

Can the reported COVID-19 data tell us the truth?

Scrutinizing the data from the measurement error models perspective

Grace Y. Yi

Canada Research Chair in Data Science (Tier 1)

Department of Statistical and Actuarial Sciences

Department of Computer Science

University of Western Ontario

Joint work with Wenqing He, Qihuang Zhang, Yayuan Zhu, and GW-DSRG

Outline

- The website of COVID-19 Canada developed by the GW-DSRG (Grace-Wenqing Data Science Research Group)
- Different estimates from different studies
 - A meta-analysis
- Time series with measurement error: sensitivity analysis
 - An additive measurement error model
 - A multiplicative measurement error model

Part 1: The Website of COVID-19 Canada

developed by the GW-DSRG (Grace-Wenqing Data Science Research Group)

<https://covid-19-canada.uwo.ca>

COVID-19 Canada Website: HOME

Western Science

Welcome to the COVID-19 Canada Site

HOME DATA VISUALIZATION STATISTICAL ANALYSIS RESEARCH MANUSCRIPTS
PUBLIC DATA SOURCES SELECTED NEWS STOP VIRUS READER'S CORNER ABOUT US

This is a clickable map that depicts real-time situations about COVID-19 for each province in Canada. Hover your mouse over the map to display a brief summary for the province of your interest. Click that province and you will be directed to another page with detailed analysis results of the province. This may take a few seconds for the program to generate updated real-time analysis results.

Total Confirmed	Total Deaths	Total Recovered
75,853 (+1240)	5,679 (+117)	37,819 (+924)

Last Update: 5/16/2020 19:00 PM

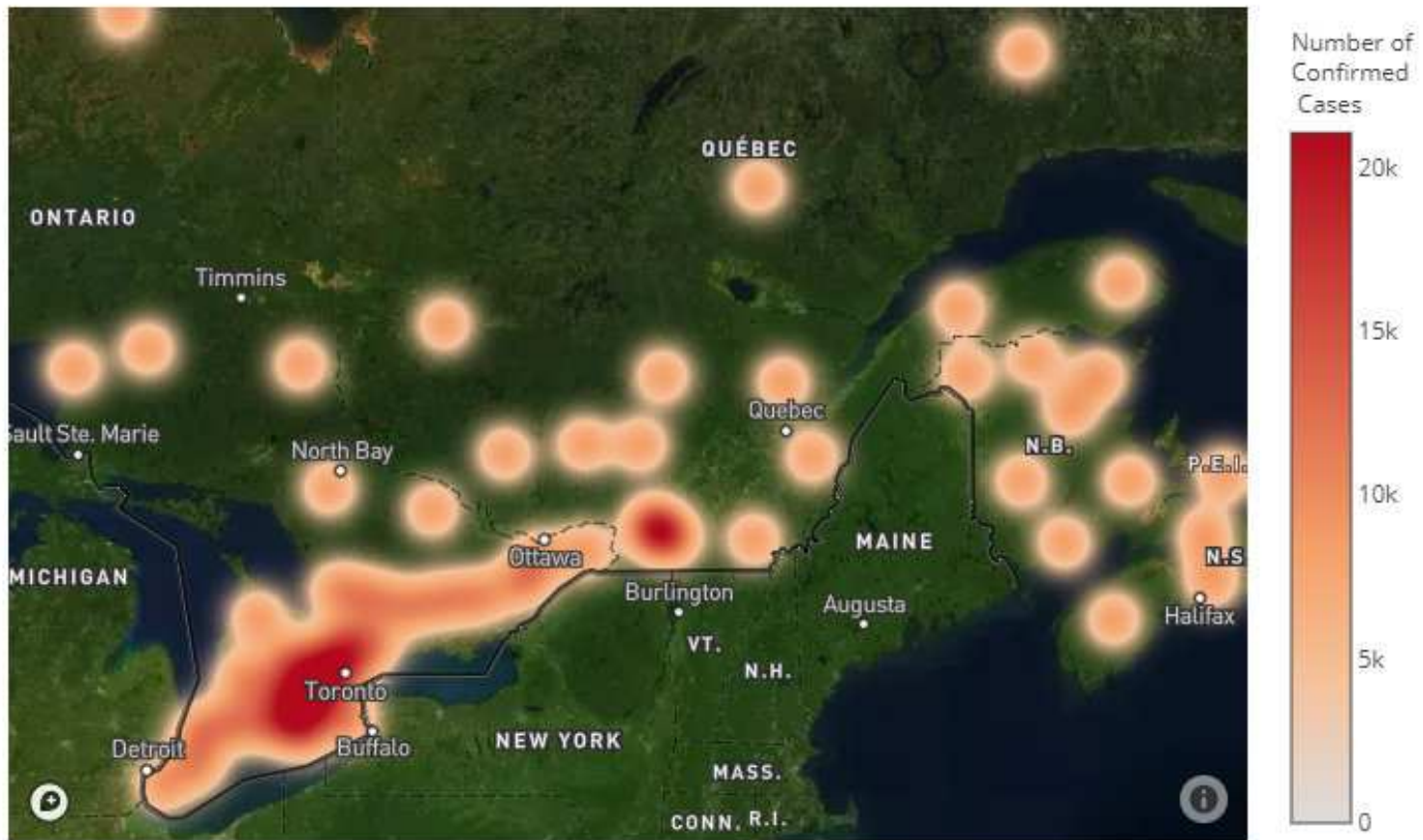
Welcome!

Welcome to the web-page created by the GW-DSRG (Grace-Wenqing Data Science Research Group) at the Faculty of Science of the University of Western Ontario.

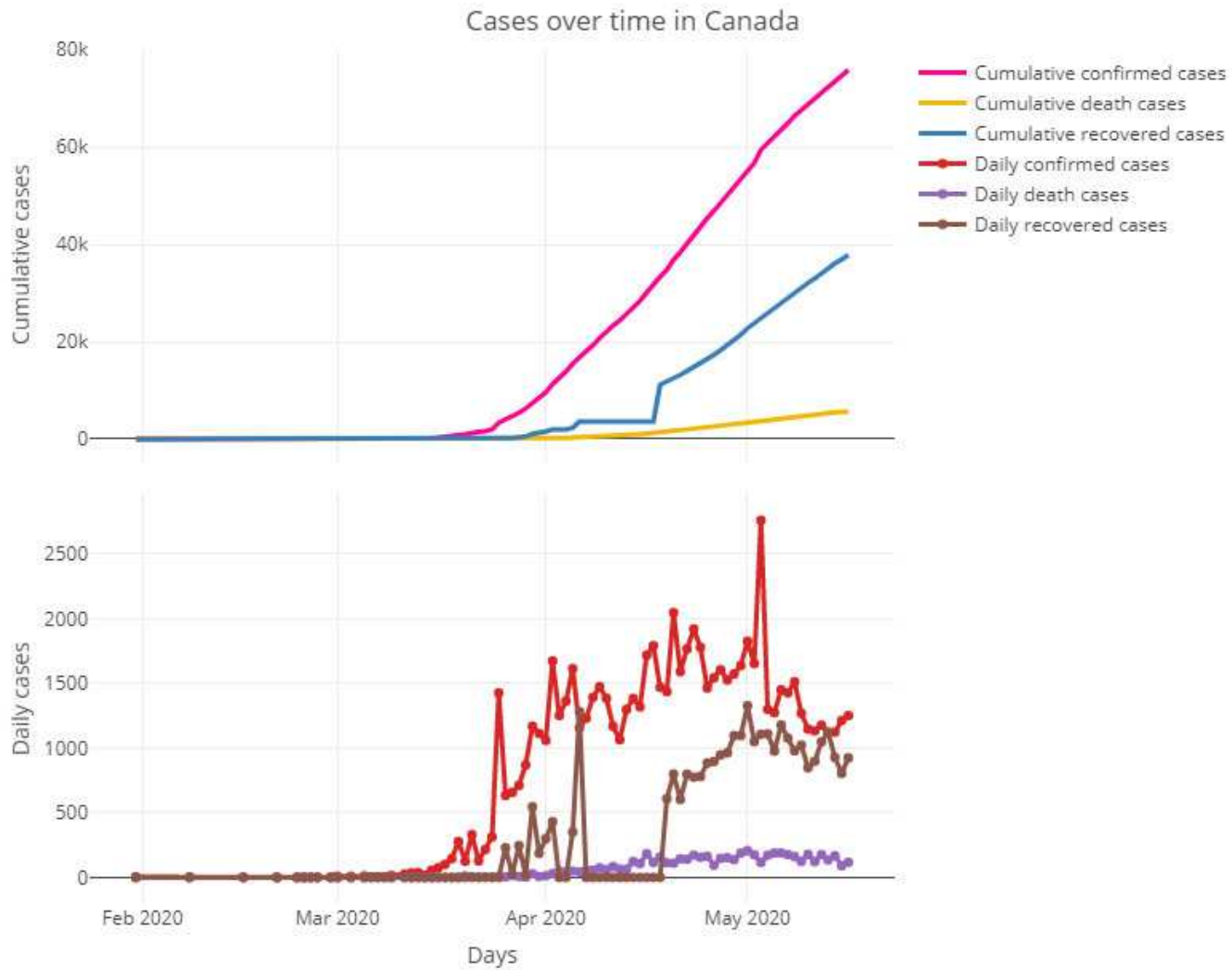
With the focus on Canadian COVID-19 data, this web-based interactive platform aims to report real-time development of COVID-19, visualize patterns, and provide state-of-the-art statistical insight from up-to-date accessible COVID-19 data. This website will be updated on a daily basis.

COVID-19 Canada Website: Heat Map of Canada

Heat map of cumulative confirmed cases by PHU in Canada

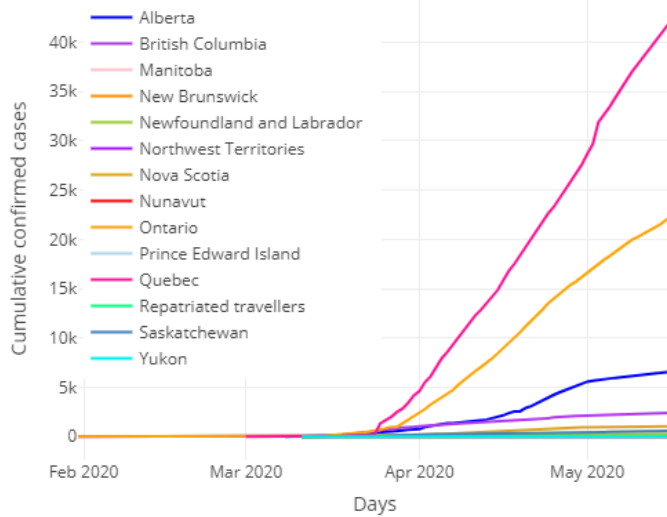


COVID-19 Canada Website: Cases Over Time in Canada

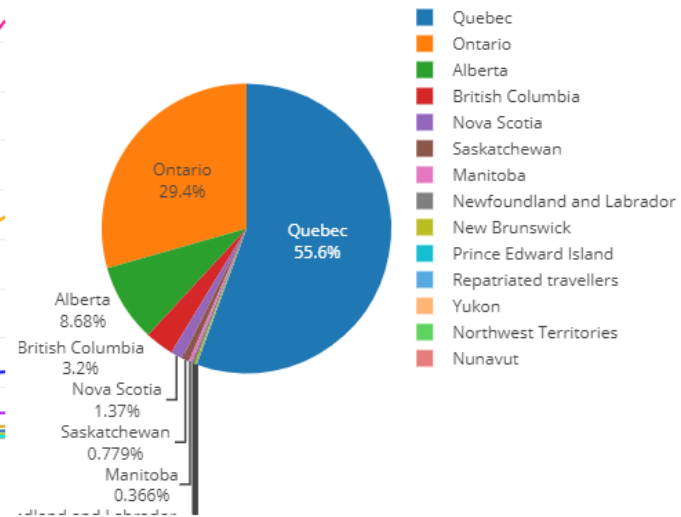


COVID-19 Canada Website: Cases by Province

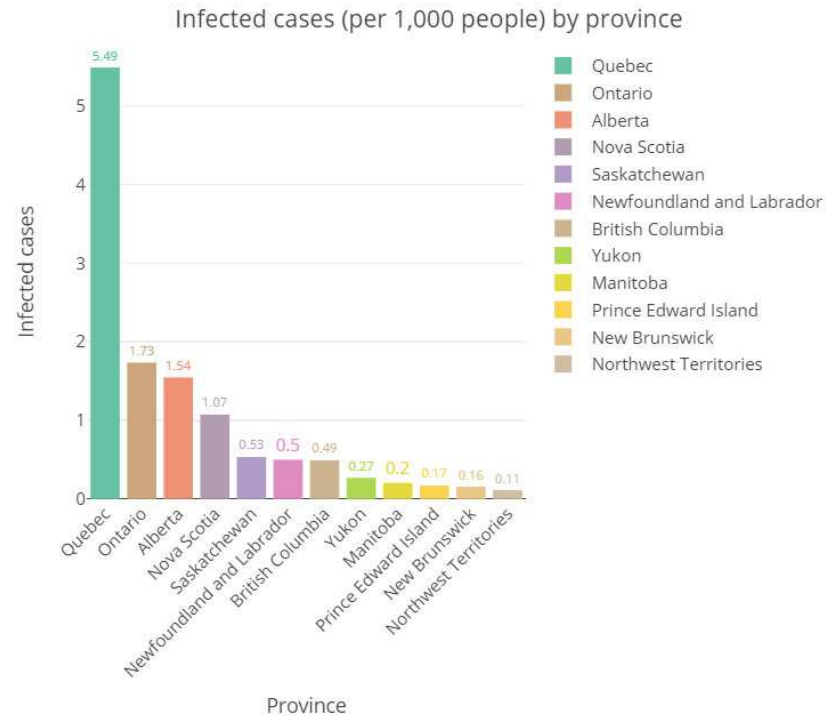
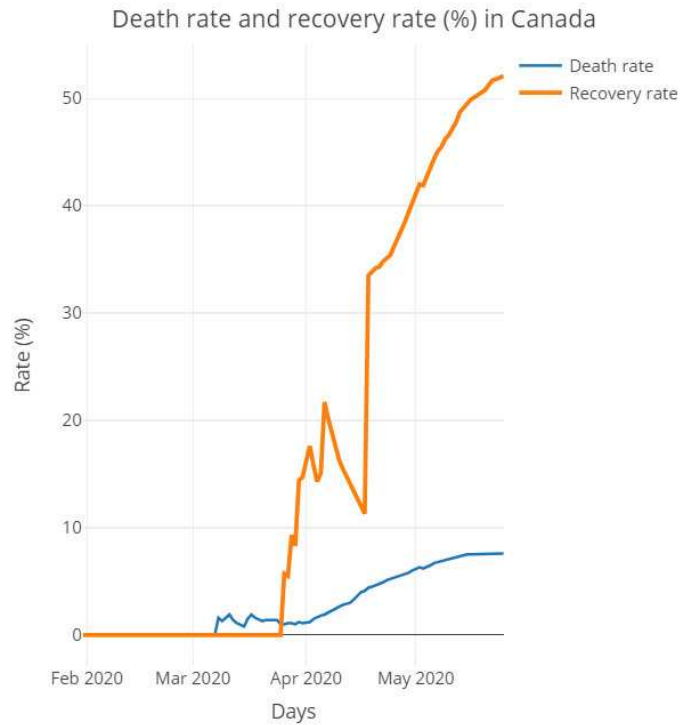
Cumulative confirmed cases by province in Canada



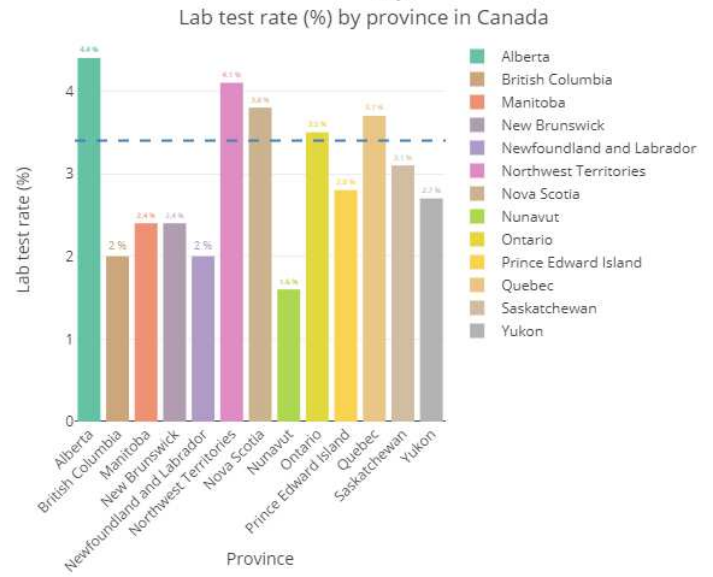
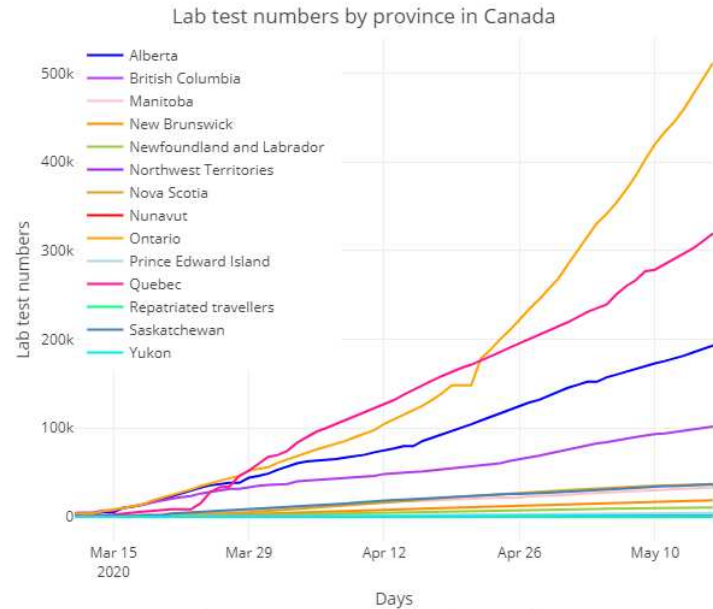
Confirmed cases by province in Canada



COVID-19 Canada Website: Died, Recovered and Infected

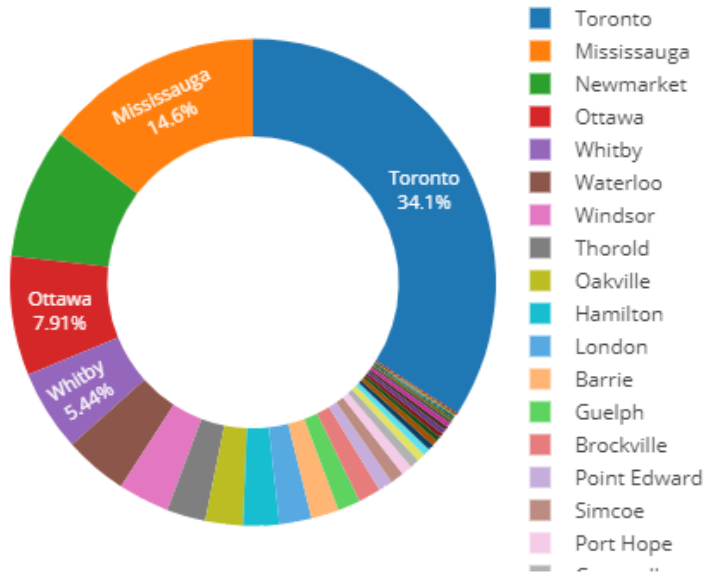


COVID-19 Canada Website: Lab Tests

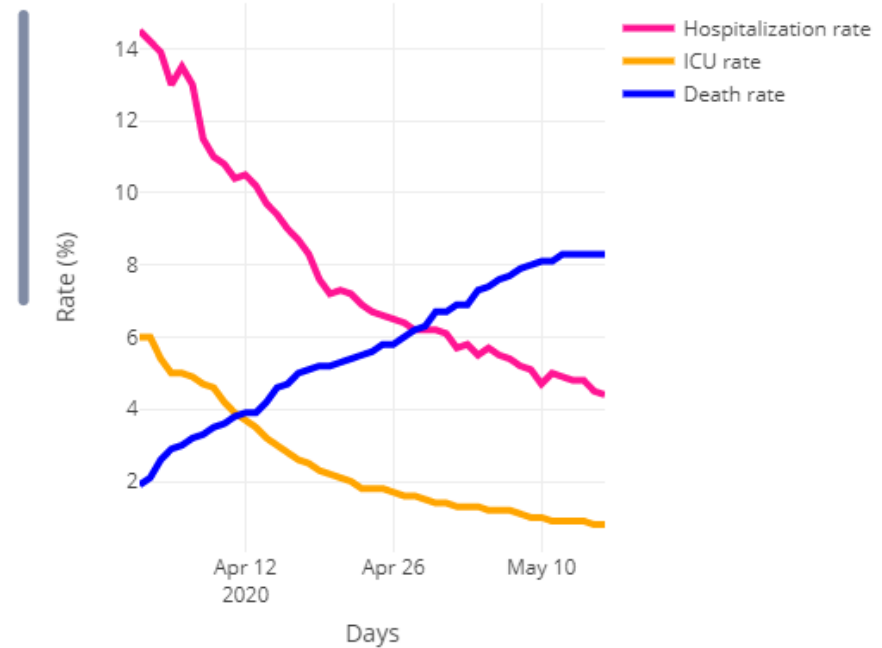


COVID-19 Canada Website: ICU and Hospitalization

Proportion of confirmed cases by city in Ontario

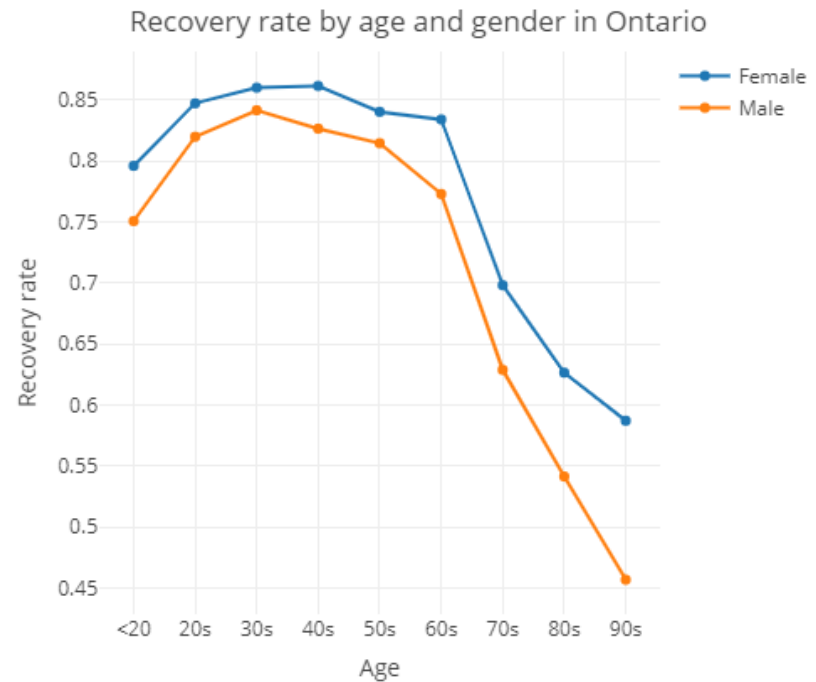
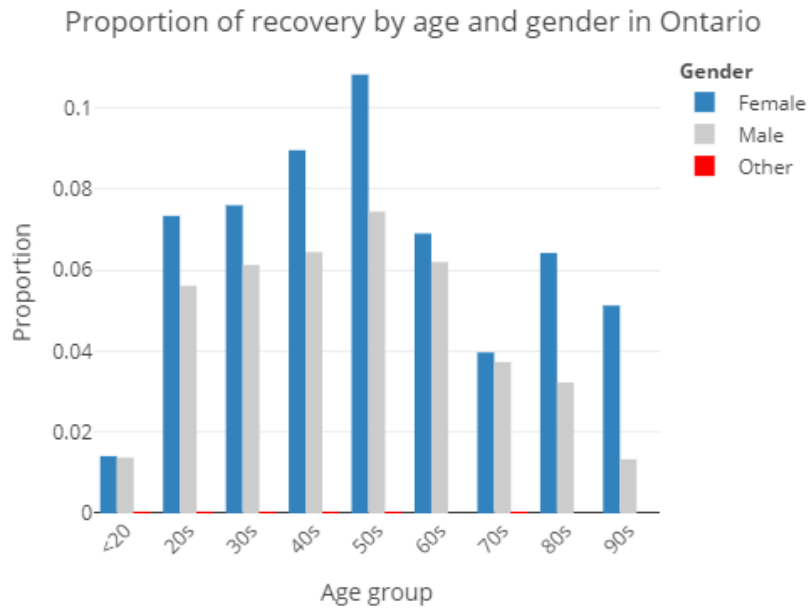


Hospitalization, ICU, death rates (%) over time in Ontario



The hospitalized rate, ICU rate and death rate are computed by the number of people who are hospitalized, placed in ICU, and dead respectively divided by the number of confirmed cases in Ontario. Data comes from Government of Ontario.

COVID-19 Canada Website: Recovery by Age/Gender



Recovery rate for each age and gender group is computed by the number of recovered cases for particular age and gender group divided by number of confirmed cases in this age and gender group.

COVID-19 Canada Website: Prediction

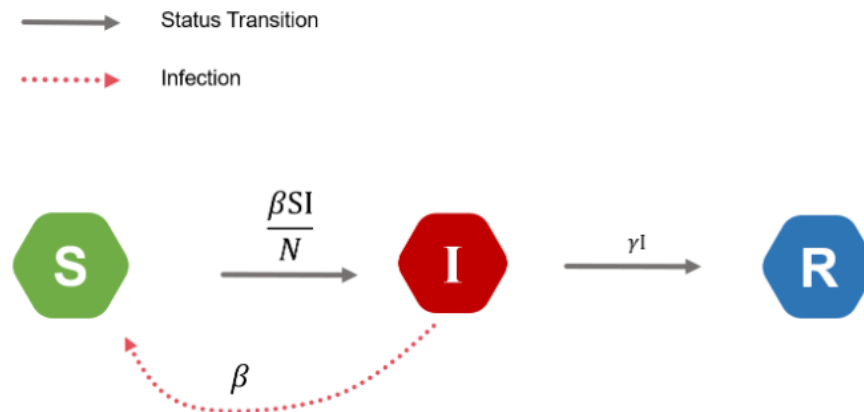
Prediction with SIR Model | Prediction with NN | Regression Analysis

Objective

We implement the SIR (susceptible-infected-resolved) model to predict cumulative numbers of infected cases with COVID-19 in Ontario, British Columbia, Quebec and Alberta. We fit the model using the data from March 18 to May 11, 2020 (<https://coronavirus.1point3acres.com/>) to predict the cases from May 12 to May 18, 2020.

Assumption and Model

The population in each province is divided as three subpopulations defined as *susceptible* (S), *infected* (I), and *resolved* (R), respectively. The status for an individual in the population may change with time: a healthy individual may become infected, and an infected patient may recover or die of the disease. Figure 1 shows the process of the status change, where β represents the average number of contacts per person per time and γ stands for the transition rate from I to R. We estimate both β and γ using the data in each province for the period of March 18 to May 11, 2020, during which all four provinces were in the "state of emergency". We assume that there are no inbound or outbound infected travellers during this period.



The process of status transition between different groups

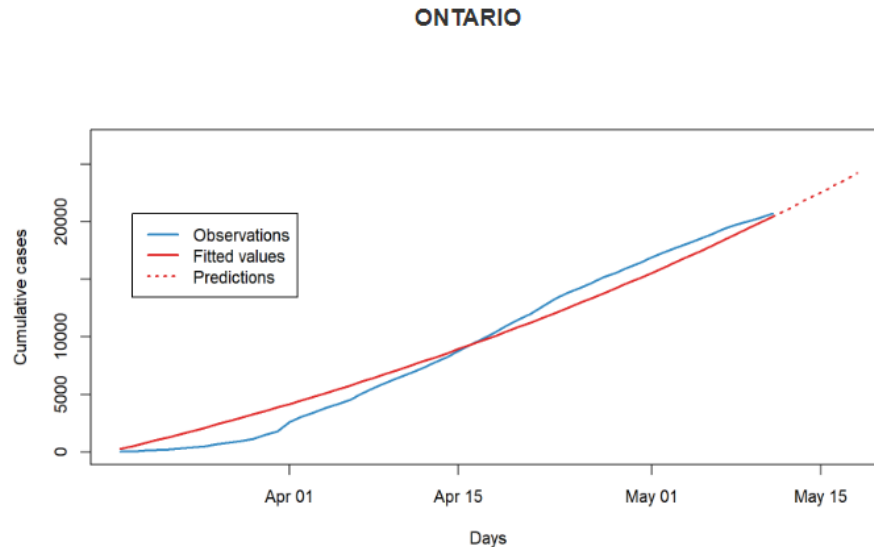
COVID-19 Canada Website: Prediction

Findings and Discussion

The following figures present the predicted cumulative number of cases (in red) together with the reported cumulative number of confirmed cases (in blue) for the four provinces. A red solid curve reports the fitted number for the period of March 18 to May 11, 2020, and its differences from the blue curve show the performance of using the SIR model to do prediction. A red dashed curve is the predicted cumulative number of cases for the period of May 12 to May 18, 2020.

For ease of visualization, here we use connecting curves instead of isolated points to display the reported or predicted cumulative numbers of cases for the four provinces. For the next few days, British Columbia shows a somewhat possibly downward trend, whereas the other three provinces exhibit an upward trend.

The prediction here focuses on showing the trend in the near future using the SIR model. The validity of the results relies on the associated assumptions of the SIR model which may be violated. For instance, unreported cases as well as the inbound and outbound travelers would make the model fitting flawed. Due to the limited testing capacity, the daily reported cumulative number of confirmed cases can differ from the predicted number with a notable discrepancy.



The comparison of the predicted cumulative number of infected cases using the SIR model (in red) versus the reported cumulative infections (in blue) in Ontario. The red dashed curve represents the prediction for the next 7 days.

COVID-19 Canada Website: Team Members

OUR TEAM MEMBERS



Grace Y. Yi



Wenqing He



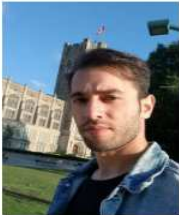
Dan Liu



Yuan Du



Jingyu Cui



Yasin Khadem Charvadeh



Gansen Deng



Li-Pang Chen



Qihuang Zhang



Kaida Cai



Joy He

Part 2: Different Estimates from Different Studies

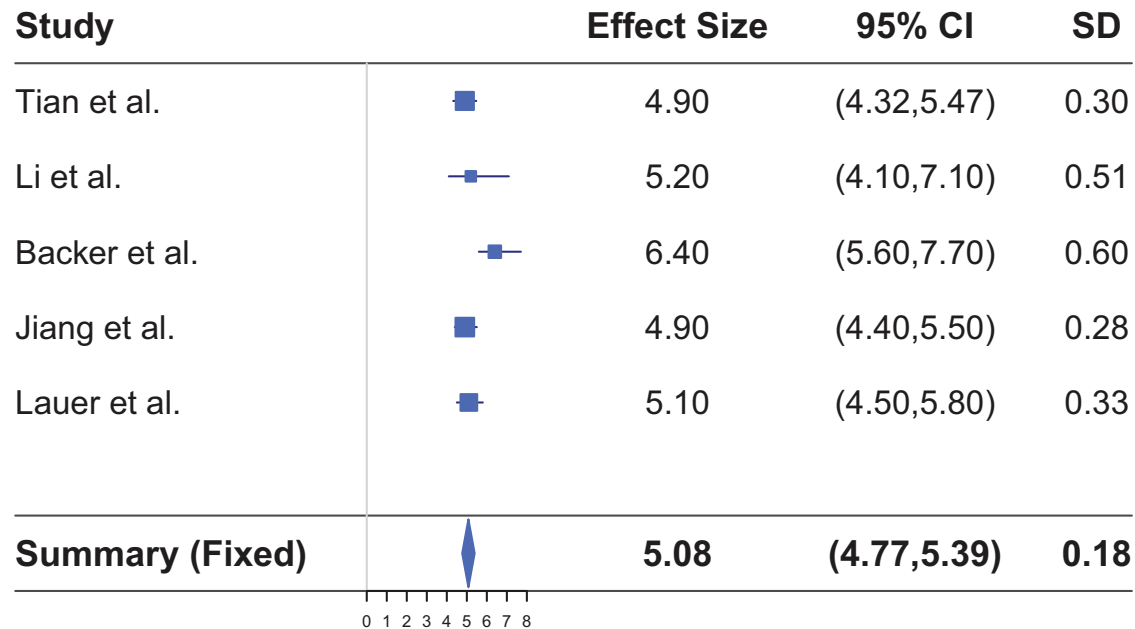
- Meta-Analysis

W He, G Y. Yi, and Y Zhu (2020). Estimation of the basic reproduction number, average incubation time, asymptomatic infection rate, and case fatality rate for COVID-19: Meta-analysis and sensitivity analysis.

To appear in *Journal of Medical Virology*.

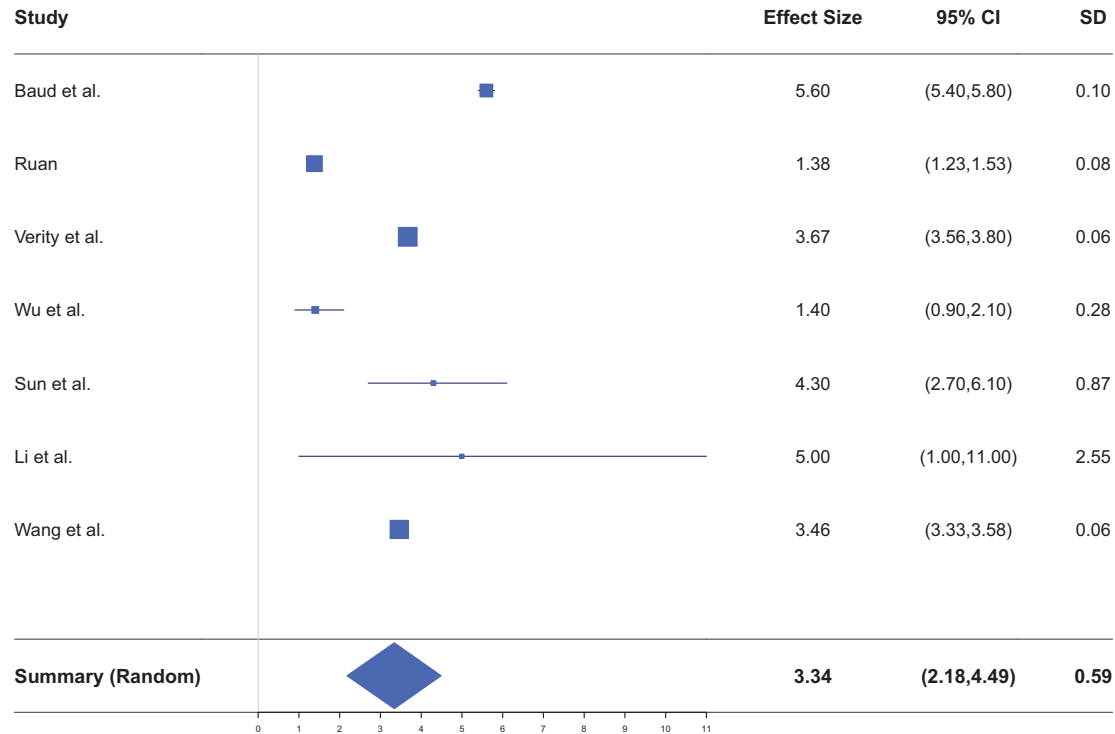
Estimate of Average Incubation Time (in Day)

	Date	Cohort
Tian et al.	12/31/2019 - 02/19/2020	Chinawide
Li et al.	up to 01/22/2020	425 cases in Wuhan
Backer et al.	01/20/2020 - 01/28/2020	88 Wuhan travellers
Jiang et al.	up to 02/08/2020	50 confirmed cases in Wuhan
Lauer et al.	up to 02/24/2020	mainly about China



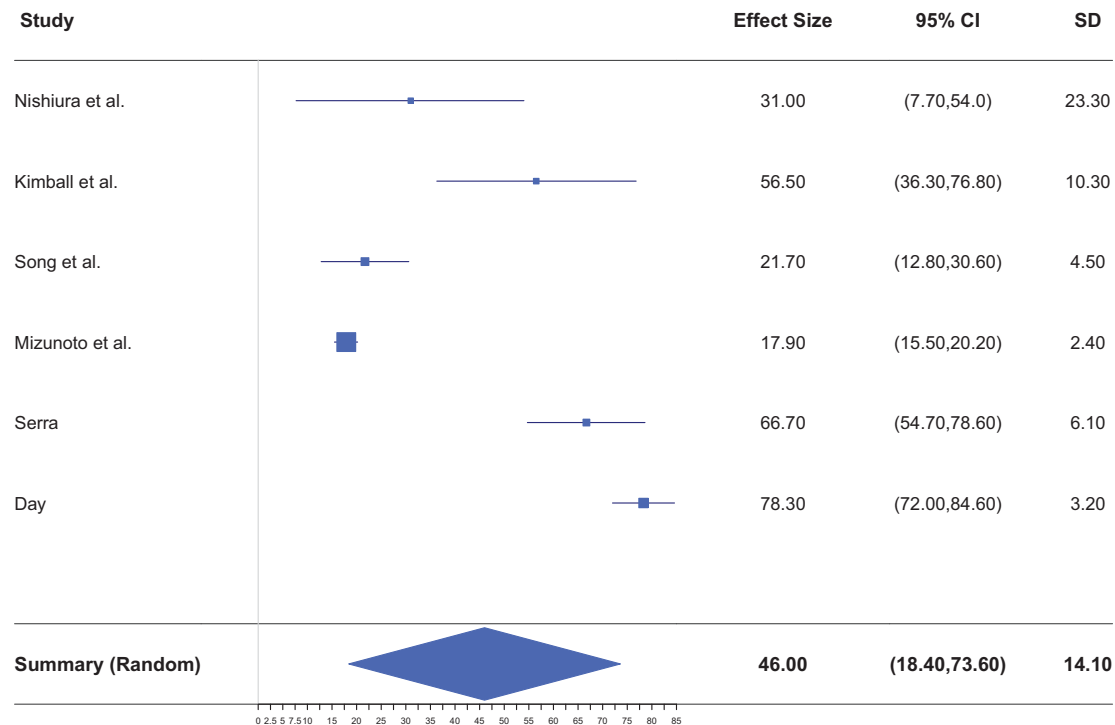
Estimate of the Fatality Rate (in Percent)

	Date	Cohort
Baud et al.	up to 03/01/2020	worldwide
Ruan	up to 03/21/2020	Chinawide
Verity et al.	01/04/2020 - 02/24/2020	outside Hubei province, China
Wu et al.	up to 02/29/2020	Wuhan, China
Sun et al.	a meta analysis with ten studies	50466 cases in China
Li et al.	12/2019 - 02/2020	China
Wang et al.	up to 02/27/2020	worldwide



Estimate of the Asymptomatic Infection Rate (in Percent)

	Date	Cohort
Nishiura et al.	up to 02/06/2020	565 Japanese nationals evacuated from Wuhan
Kimball et al.	03/13-20/2020	13 long-term care residents in King County, Washington
Song et al.	up to 03/06/2020	a single-centre study in Daofu county, Sichuan
Mizumoto et al.	up to 02/21/2020	3,711 people on the Diamond Princess cruise ship
Serra	04/02/2020	Northern Italy, 60 volunteer blood donors
Day	04/01/2020	166 new infections



Part 3: Error-Prone COVID-19 Data

Some Error Sources of COVID-19 Data

Error-Contaminated COVID-19 Data

Confirmed cases may be under-reported:

Asymptomatic patients

- may not be detected
- may be delayed to be found

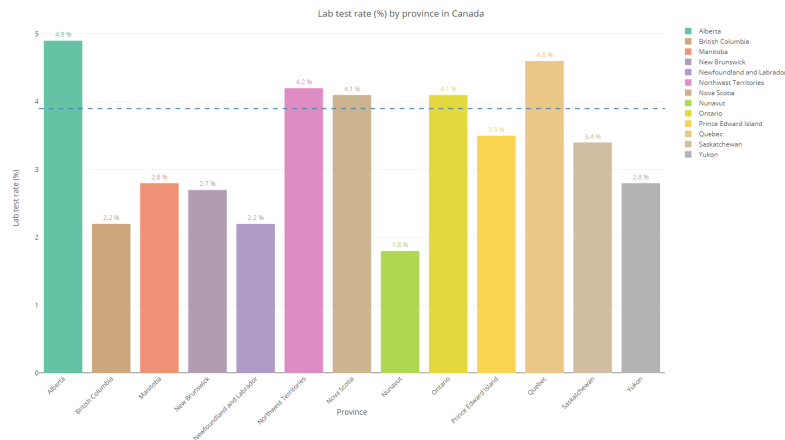
e.g.,

Miszunoto et al. (2020): the asymptomatic infection rate $\approx 17.90\%$

Day (2020): the asymptomatic infection rate $\approx 78.30\%$

Limited test capacity does not allow infected patients with light symptoms to be tested.

e.g., As of May 25, 2020



Error-Contaminated COVID-19 Data

- Confirmed cases or discharged recoveries may be error-prone due to testing error:

- The accuracy of the current COVID-19 tests is not precisely known.

Example:

Based on the test performance in China and the performance of the influenza tests, Hutchison (2020) suggested that the sensitivity and specificity of COVID-19 tests were estimated to be 60% and 90%, respectively.

- Example:

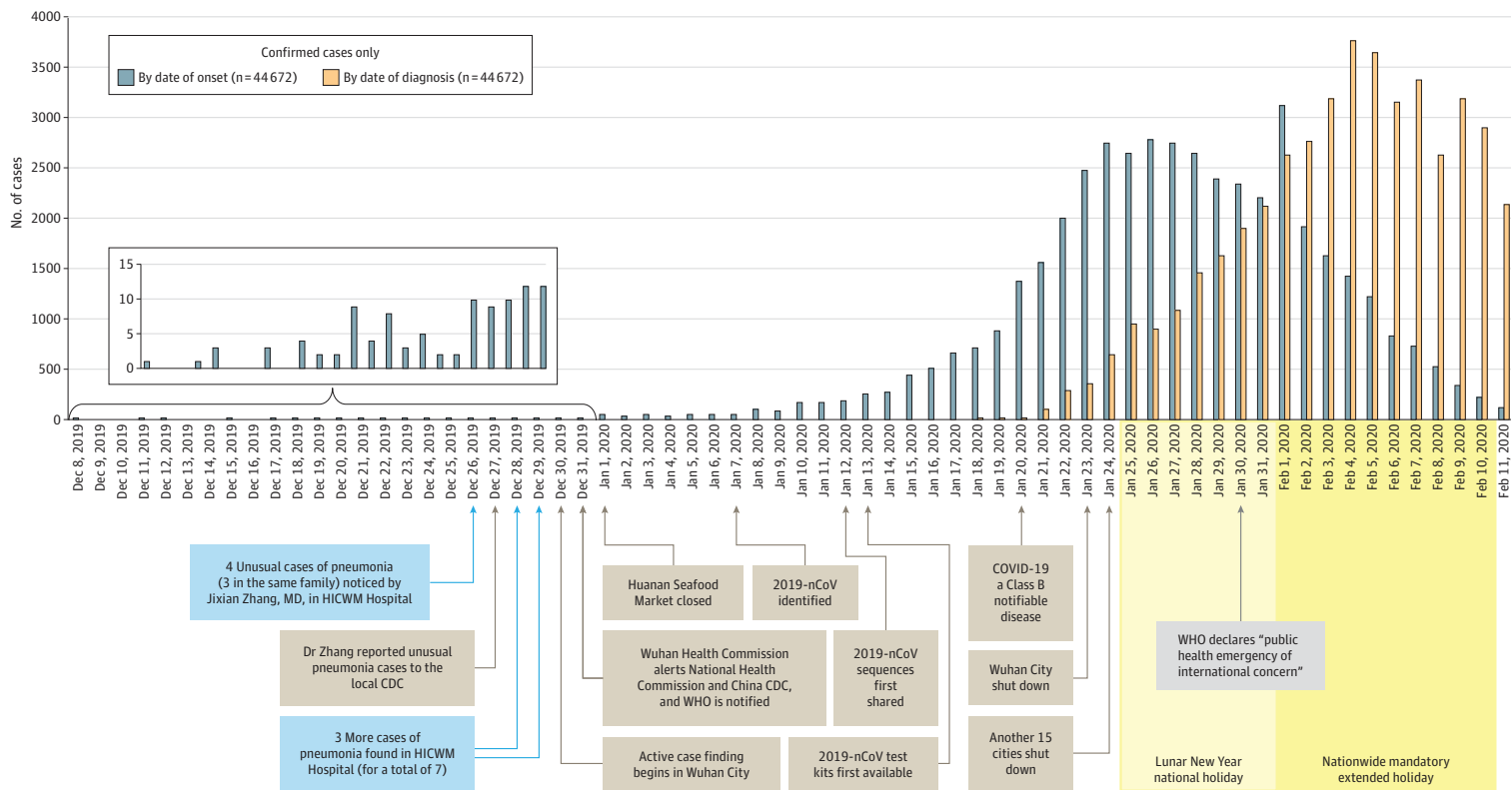
In China's Guangdong province, 14% of people who recovered in the province were later retested to be positive. Similar cases have been reported in other countries (Guzman 2020).

Error-Contaminated COVID-19 Data

 The incubation time varies from patient to patient.

Source: Wu and McGoogan (2020)

Figure 1. Epidemic Curve of the Confirmed Cases of Coronavirus Disease 2019 (COVID-19)



Opinion Viewpoint

Daily numbers of confirmed cases are plotted by date of onset of symptoms (blue) and by date of diagnosis (orange). Because, on retrospective investigation, so few cases experienced illness in December, these cases are

symptoms in December were only discovered when retrospectively investigated. Major epidemic response actions taken by the Chinese government are shown in brown boxes. The normally scheduled Lunar New Year

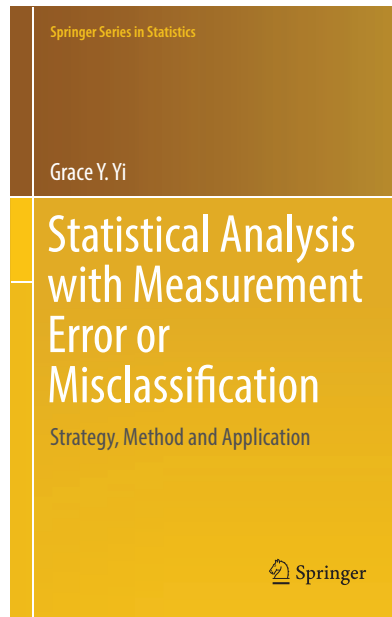
Error-Contaminated COVID-19 Data

- **Heterogeneity:**

Since the outbreak of the disease, a large body of research on COVID-19 has been done and many articles have been published in scientific journals or shared on platforms such as bioRxiv and medRxiv.

 - **Different studies** have been carried out
 - on different patients
 - under different conditions,
 - and different authors may make different model assumptions.
 - **Simulations** of the epidemic have been published under **various assumptions** to delineate hidden transmissions of the virus.
- Interpretation of the available findings must be coupled with the associated features of the studies.

General Framework of Data with Measurement Error



Some Sources of Measurement Error (Yi 2017)

- Measurement error may refer to **random noise, sampling error**, or **uncertainty/variation** in the measuring process.

Measurement Error = reading error + biological variability + sampling error
reporting error + detecting error + others

- **Flawed** or **mismanaged** data collection procedures result in imprecise measurements.
- Variables are not accurately measured due to **reporting errors** for sensitive questions.
- Variables are not accurately measured due to **recall bias**.
- Variables are **impossible** to measure precisely.
- Variables represent **averages** of certain quantities over time.
- Variables may be **manipulated** artificially.
- Variables are too **expensive** or time consuming to measure precisely.

Impact of Measurement Error

Remarks (Carroll et al. 2006; Yi 2017)

- Measurement error in variables may
 - change the structure of the response model
 - cause bias in parameter estimation
 - lead to a loss of power for detecting interesting relationship among variables
 - mask the features of the data

- The effects of measurement error are very complex, depending on the form of
 - the inference method
 - the measurement error model
 - the response model
 - the association of the covariates

Part 4: Time Series Data with Measurement Error

- Sensitivity Analysis

Objectives and Challenges

Objectives

- Model the daily quantities of interest
 - daily confirmed cases
 - daily deaths/fatality rates
 - daily recoveries
- Forecast the trends using historical data

Challenges

- **Reported data are error-contaminated** due to multiple reasons including
 - limited test capacity / shortage of healthcare personnel
 - undetected asymptomatic infections / unreported infections with light symptoms
 - delayed detection due to the incubation period
 - reporting errors and recall bias
- It is **difficult to characterize the measurement error process** due to
 - unknown error form
 - unknown error degree

Time Series Data in Error-Free Settings

Autoregressive Model

Notation

discrete time series: $\{X_t : t = 1, \dots, T\}$

AR(p) Model

$$X_t = \phi_0 + \sum_{j=1}^p \phi_j X_{t-j} + \epsilon_t$$

where

p is an integer smaller than T

the ϵ_t are independent of each other and of the X_t

each ϵ_t has zero mean and variance σ_ϵ^2

Interest

estimation of $\phi = (\phi_1, \dots, \phi_p)^\top$ and ϕ_0

prediction for a short period after T

Remark

The parameters are constrained for stationary time series.

Least Squares Method

Estimation of $\phi = (\phi_1, \dots, \phi_p)^\top$

• Let $S(\phi) = \sum_{t=p+1}^T \{X_t - (\phi_0 + \sum_{j=1}^p \phi_j X_{t-j})\}^2$

• Estimator:

$$\begin{aligned}\hat{\phi}^{(LS)} &= \operatorname{argmin}_{\phi} S(\phi) \\ &= \left(\sum_{t=p+1}^T \left\{ \tilde{X}_{t-1} - E(\tilde{X}_{t-1}) \right\} \left\{ \tilde{X}_{t-1} - E(\tilde{X}_{t-1}) \right\}^\top \right)^{-1} \\ &\quad \times \sum_{t=p+1}^T \left\{ \tilde{X}_{t-1} - E(\tilde{X}_{t-1}) \right\} \{X_t - E(X_t)\},\end{aligned}$$

where $\tilde{X}_{t-1} = (X_{t-1}, \dots, X_{t-p})^\top$ and $E(X_t)$ is estimated by $\hat{\mu} = \frac{1}{T} \sum_{t=1}^T X_t$

• $\hat{\phi}_0^{(LS)} = \hat{\mu}(1 - \sum_{j=1}^p \hat{\phi}_j)$

• $\hat{\sigma}_\epsilon^{2(LS)} = \frac{1}{T-p} S(\hat{\phi})$

Estimating Equation

Notation

$\gamma_k = \text{Cov}(X_t, X_{t-k})$ for $t > p$ and $k = 0, \dots, p$

$\gamma = (\gamma_1, \dots, \gamma_p)^\top$

Γ : the $p \times p$ autocovariance matrix

Alternative Method

Estimators $\hat{\phi}$, $\hat{\phi}_0$ and $\hat{\sigma}_\epsilon^2$ solve

$$\begin{aligned}\phi &= \hat{\Gamma}^{-1} \hat{\gamma}; \\ \phi_0 &= \left(1 - \sum_{i=1}^p \phi_i\right) \hat{\mu}; \\ \sigma_\epsilon^2 &= \hat{\gamma}_0 - 2\phi^\top \hat{\gamma} + \phi^\top \hat{\Gamma} \phi\end{aligned}$$

Remark



Asymptotic equivalence:

$$\hat{\phi} - \hat{\phi}^{(\text{LS})} \xrightarrow{p} 0, \hat{\phi}_0 - \hat{\phi}_0^{(\text{LS})} \xrightarrow{p} 0 \text{ and } \hat{\sigma}_\epsilon^2 - \hat{\sigma}_\epsilon^{2(\text{LS})} \xrightarrow{p} 0 \text{ as } T \rightarrow \infty$$



This offers a unified estimation framework in its connections with the least squares estimation, the maximum likelihood method with the assumption of Gaussian error, and the Yule-Walker method.

Measurement Error Models

Measurement Error Models

Notation

X_t : the **true** variable of interest on day t , which is **unobserved**

X_t^* : the **reported** value of X_t

Goal

use the **observed** $\{X_t^* : t = 1, \dots, T\}$ to understand / infer the **true** process of $\{X_t : t = 1, \dots, T\}$

Additive Measurement Error Model

Additive Measurement Error Model

$$X_t^* = \alpha_0 + \alpha_1 X_t + e_t$$

for $t = 1, \dots, T$, where

e_t : independent of each other as well as of X_t ; has mean 0 and variance σ_e^2

α_0 represents the systematic error

α_1 represents the constant inflation (or shrinkage) due to the measurement error

Remark

- If $\alpha_0 = 0$ and $\alpha_1 < 1$, the measurement error process implies that X_t^* tends to be smaller than X_t if the noise term is ignored.
- If $\alpha_0 = 0$ and $\alpha_1 > 1$, the measurement error process implies that X_t^* tends to be greater than X_t if the noise term is ignored.
- This model generalizes the classical additive model considered by Staudenmayer & Buonaccorsi (2005) who considered the case with $\alpha_0 = 0$ and $\alpha_1 = 1$

Multiplication Measurement Error Model

Multiplicative Measurement Error Model

$$X_t^* = \beta_0 u_t X_t$$

for $t = 1, \dots, T$, where

β_0 : a scaling parameter

u_t : independent of each other and of the X_t ; has mean 1 and variance σ_u^2

Remark

- These models are examples of describing the discrepancies between X_t and X_t^* .
- Other flexible models may be considered. For example,

- **two-piece additive model:**

$$X_t^* = \alpha_{10} + \alpha_{11} X_t + e_{1t} \quad \text{for } t = 1, \dots, T_1$$

$$X_t^* = \alpha_{20} + \alpha_{21} X_t + e_{2t} \quad \text{for } t = T_1 + 1, \dots, T$$

where T_1 represents the day on which measures of curbing the virus start

- **mixture of additive and multiplicative models:**

$$X_t^* = \alpha_0 + \beta_0 u_t X_t + e_t \quad \text{for } t = 1, \dots, T$$

Bias Analysis for Naive Estimators

Bias Analysis for the AR(p) model

Naive Analysis

If naively replacing X_t in the AR(p) model with X_t^* , then the working model is

$$X_t^* = \phi_0^* + \sum_{j=1}^p \phi_j^* X_{t-j}^* + \epsilon_t^*$$

where

$(\phi_0^*, \phi_1^*)^T$ and ϵ_t^* show possible difference from the corresponding quantity in the AR(p) model.

Naive Estimator

$\widehat{\phi}_1^*$ and $\widehat{\phi}_0^*$ are the least squares estimators of ϕ_1^* and ϕ_0^*

Bias Analysis with AR(1) Model

Theorem 0 Assume that the times series $\{X_t : t = 1, \dots, T\}$ is **AR(1)** stationary. Then

$$\widehat{\phi}_1^* \xrightarrow{p} \phi_1^* \text{ and } \widehat{\phi}_0^* \xrightarrow{p} \phi_0^* \text{ as } T \rightarrow \infty$$

where

(a) Under the additive measurement error model:

$$\phi_1^* = \phi_1 \omega_1 \text{ and } \phi_0^* = \left(\alpha_0 + \frac{\alpha_1 \phi_0}{1 - \phi_1} \right) (1 - \phi_1 \omega_1)$$

$$\text{with } \omega_1 = \left\{ 1 + \frac{\sigma_e^2 (1 - \phi_1^2)}{\alpha_1^2 \sigma_\epsilon^2} \right\}^{-1}$$

(b) Under the multiplicative measurement error model:

$$\phi_1^* = \phi_1 \omega_2 \text{ and } \phi_0^* = \frac{\beta_0 \phi_0}{1 - \phi_1} (1 - \omega_2 \phi_1)$$

$$\text{with } \omega_2 = \left\{ 1 + \sigma_u^2 + \frac{\sigma_u^2 \phi_0^2}{\sigma_\epsilon^2} \frac{1 + \phi_1}{1 - \phi_1} \right\}^{-1}$$

Bias Analysis with AR(p) Model with $p \geq 2$

Theorem 0 Assume that the times series $\{X_t : t = 1, \dots, T\}$ is AR(p) stationary with $p \geq 2$. Then

$$\widehat{\phi}_1^* \xrightarrow{p} \phi_1^* \text{ and } \widehat{\phi}_0^* \xrightarrow{p} \phi_0^* \text{ as } T \rightarrow \infty$$

where

(a) Under the additive measurement error model:

$$\phi^* = \alpha_1^2 (\alpha_1^2 \Gamma + \sigma_e^2 I_p)^{-1} \gamma$$

$$\phi_0^* = (1 - \phi^* \mathbf{1}_p) (\alpha_0 + \alpha_1 \mu)$$

(b) Under the multiplicative measurement error model:

$$\phi^* = \{\Gamma + \sigma_u^2 (\gamma_0 + \mu^2) I_p\}^{-1} \gamma$$

$$\phi_0^* = \beta_0 (1 - \phi^{*\top} \mathbf{1}_p) \mu$$

Adjustment of Measurement Error Effects

Adjustment of Measurement Error Effects

Basic Idea

find $\tilde{\mu}$ and the $\tilde{\gamma}_k$ which are expressed in terms of the observed data X_t^* such that

$\tilde{\mu}$ and $\hat{\mu}$ have the same limit in probability

$\tilde{\gamma}_k$ and $\hat{\gamma}_k$ have the same limit in probability for $k = 0, \dots, p$

Estimating Equation

Estimators $\tilde{\phi}, \tilde{\phi}_0$ and $\tilde{\sigma}_\epsilon^2$ are obtained by solving

$$\phi = \tilde{\Gamma}^{-1} \tilde{\gamma}$$

$$\phi_0 = \left(1 - \sum_{i=1}^p \phi_i\right) \tilde{\mu}$$

$$\sigma_\epsilon^2 = \tilde{\gamma}_0 - 2\phi^T \tilde{\gamma} + \phi^T \tilde{\Gamma} \phi$$

Theorem 0 *Assume regularity conditions. Then as $T \rightarrow \infty$*

(1) $\tilde{\phi} \xrightarrow{p} \phi, \tilde{\phi}_0 \xrightarrow{p} \phi_0, \text{ and } \tilde{\sigma}_\epsilon^2 \xrightarrow{p} \sigma_\epsilon^2;$

(2) $\sqrt{n}(\tilde{\phi} - \phi) \xrightarrow{d} N(0, GQG^T),$

where G is the matrix of derivatives of $\tilde{\phi}$ with respect to the components of $(\hat{\gamma}_0^*, \hat{\gamma}^{*T})^T$, and the form of Q depends on the measurement error model.

Adjustment of Measurement Error Effects

Forecasting

$$\widehat{X}_{T+h} = \hat{\phi}_0 + \hat{\phi}_1 \widehat{X}_{T+h-1} + \dots + \hat{\phi}_p \widehat{X}_{T+h-p}$$

with \widehat{X}_j estimated from the measurement error model

Prediction Error

the h -step prediction error with the AR(1) model is

$$E\{(\widehat{X}_{T+h} - X_{T+h})^2\}$$
$$= \begin{cases} \frac{\phi_1^{2h} \sigma_e^2}{\alpha_1^2} + \frac{1-\phi_1^{2h}}{1-\phi_1^2} \sigma_e^2 & \text{for the additive error model} \\ \phi_1^{2h} \left(\frac{\sigma_e^2}{1-\phi_1^2} + \frac{\phi_0^2}{(1-\phi_1)^2} \right) \sigma_u^2 + \frac{1-\phi_1^{2h}}{1-\phi_1^2} \sigma_e^2 & \text{for the multiplicative error model} \end{cases}$$

Remark

- These procedures are developed under the assumption that the measurement error model and the associated parameters are **known**.
- This assumption is **usually untrue** in applications. But the developed procedures are useful for conducting **sensitivity analyses** to understand the impact of measurement error on inference results.

Sensitivity Analysis for the Fatality Rate in Canadian Provinces: Ontario and British Columbia

Uncertainty in Defining the Fatality Rate

Definition 1: $Y_t = \frac{\text{the \# of deaths on day } t}{\text{the \# of cases up to day } t}$

Definition 2: $Y_t = \frac{\text{the \# of deaths on day } t}{\text{the \# of cases up to day } t-10}$

Definition 3: $Y_t = \frac{\text{the \# of deaths on day } t}{\text{the \# of cases up to day } t-14}$

Remark

- Definition 1 is conventionally used.
- Definition 2 is proposed based on the fact that the median time from onset of symptoms to ICU admission is around 10 days (Huang et al. 2020)
- Definition 3 is taken by Baud et al. (2020) with the consideration that the maximum incubation period is assumed to be up to 14 days.

Source: Baud et al. (2020)

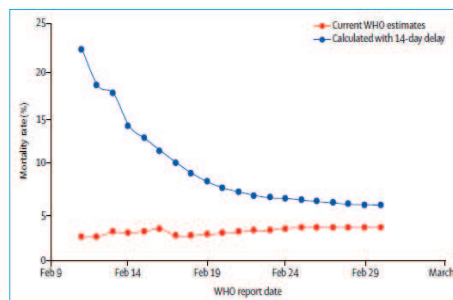
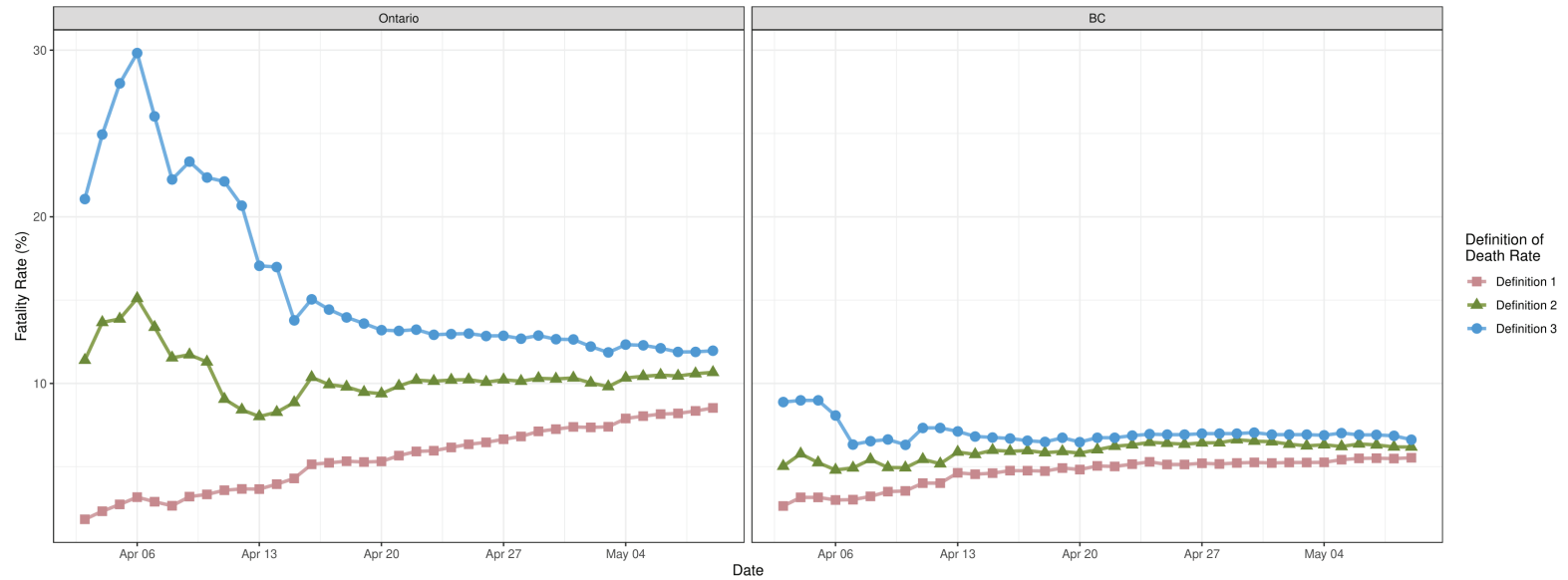


Figure: Global COVID-19 mortality rates (Feb 11 to March 1, 2020)
Current WHO mortality estimates (total deaths divided by total confirmed cases), and mortality rates calculated by dividing the total number of deaths by the total number of confirmed cases 14 days previously.

Comparison of the Fatality Rate from Different Definitions

Fatality Rate of Ontario and British Columbia (BC)



Sensitivity Analysis of the Fatality Rate

Data

We use the COVID-19 data in British Columbia and Ontario containing the daily confirmed cases and deaths from April 10 to May 4, 2020.

Goal

forecast the fatality rate for May 5-9 of 2020 in British Columbia and Ontario

Sensitivity Analysis

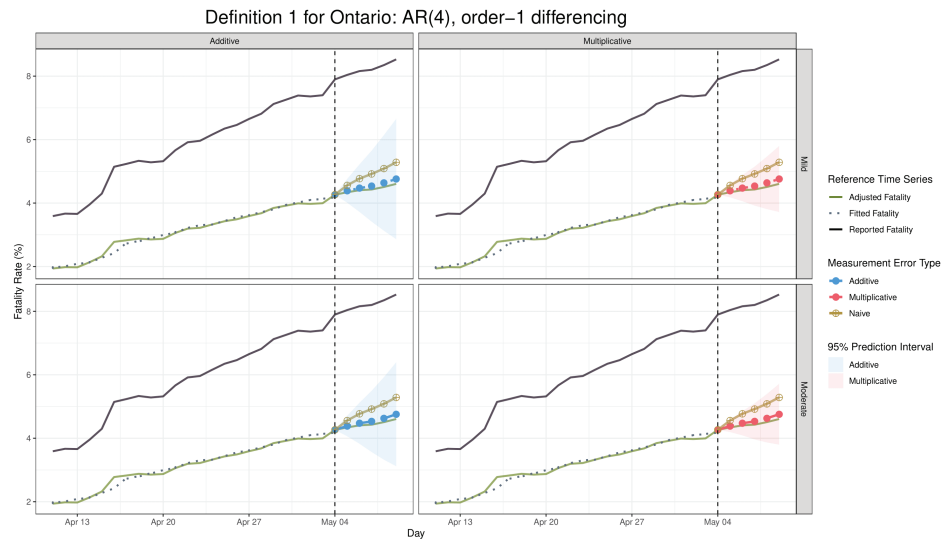
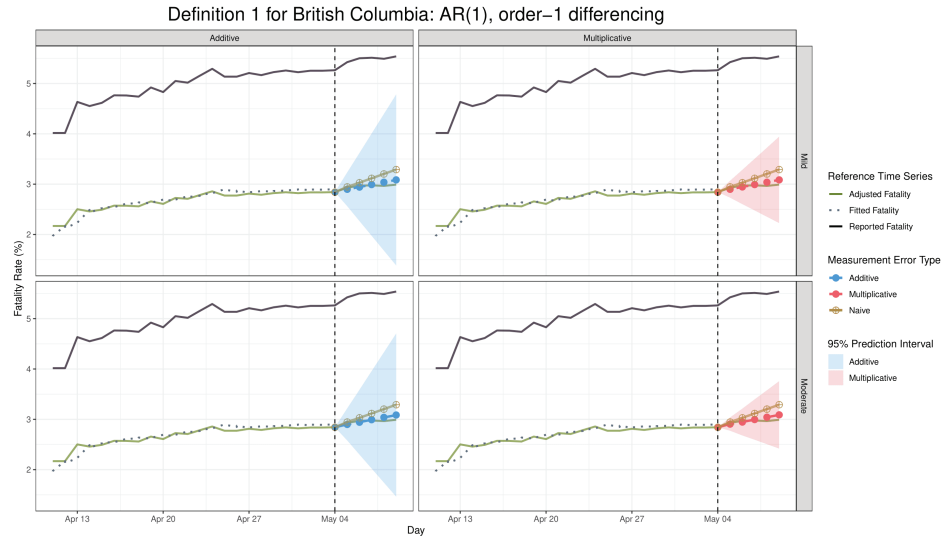
- Starting Point:
If under-reported cases are only caused from undetected asymptomatic cases, then

$$Y_{t-i} = (1 - \tau)Y_t^* \quad \text{for } i = 0, 10, 14$$

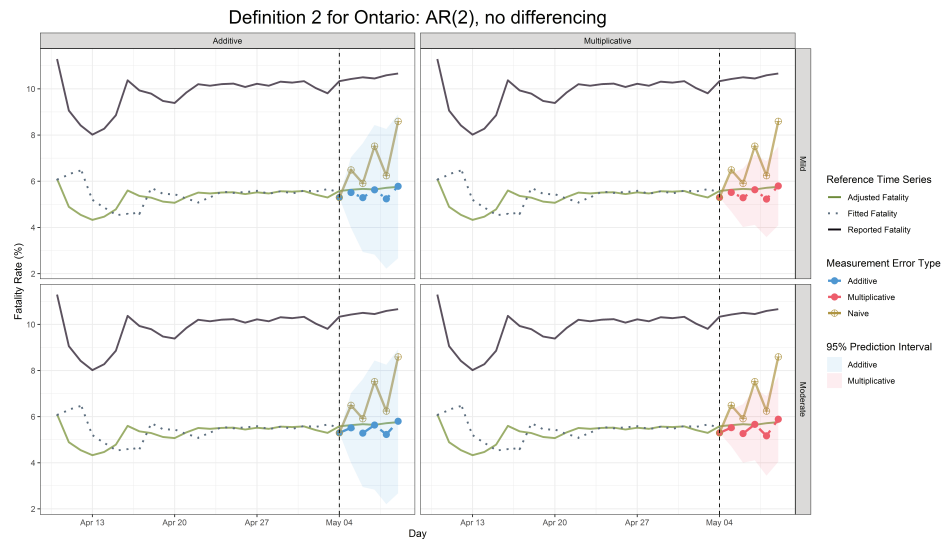
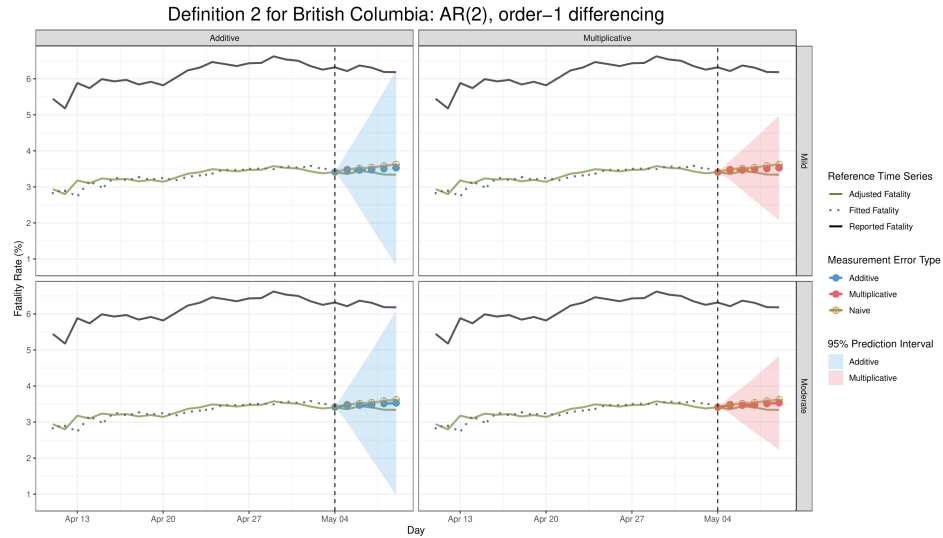
where τ represents the average rate of asymptomatic infections; set $\tau = 46\%$ as the estimate from the meta-analysis of He, Yi and Zhu (2020).

- Multiplicative Model: $\beta_0 = \frac{1}{1-\tau}$; $\sigma_u^2 = 0.2^2$ or 0.5^2
- Additive Model: $\alpha_0 = 0$ and $\alpha_1 = \frac{1}{1-\tau}$; $\sigma_e^2 = 0.1^2$ or 0.3^2
- Note: Y_t may be transformed by “differencing” to make the series (nearly) stationary.

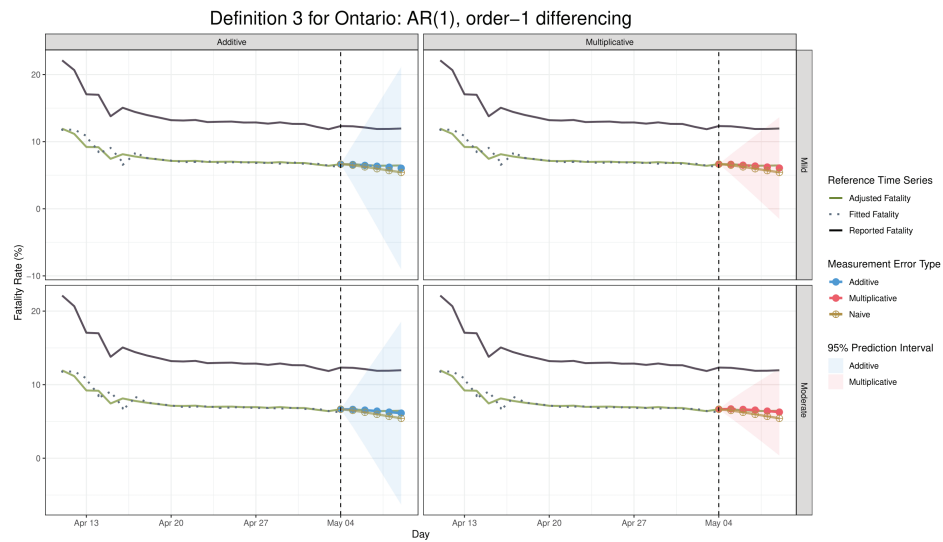
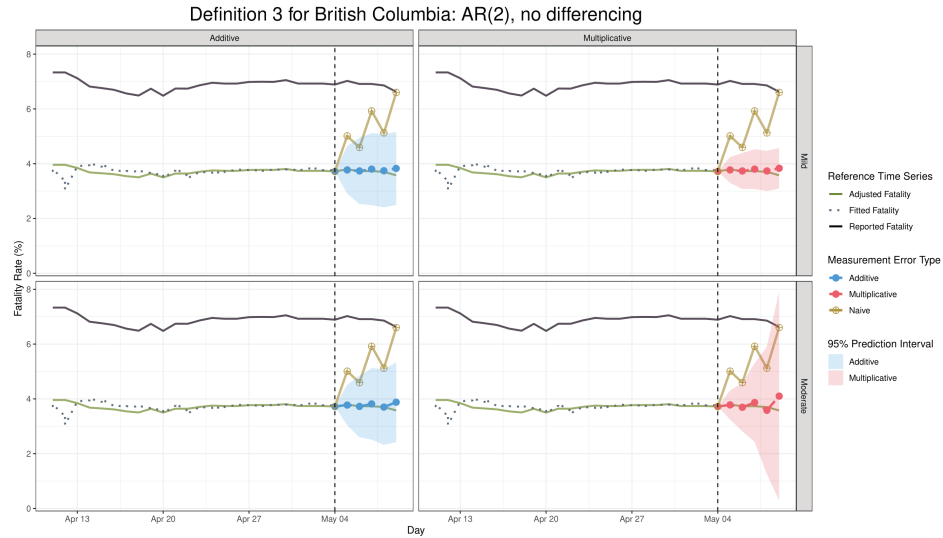
Sensitivity Analysis: Definition 1 - BC vs ON



Sensitivity Analysis: Definition 2 - BC vs ON



Sensitivity Analysis: Definition 3 - BC vs ON



Concluding Remarks

Take Home Messages

- Confirmed cases are commonly error-contaminated:
 - under-reported:
 - limited test capacity
 - asymptomatic patients are not detected.
 - testing error:
 - false positive
 - false negative
 - Incubation times varies from individual to individual:
 - recall bias
 - reporting bias

- Fatality rate estimates based on Definition 1 may not be representative of the actual death rate.
 - reported confirmed cases on a given day usually contracted the virus at an earlier time

Take Home Messages

- COVID-19 data are available from multiple sources, and many methods have been developed:
 - statistical validity/efficiency
 - modeling complexity/validity
 - computation complexity/feasibility

Take Home Messages

- COVID-19 data are available from multiple sources, and many methods have been developed:
 - statistical validity/efficiency
 - modeling complexity/validity
 - computation complexity/feasibility
- An issue is often overlooked:
 - EXAMINING DATA PROVENANCE AND QUALITY IS CRUCIAL!

Take Home Messages

- COVID-19 data are available from multiple sources, and many methods have been developed:
 - statistical validity/efficiency
 - modeling complexity/validity
 - computation complexity/feasibility
- An issue is often overlooked:
 - EXAMINING DATA PROVENANCE AND QUALITY IS CRUCIAL!
- Reported COVID-19 data are not automatically useful:
 - Ignoring error in COVID-19 data can yield misleading results!
 - Carefully scrutinizing COVID-19 is needed!

Take Home Messages

- COVID-19 data are available from multiple sources, and many methods have been developed:
 - statistical validity/efficiency
 - modeling complexity/validity
 - computation complexity/feasibility
- An issue is often overlooked:
 - EXAMINING DATA PROVENANCE AND QUALITY IS CRUCIAL!
- Reported COVID-19 data are not automatically useful:
 - Ignoring error in COVID-19 data can yield misleading results!
 - Carefully scrutinizing COVID-19 is needed!
- Characterizing the form and degree of error in COVID-19 data can be difficult:
 - Conducting sensitivity analyses may be a viable strategy!

Acknowledgements

 Funding Agency:

NSERC

Canada Research Chairs Program