Ch 8. Latent Variables and State-Space Models

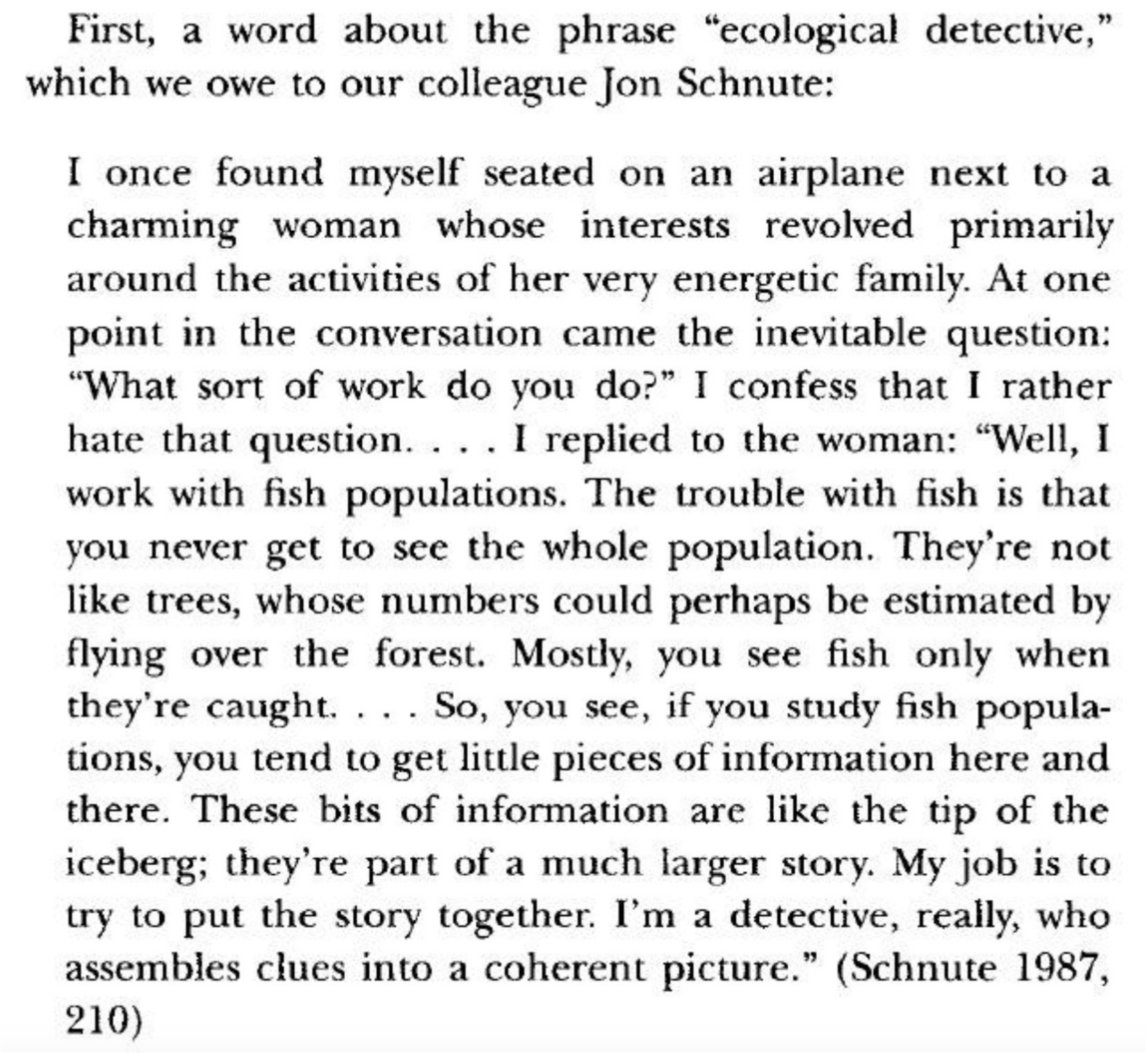
Amy Hurford
Memorial University

Latent variables

Latent variables: The variable of interest is unobserved or estimated with uncertainty

Dietz 2017. Ecological forecasting.

The ecological detective



First, a word about the phrase "ecological detective," which we owe to our colleague Jon Schnute:

> I once found myself seated on an airplane next to a charming woman whose interests revolved primarily around the activities of her very energetic family. At one point in the conversation came the inevitable question: "What sort of work do you do?" I confess that I rather hate that question. . . . I replied to the woman: "Well, I work with fish populations. The trouble with fish is that you never get to see the whole population. They're not like trees, whose numbers could perhaps be estimated by flying over the forest. Mostly, you see fish only when they're caught. . . . So, you see, if you study fish populations, you tend to get little pieces of information here and there. These bits of information are like the tip of the iceberg; they're part of a much larger story. My job is to try to put the story together. I'm a detective, really, who assembles clues into a coherent picture." (Schnute 1987, 210)

Mangel, M. and C. Clark. 1997. *The Ecological Detective: confronting models with data*. Princeton Monographs.

Latent variables

Latent variables: The variable of interest is unobserved or estimated with uncertainty

1. Random and systematic observation errors
2. Proxy data
3. Missing data
4. Unobserved variables

Dietz 2017. Ecological forecasting.

Examples

- Random & systematic: sensor drift
- Proxy: Total Domain Reflectometry (TDR) – 2 electric probes that measures soil impedence estimate soil moisture
- Proxy: NVDI - net primary production
- Proxy: O^{18} in water – temperature, evaporation, and atmospheric cipitation
- Unobserved: resource allocation to growth, fecundity, allometry

Dietz 2017. Ecological forecasting.

Canadian vaccination data

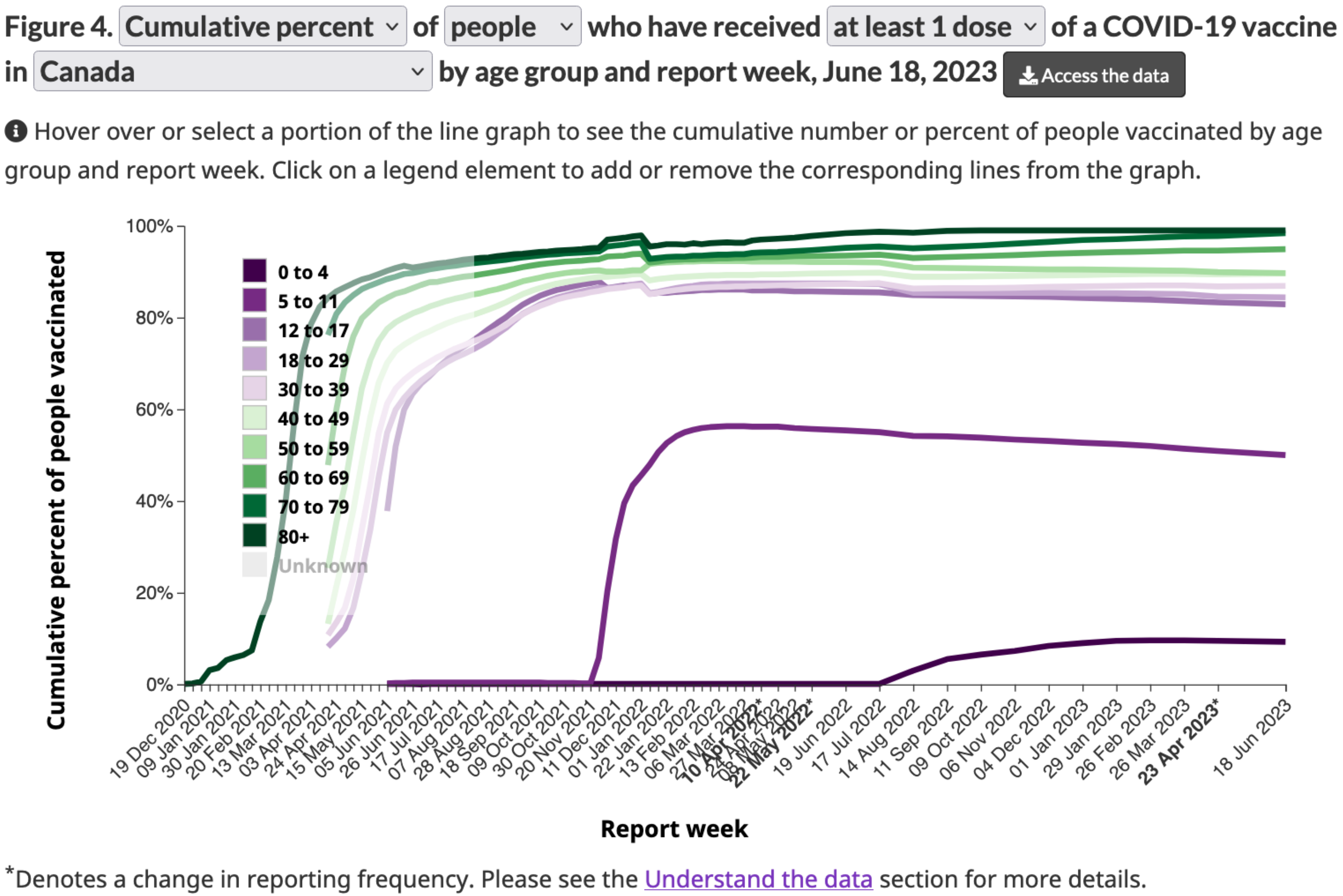


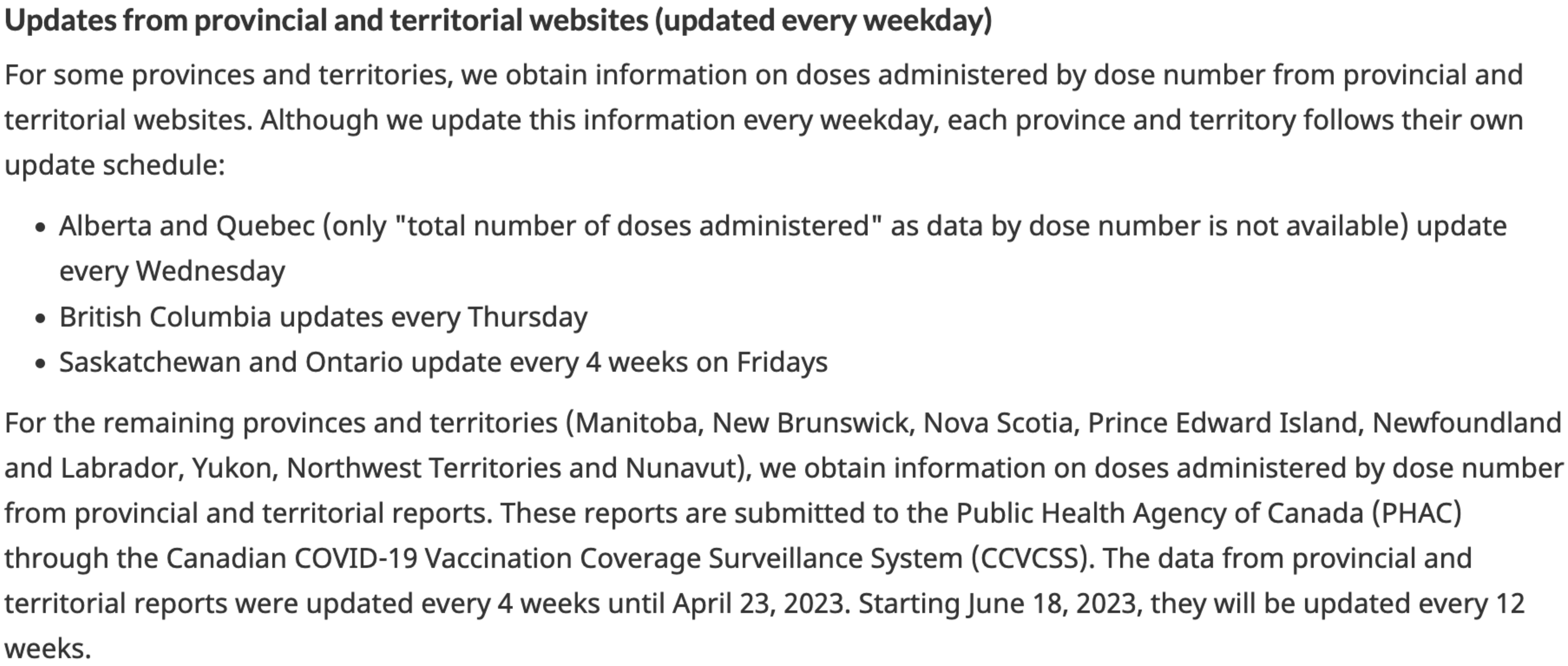
Figure 4. Cumulative percent ⌄ of people ⌄ who have received at least 1 dose ⌄ of a COVID-19 vaccine in Canada ⌄ by age group and report week, June 18, 2023

Hover over or select a portion of the line graph to see the cumulative number or percent of people vaccinated by age group and report week. Click on a legend element to add or remove the corresponding lines from the graph.

*Denotes a change in reporting frequency. Please see the Understand the data section for more details.

Government of Canada https://health-infobase.canada.ca/covid-19/vaccination-coverage/

Reporting frequency (missing data)



Updates from provincial and territorial websites (updated every weekday)

For some provinces and territories, we obtain information on doses administered by dose number from provincial and territorial websites. Although we update this information every weekday, each province and territory follows their own update schedule:

- Alberta and Quebec (only "total number of doses administered" as data by dose number is not available) update every Wednesday
- British Columbia updates every Thursday
- Saskatchewan and Ontario update every 4 weeks on Fridays

For the remaining provinces and territories (Manitoba, New Brunswick, Nova Scotia, Prince Edward Island, Newfoundland and Labrador, Yukon, Northwest Territories and Nunavut), we obtain information on doses administered by dose number from provincial and territorial reports. These reports are submitted to the Public Health Agency of Canada (PHAC) through the Canadian COVID-19 Vaccination Coverage Surveillance System (CCVCSS). The data from provincial and territorial reports were updated every 4 weeks until April 23, 2023. Starting June 18, 2023, they will be updated every 12 weeks.

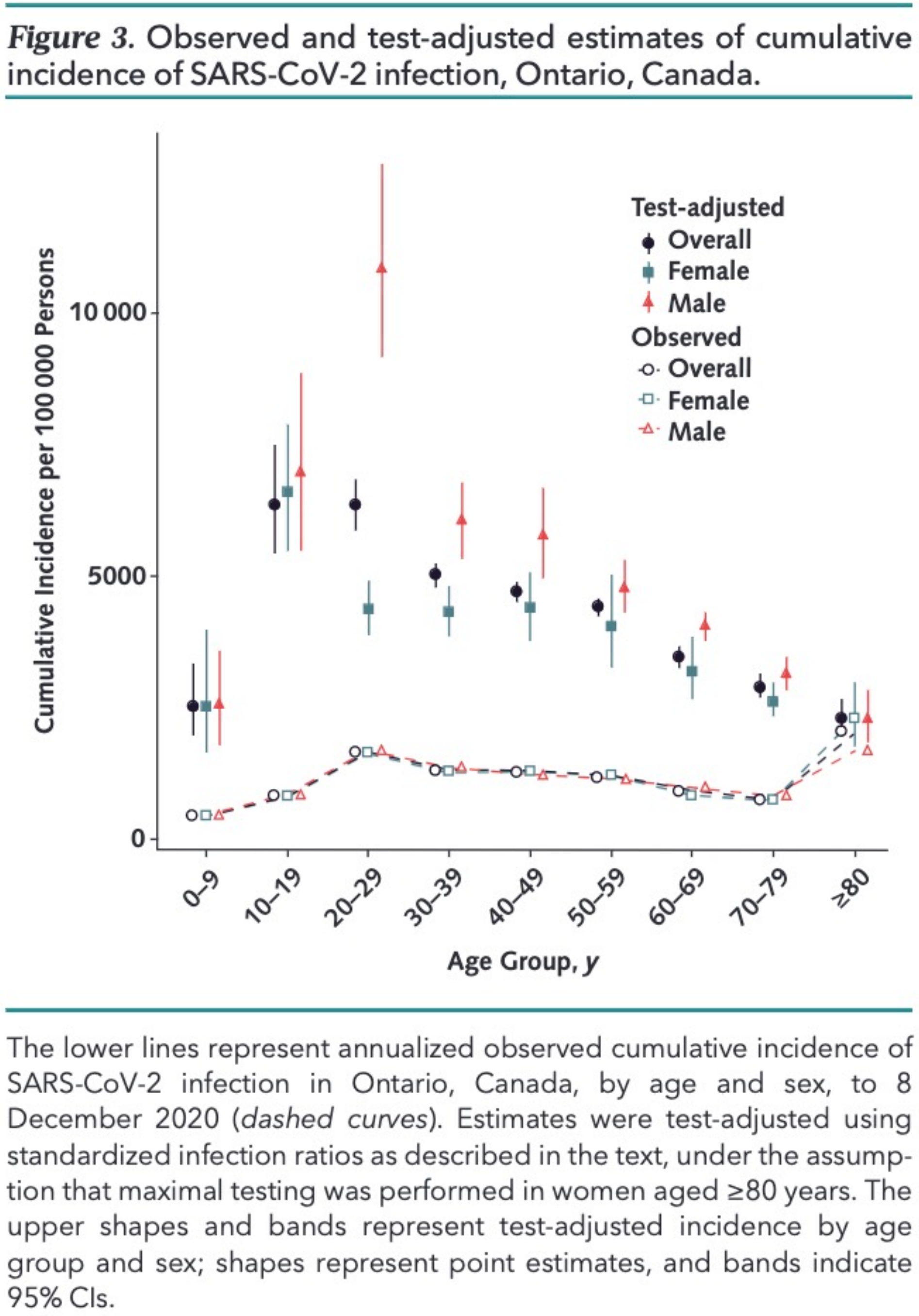


Figure 3. Observed and test-adjusted estimates of cumulative incidence of SARS-CoV-2 infection, Ontario, Canada.

The lower lines represent annualized observed cumulative incidence of SARS-CoV-2 infection in Ontario, Canada, by age and sex, to 8 December 2020 (*dashed curves*). Estimates were test-adjusted using standardized infection ratios as described in the text, under the assumption that maximal testing was performed in women aged ≥80 years. The upper shapes and bands represent test-adjusted incidence by age group and sex; shapes represent point estimates, and bands indicate 95% CIs.

COVID-19 Case Age Distribution: Correction for Differential Testing by Age David N. Fisman, Amy L. Greer, et al. Annals of Internal Medicine

State space models

- For time series analysis, also referred to as hidden Markov models
- hidden => latent; Markov => next state depends only on current state

Dietz 2017. Ecological forecasting.

$y_t \sim N(x_t, \tau_{obs})$ Data model

$x_{t+1} \sim N(x_t, \tau_{add})$ Process model

$1/\tau_{obs} \sim Gamma(a_{obs}, r_{obs})$ Observation prior

$1/\tau_{add} \sim Gamma(a_{add}, r_{add})$ Process prior

$x_0 \sim N(x_{IC}, \tau_{IC})$ Initial condition prior

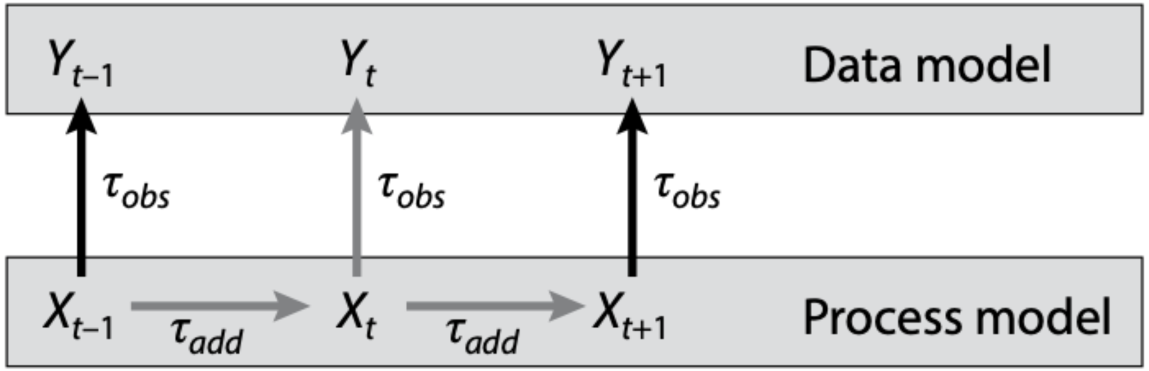


FIGURE 8.1. State-space model describing the evolution of the latent state variable, X, conditional on the observations, Y. In this random-walk example the only components are observation error, τ_{obs}, and process error, τ_{add}. The gray arrows indicate the connections relevant for estimating the posterior distribution for X_t.

Dietz 2017. Ecological forecasting.

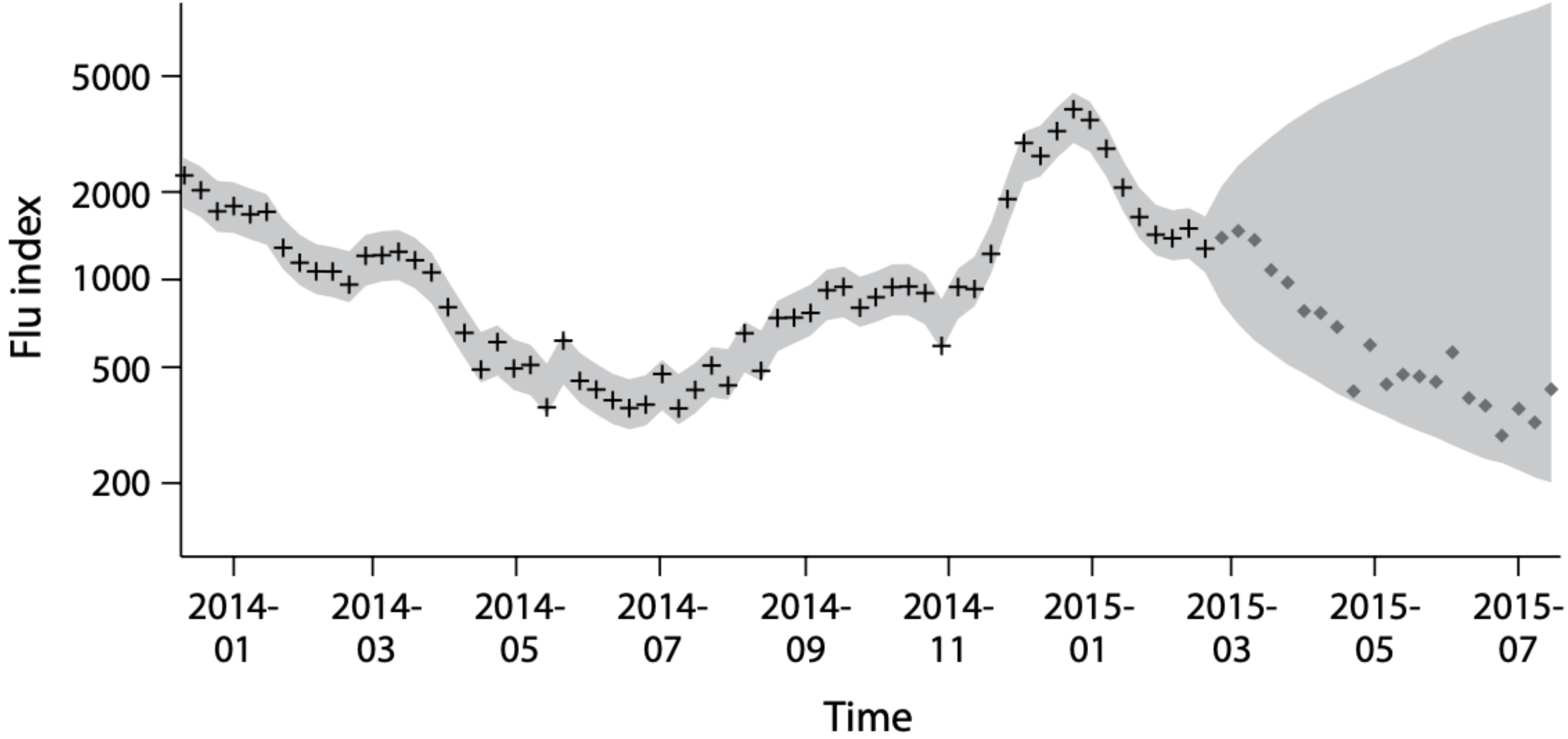


FIGURE 8.3. Forecast of the Google flu trend data compared to observations. Gray diamonds (starting March 2015) were not assimilated into the forecast.

Dietz 2017. Ecological forecasting.

That process model was not SIR – that was a random walk…

$$x_{t+1} = x_t + \epsilon$$

$$\epsilon \sim N(0, \sigma^2)$$

…. much less epidemiological detail than McMasterPandemic

https://canmod.github.io/macpan-book/index.html#history-and-motivation

Key concepts

3. Because of their capacity to flexibly capture and partition a wide range of uncertainties and address the complexities of real data, Dietz recommends state space models as the basis of forecasting.

5. Missing data gaps and irregularly spaced data are handled automatically with uncertainties increasing with distance to the nearest observation.

Dietz 2017. Ecological forecasting.

Ch 9 Fusing Data Sources

“Balancing the information provided by different data sources remains among the most debated topics in ecological model-data fusion”

Dietz 2017. Ecological forecasting.

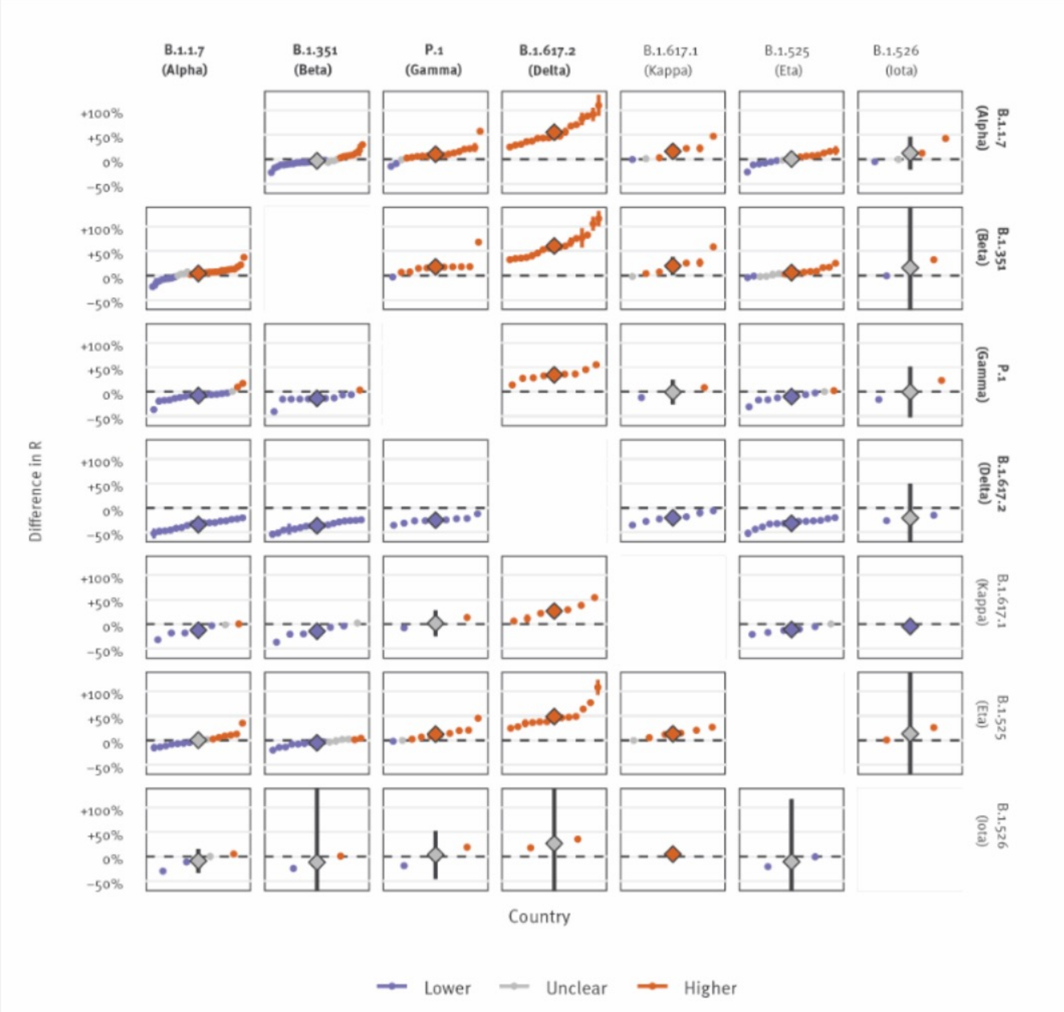
Fusing data sources

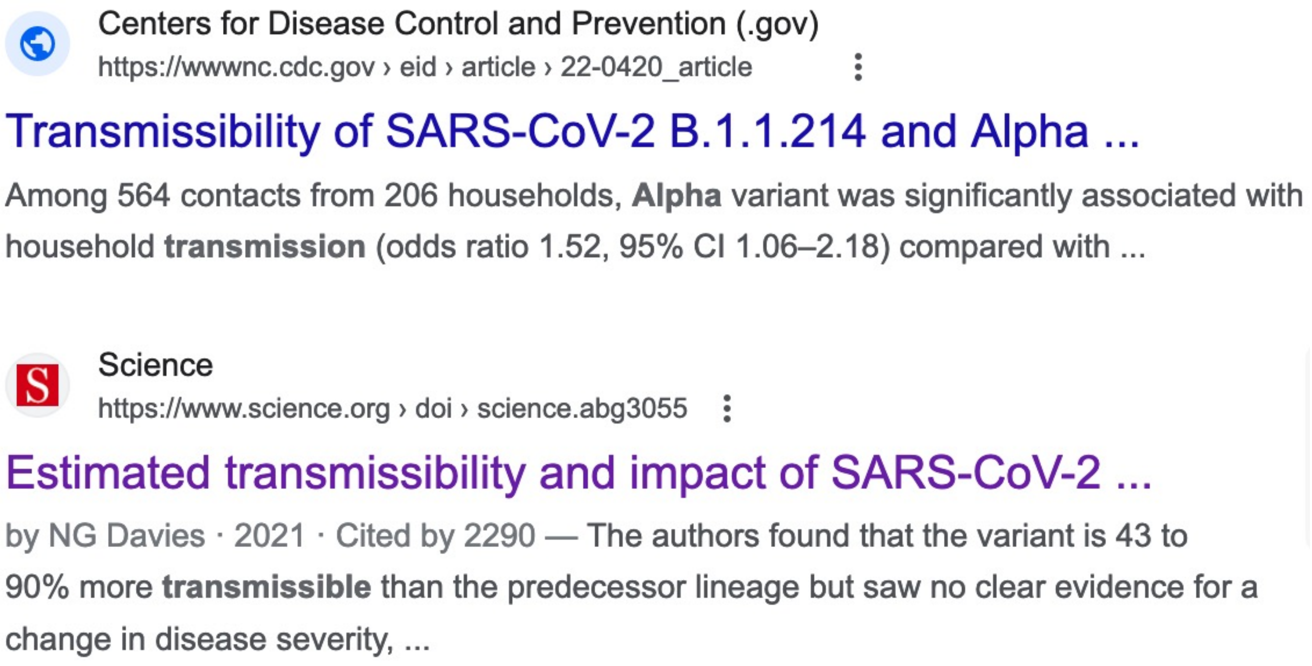
- Fusing data involves more than joining data sources
- Naïve interpolation or extrapolation can lead to biased and overconfident results
- Covariances are critical to leverage complementary data types

Dietz 2017. Ecological forecasting.

Fusing data sources

Figure 3. Effective reproduction number of SARS-CoV-2 variants of concern/interest compared against each other, 64 countries, data until 3 June 2021





Centers for Disease Control and Prevention (.gov)
https://wwwnc.cdc.gov › eid › article › 22-0420_article

Transmissibility of SARS-CoV-2 B.1.1.214 and Alpha ...

Among 564 contacts from 206 households, **Alpha** variant was significantly associated with household **transmission** (odds ratio 1.52, 95% CI 1.06–2.18) compared with ...

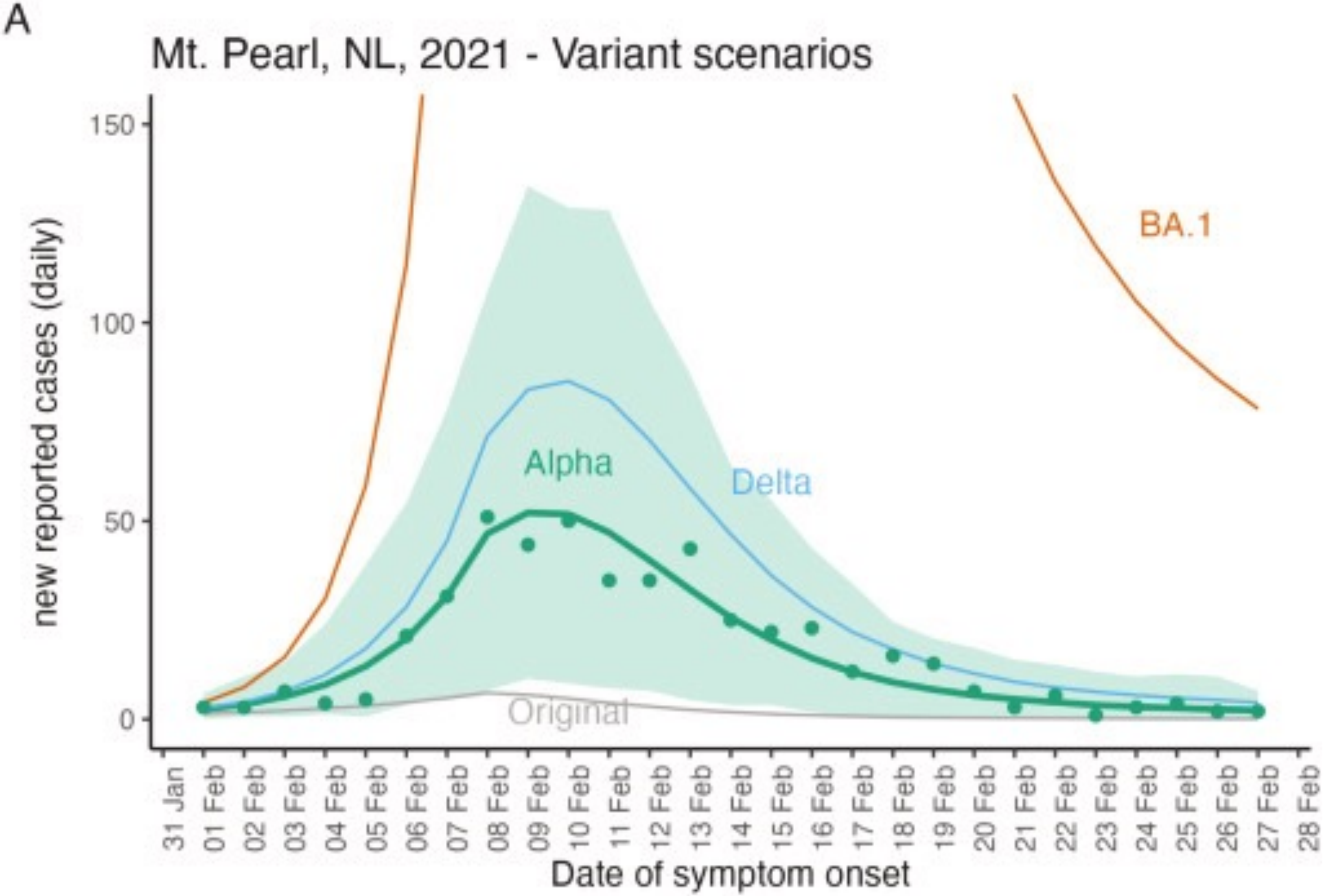
Science
https://www.science.org › doi › science.abg3055

Estimated transmissibility and impact of SARS-CoV-2 ...

by NG Davies · 2021 · Cited by 2290 — The authors found that the variant is 43 to 90% more **transmissible** than the predecessor lineage but saw no clear evidence for a change in disease severity, ...

Increased transmissibility and global spread of SARS-CoV-2 variants of concern as at June 2021
Campbell et al. (2021), Eurosurveillence. https://doi.org/10.2807/1560-7917.ES.2021.26.24.2100509

Fusing data sources - counterfactuals



Hurford et al. 2023. Pandemic modelling for regions implementing an elimination strategy

Meta-analysis

- Meta-analyses combine information, usually in the form of summary statistics from independent studies
- Reporting bias is a challenge for high quality meta-analyses; that is, that negative results are not reported
- While less common, meta-analysis can also be used to directly estimate priors for parameters in a larger model (LeBauer et al. 2013)

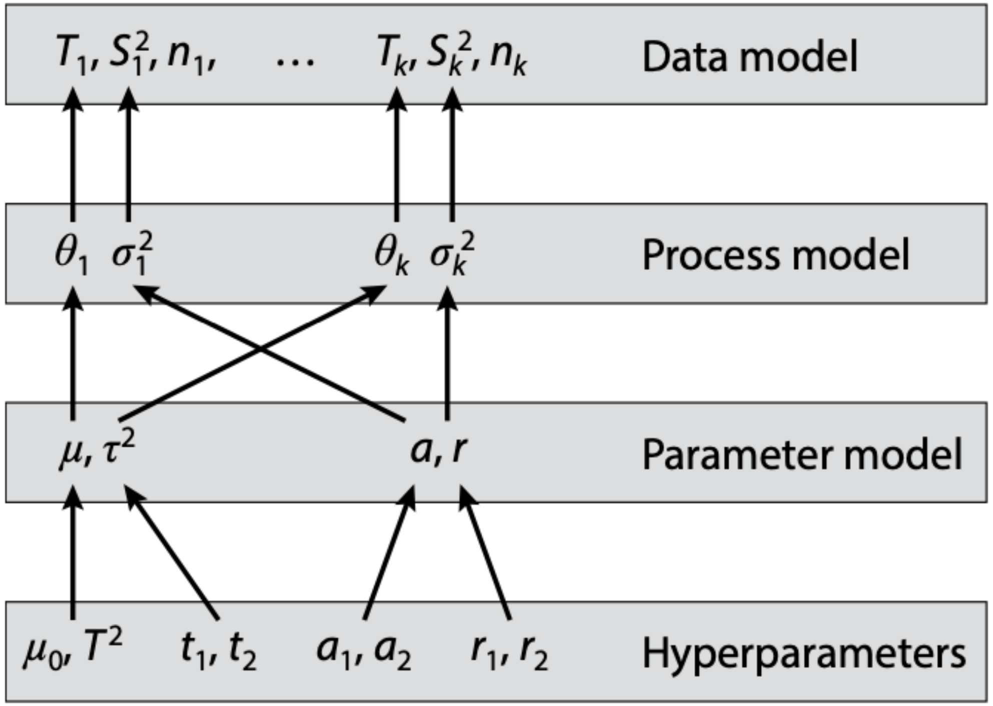
Dietz 2017. Ecological forecasting.

Meta-analysis

- Compared to parameterizing a model from a single study or site, a meta-analysis provides both greater constraint and the ability to account for the real ecological variability among multiple studies.
- What distinguishes a meta-analytical model from other models is that the observations are typically summary statistics, and thus there is a need for *different studies to have different weights based on their sample sizes and variability*

Dietz 2017. Ecological forecasting.

Meta-analysis



Reported summary statistics in study, *i*

True summary statistics in study, *i*

True across study mean

FIGURE 9.3. Hierarchical Bayes meta-analysis model. In a meta-analysis the data are not raw observations but are summary statistics from each of k publications: sample mean (T), sample standard deviation (S), and sample size (n).

Dietz 2017. Ecological forecasting.

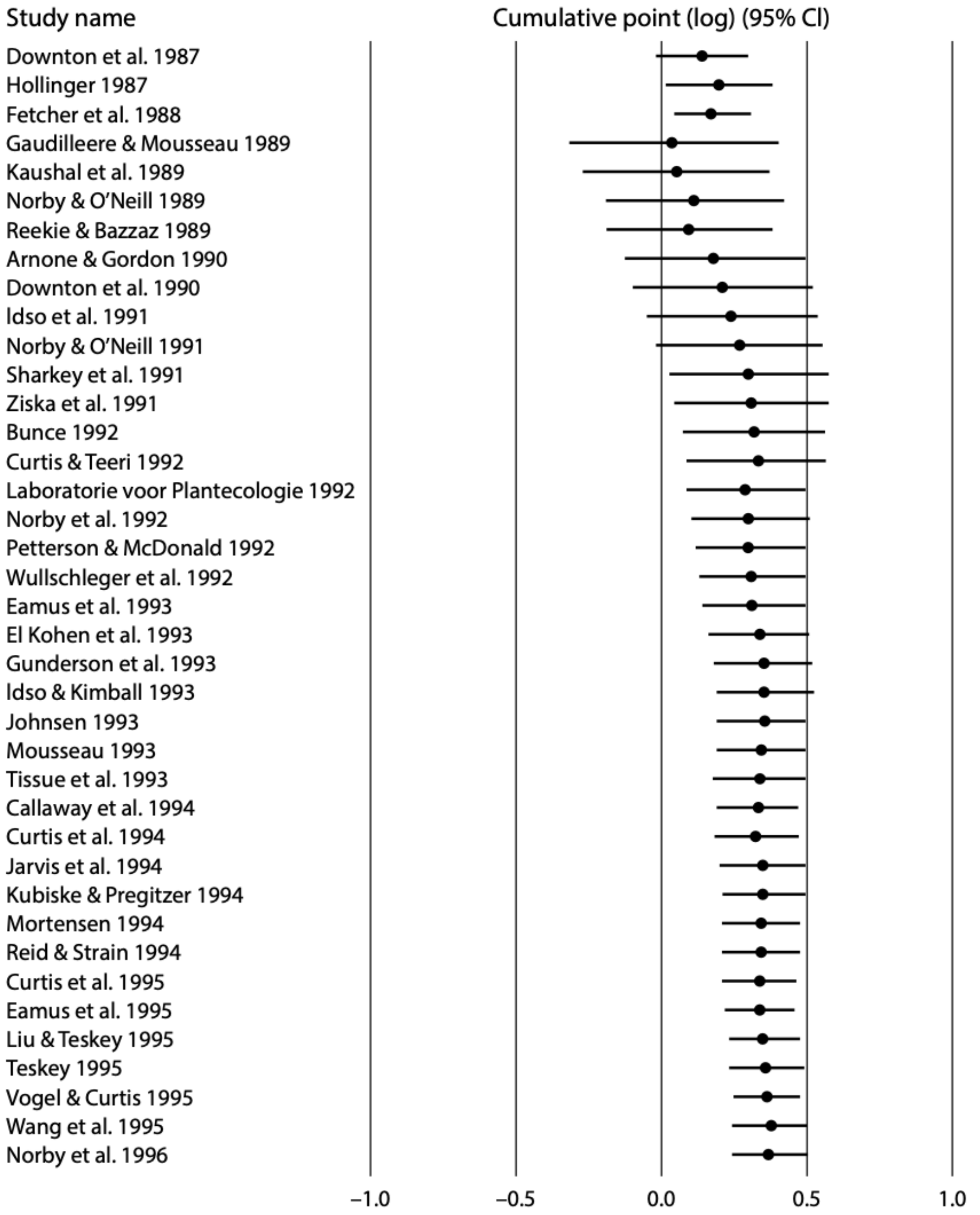


FIGURE 9.2. Cumulative meta-analysis of elevated CO_2 effects on net assimilation in woody plants. Data from Curtis and Wang (1998); figure from *Handbook of Meta-analysis in Ecology and Evolution*, edited by Julia Koricheva, Jessica Gurevitch, and Kerrie Mengersen. Copyright © 2013 by Princeton University Press. Reprinted by permission.

Dietz 2017. Ecological forecasting.

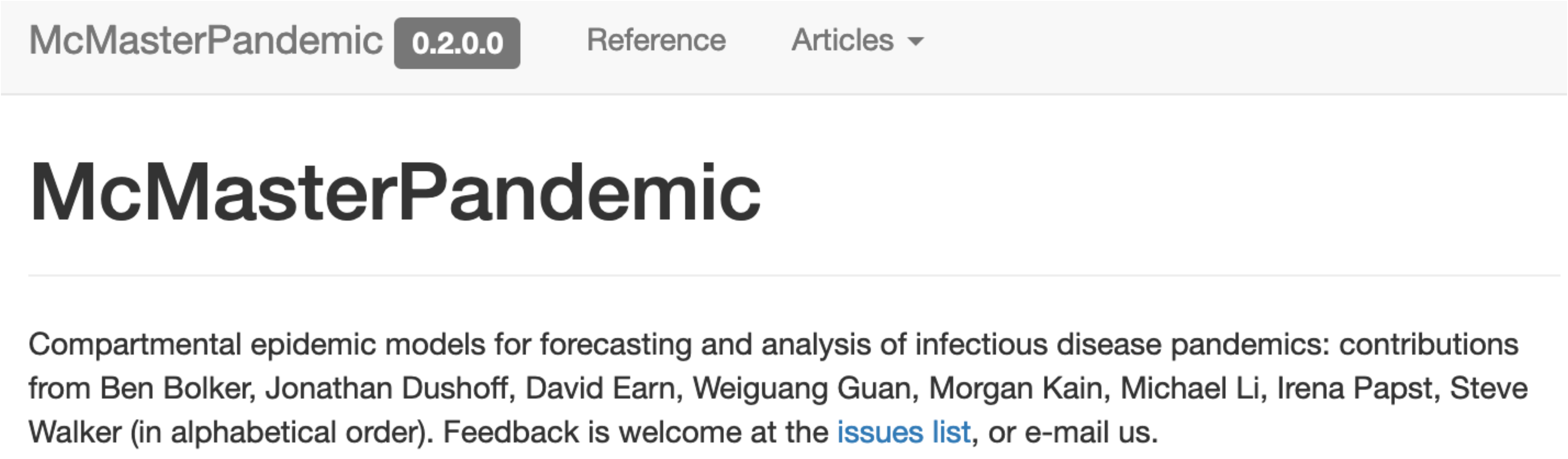
PHAC – Emerging Sciences Group

Parameter estimation tables

IFR, CFR, Incubation period, Latent period, Infectious period, Asymptomatic-Inf, Serial interval

Epi+Model Data Parameter Tables 2020-06-29.xls

Parameter estimates



McMasterPandemic

Compartmental epidemic models for forecasting and analysis of infectious disease pandemics: contributions from Ben Bolker, Jonathan Dushoff, David Earn, Weiguang Guan, Morgan Kain, Michael Li, Irena Papst, Steve Walker (in alphabetical order). Feedback is welcome at the issues list, or e-mail us.

https://canmod.github.io/macpan-book/index.html#history-and-motivation

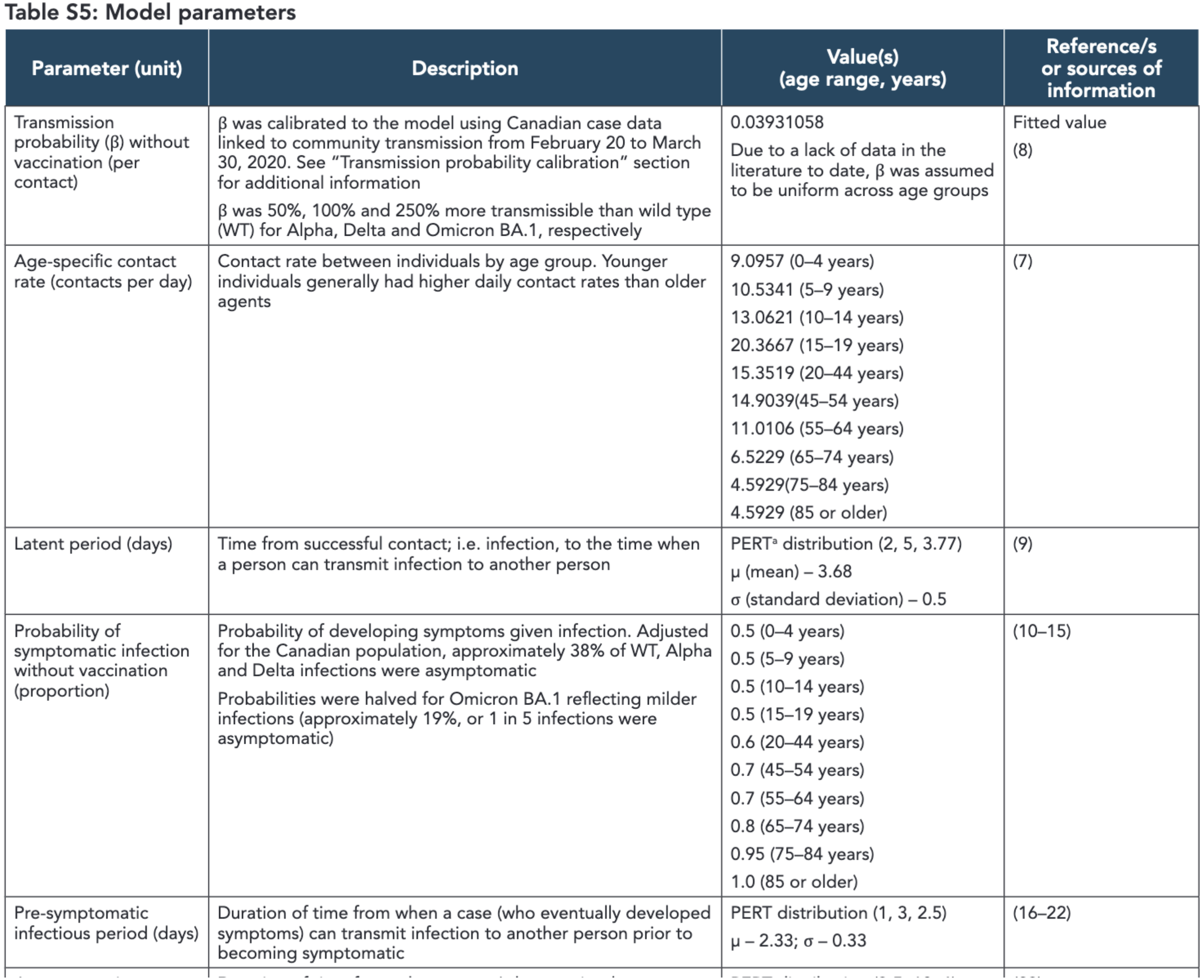
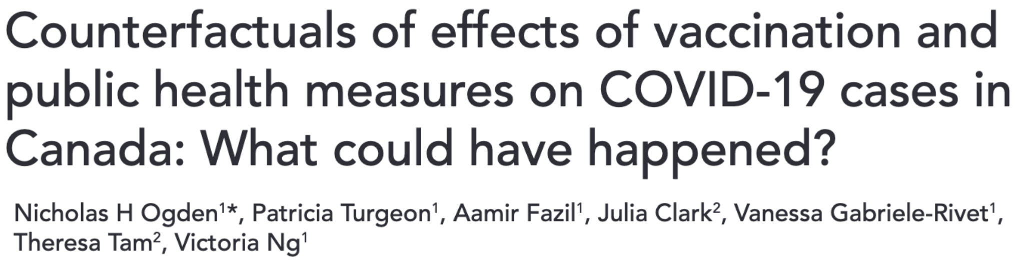


Table S5: Model parameters

Parameter (unit)	Description	Value(s) (age range, years)	Reference/s or sources of information
Transmission probability (β) without vaccination (per contact)	β was calibrated to the model using Canadian case data linked to community transmission from February 20 to March 30, 2020. See "Transmission probability calibration" section for additional information β was 50%, 100% and 250% more transmissible than wild type (WT) for Alpha, Delta and Omicron BA.1, respectively	0.03931058 Due to a lack of data in the literature to date, β was assumed to be uniform across age groups	Fitted value (8)
Age-specific contact rate (contacts per day)	Contact rate between individuals by age group. Younger individuals generally had higher daily contact rates than older agents	9.0957 (0–4 years) 10.5341 (5–9 years) 13.0621 (10–14 years) 20.3667 (15–19 years) 15.3519 (20–44 years) 14.9039(45–54 years) 11.0106 (55–64 years) 6.5229 (65–74 years) 4.5929(75–84 years) 4.5929 (85 or older)	(7)
Latent period (days)	Time from successful contact; i.e. infection, to the time when a person can transmit infection to another person	PERT[a] distribution (2, 5, 3.77) μ (mean) – 3.68 σ (standard deviation) – 0.5	(9)
Probability of symptomatic infection without vaccination (proportion)	Probability of developing symptoms given infection. Adjusted for the Canadian population, approximately 38% of WT, Alpha and Delta infections were asymptomatic Probabilities were halved for Omicron BA.1 reflecting milder infections (approximately 19%, or 1 in 5 infections were asymptomatic)	0.5 (0–4 years) 0.5 (5–9 years) 0.5 (10–14 years) 0.5 (15–19 years) 0.6 (20–44 years) 0.7 (45–54 years) 0.7 (55–64 years) 0.8 (65–74 years) 0.95 (75–84 years) 1.0 (85 or older)	(10–15)
Pre-symptomatic infectious period (days)	Duration of time from when a case (who eventually developed symptoms) can transmit infection to another person prior to becoming symptomatic	PERT distribution (1, 3, 2.5) μ – 2.33; σ – 0.33	(16–22)

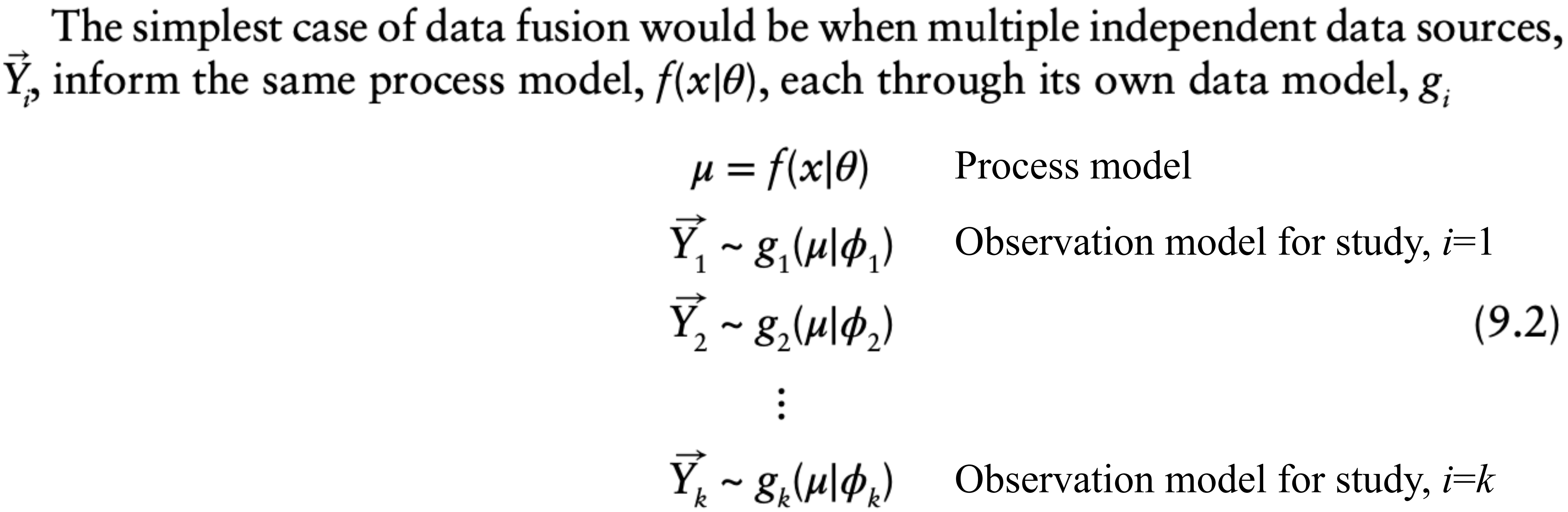


Counterfactuals of effects of vaccination and public health measures on COVID-19 cases in Canada: What could have happened?

Nicholas H Ogden[1]*, Patricia Turgeon[1], Aamir Fazil[1], Julia Clark[2], Vanessa Gabriele-Rivet[1], Theresa Tam[2], Victoria Ng[1]

https://www.canada.ca/content/dam/phac-aspc/documents/services/reports-publications/canada-communicable-disease-report-ccdr/monthly-issue/2022-48/issue-7-8-july-august-2022/ccdrv48i78a01s-eng.pdf

Combining data

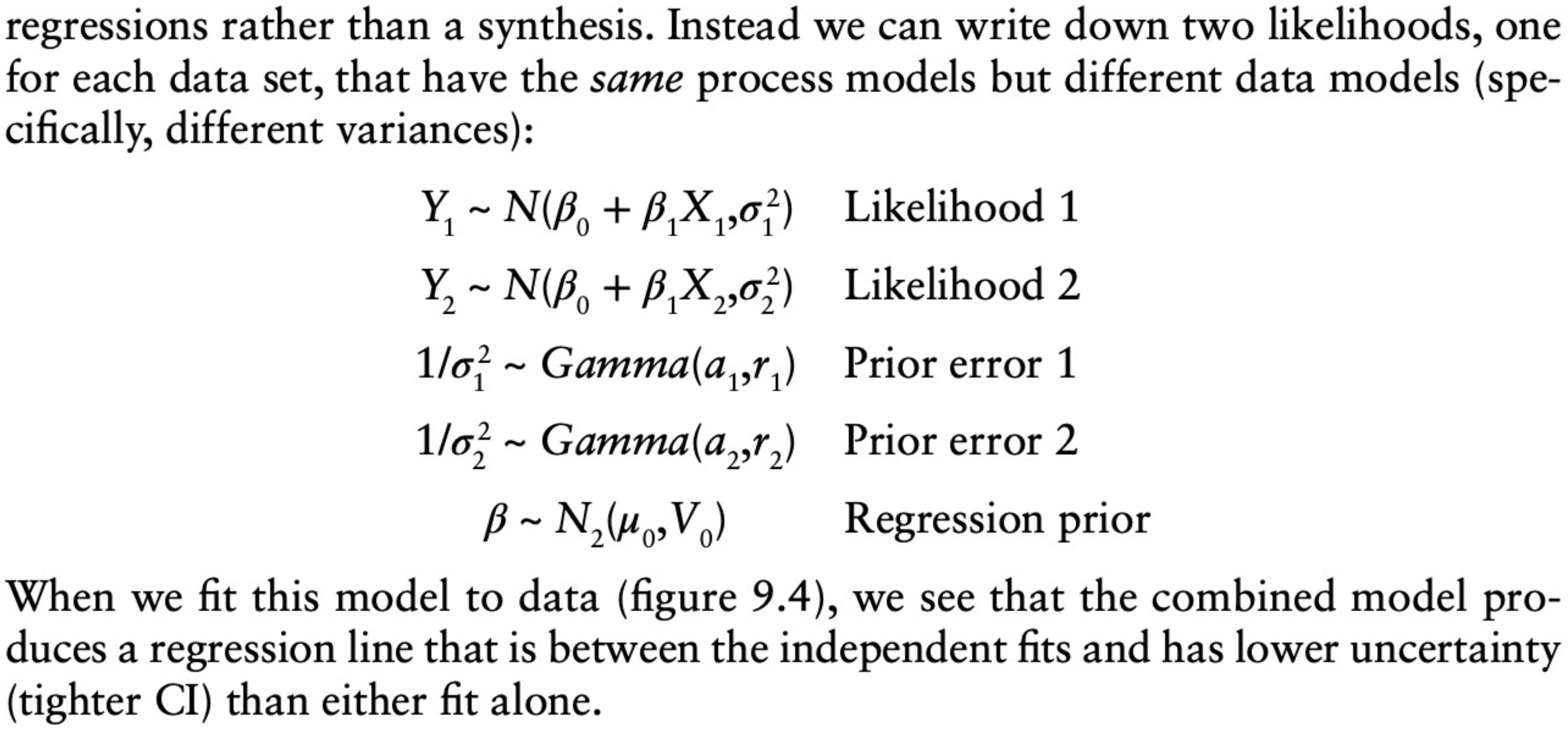


The simplest case of data fusion would be when multiple independent data sources, $\vec{Y}_i$, inform the same process model, $f(x|\theta)$, each through its own data model, g_i

$$\begin{aligned}
\mu &= f(x|\theta) && \text{Process model} \\
\vec{Y}_1 &\sim g_1(\mu|\phi_1) && \text{Observation model for study, } i\text{=1} \\
\vec{Y}_2 &\sim g_2(\mu|\phi_2) && \qquad\qquad\qquad\qquad\qquad\qquad (9.2) \\
&\;\vdots \\
\vec{Y}_k &\sim g_k(\mu|\phi_k) && \text{Observation model for study, } i\text{=}k
\end{aligned}$$

Dietz 2017. Ecological forecasting.

Combining data



regressions rather than a synthesis. Instead we can write down two likelihoods, one for each data set, that have the *same* process models but different data models (specifically, different variances):

$$\begin{array}{rl}
Y_1 \sim N(\beta_0 + \beta_1 X_1, \sigma_1^2) & \text{Likelihood 1} \\
Y_2 \sim N(\beta_0 + \beta_1 X_2, \sigma_2^2) & \text{Likelihood 2} \\
1/\sigma_1^2 \sim Gamma(a_1, r_1) & \text{Prior error 1} \\
1/\sigma_2^2 \sim Gamma(a_2, r_2) & \text{Prior error 2} \\
\beta \sim N_2(\mu_0, V_0) & \text{Regression prior}
\end{array}$$

When we fit this model to data (figure 9.4), we see that the combined model produces a regression line that is between the independent fits and has lower uncertainty (tighter CI) than either fit alone.

Dietz 2017. Ecological forecasting.

Combining data

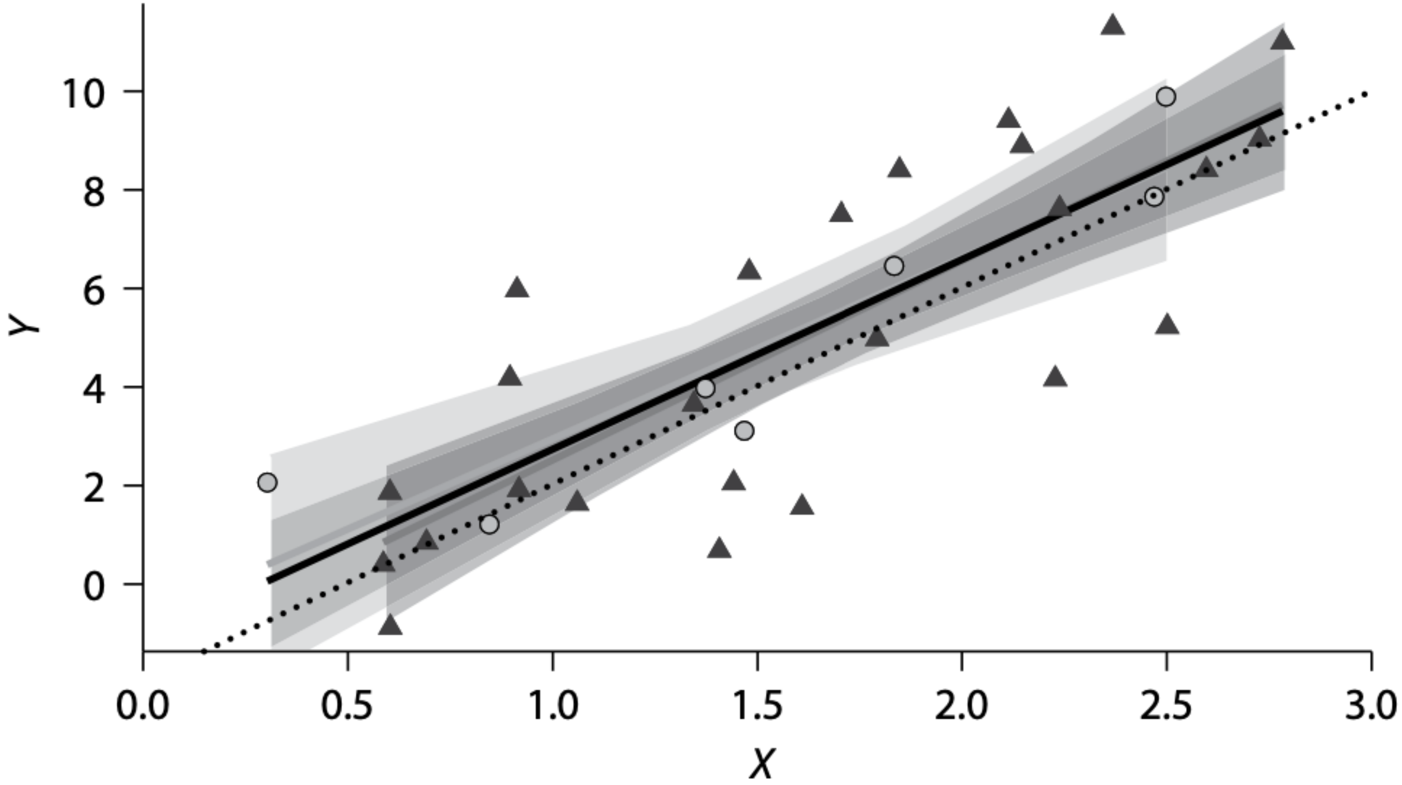


FIGURE 9.4. Fusing two regression models. Comparison of independent fits to independent data (gray lines: diamonds = common but noisy; circles = precise but expensive) and combined fit (black line).

Dietz 2017. Ecological forecasting.

Key concepts

3. Fusing data involves more than concatenating files. Considering uncertainty is essential.

7. When combining likelihoods avoid ad hoc weightings, instead use a model

9. State-space models can allow us to combine spatial and temporal information that operate at different scales, even if scales are misaligned.

Dietz 2017. Ecological forecasting.