

The C.D.C. director says new data about people who are infected but symptom-free could lead the agency to recommend broadened use of masks.



CZ Biohub and COVID-19

- Nonprofit biomedical research organization in San Francisco, California founded in 2017
- COVID-19 clinical testing lab and SARS-CoV2 viral sequencing capacity
- Data science contributions:
 - Host response
 - Generating whole genome sequences of SARS-CoV2 viruses
 - Identifying co-infecting pathogens



Chan Zuckerberg Biohub @czbiohub · Mar 19 A newly-formed partnership between @czbiohub, @ChanZuckerberg, @UCSF and @CAgovernor aims to increase local #COVID19 testing to 1,000 tests per day within the coming days thru operational support, informatics and protocol development



New Partnerships Expand COVID-19 Clinical Testing CZI, CZ Biohub & UCSF partner to expand local testing, aiming for 1,000 tests a day & chanzuckerberg.com

Acknowledgements

- AWS Diagnostics Development Initiative for supporting the computational analyses
- GISAID for sequence data
- Johns Hopkins CSSE time series data
- CZ Biohub
 - Patrick Ayscue Senior Biosecurity Fellow
 - Data Science team: Jack Kamm, Phoenix Logan, Samantha Hao, David Dynerman, Josh Batson, Aaron McGeever, Olga Botvinnik, Jim Karkanias