Introduction to Modeling for Public Health

Dr. Rachel Sippy
University of Cambridge

Objectives

- Understand the types of models commonly used in public health research
- Learn some modeling vocabulary
- Identify modeling goals and questions
- Understand sources of uncertainty and how to represent uncertainty
- Practice manuscript evaluation & interpretation

Post Questions in the Chat!

(we will have breaks to answer these during the workshop)

Workshop Schedule

Time	Topics
2:00—2:10 pm	Outline & Introduction
2:10—3:00 pm	Defining Models & Modeling Terms
3:00—3:30 pm	Types of Models
3:30—3:40 pm	Break
3:40—4:00 pm	Modeling Goals & Questions
4:00—4:20 pm	Uncertainty in Models
4:20—4:40 pm	Interpreting & Evaluating Modeling Papers
4:40—5:00 pm	Questions & Discussion

Introductions!

- Dr. Rachel Sippy
- Postdoctoral Researcher at University of Cambridge

Defining Models

& Modeling Glossary

What is a Model?

What is a Model?

- A representation to analyze or explain something
- A simplification of a complex phenomenon

- A way to examine and understand patterns in data
- An equation or algorithm to represent relationships between variables

What Types of Models Are There?

What Types of Models Are There?

Following Lowe *et al.* (2017), a Bayesian hierarchical mixed model was fitted to counts of dengue cases from

quintiles. Specifically, we fit the following log-linear model with only an interaction term between HAQI and SDIq (that is, no main effect terms):

$$\log_{x_{min}} \approx \mathrm{HAQI} \times \mathrm{SDIq} + \varepsilon$$

We implemented a standard deterministic SEIR model with compartments for susceptible (S), exposed-not-infectious (E), infectious (I) and recovered (R). This is the model we use to

Effects of pre-existing immunity on seroconversion to recent strains. We predicted the odds of seroconversion ($c_{i,j}$) to one of four recent strains i (i.e. A/HongKong/2014, A/Texas/2012, A/Victoria/2009 or A/Perth/2009) by fitting logistic regression with predictors that

Model Types & Model Methods

Type

- Linear regression
- SIR model
- Machine learning
- Survival analysis

Method

- Ordinary least squares
- Deterministic with difference equations
- k-nearest neighbors
- Kaplan-Meier estimation

Mix & Match Models & Methods

- The same model could be calculated with multiple methods
  - Example: linear regression
    - Ordinary least squares
    - Maximum likelihood
    - Generalized least squares
- The same method could be used with multiple models
  - Example: k-nearest neighbors
    - Machine learning
    - Spatial interpolation
- Some models are only used with one method
  - Example:

Why Choose a Certain Model Type or Method?

Why Choose a Certain Model Type or Method?

- Appropriate for the data
  - Data availability or type of data (e.g. binary outcomes)
- Question of interest
  - Predict future epidemic size or understand epidemic cause?
- Previous work with similar data used this model/method
  - If we know it works, use it again!
- Comparison of models found one to be superior
  - Methods get refined/improved over time
  - Some approaches more popular?

Why Choose a Certain Model Type or Method?

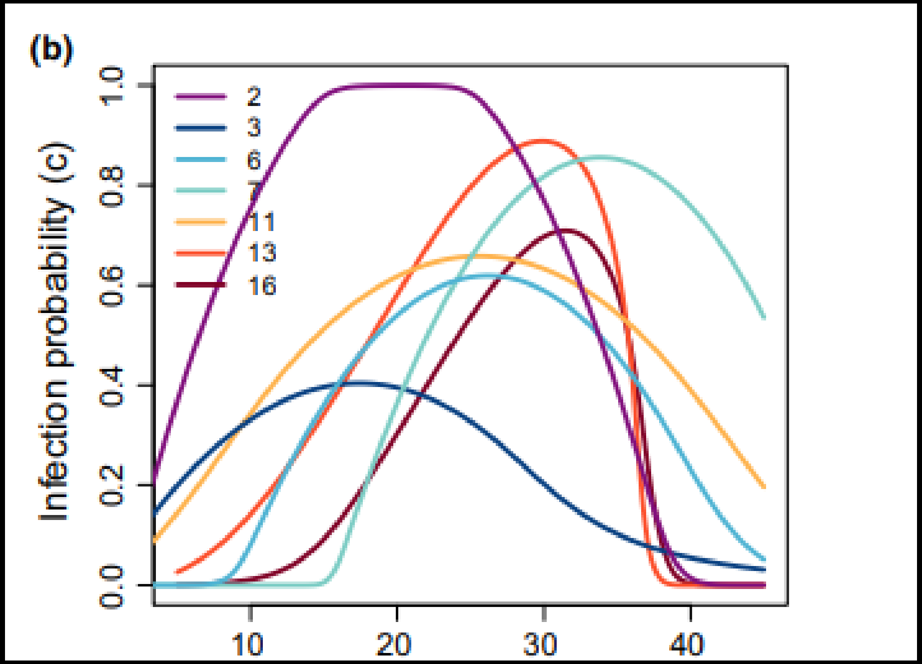
- Software/programs
  - Some are more user-friendly
- Computational power
  - More complicated models take more time
- Favorite model/method
  - Familiarity & comfort with method!

Why Choose a Certain Model Type or Method?

- If results represent a true characteristic of the data and the model is appropriate, then many models/methods will give the same results!

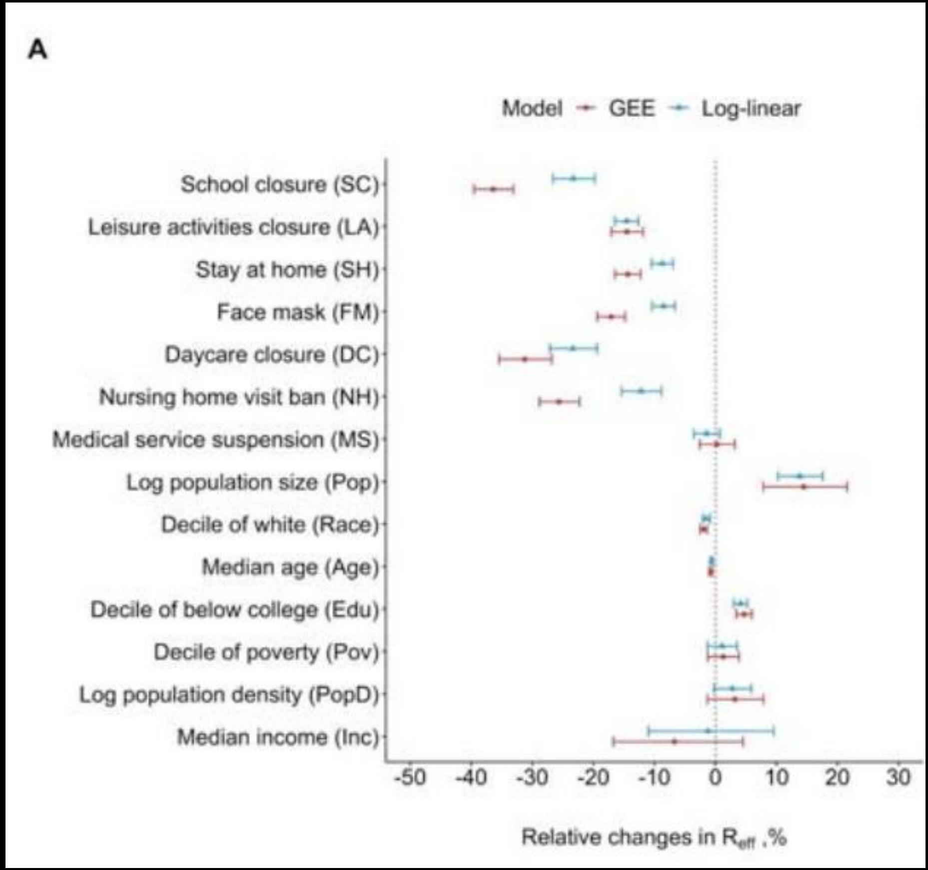
Modeling Terms & Glossary

- Outcomes: y variable, dependent variables, event, outputs, response variable
  - An unknown!
  - What we want to predict
  - What we want to estimate



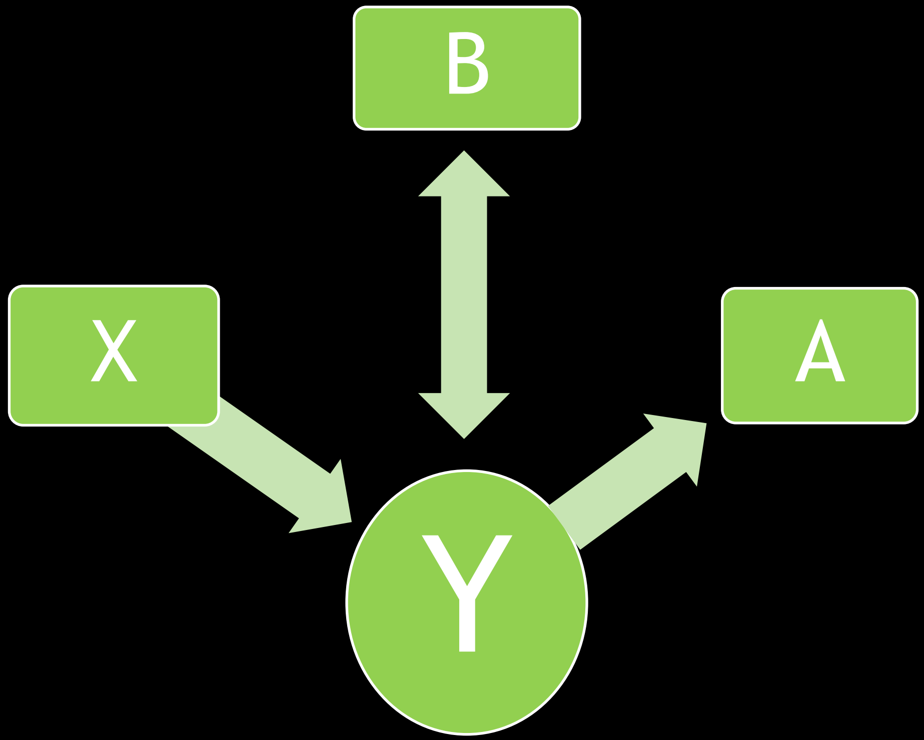
Modeling Terms & Glossary

- Input variable: x variables, descriptors, predictors, exposures, factors, inputs, independent variables , features, explanatory variable
  - Known variables!
  - Data available to put into model
  - May be main purpose of study or included for adjustment



Modeling Terms & Glossary

- Association: A relationship between two variables
  - When values of the first variable change, the values of the second variable also change in some pattern.



Modeling Terms & Glossary

- Parameter: an inherent characteristic of a system or population
  - Sometimes we use models to estimate parameters
  - Sometimes we already know parameters and put them into models

Table 1. Variable and parameters common for all geographic locations

Parameter, meaning	Value
β, basal transmission rates	optimized to fit data
Factors modifying transmission rate	
ε, asymptomatic transmission	0.75
ρ, reduced healthcare worker interactions	0.8
ρ_v, reduced visitor-community interaction	0.5
γ, quarantine	0.2
γ_v, quarantine for visitor	0.3
κ, hospital precautions	0.5
η, healthcare worker precautions	0.2375
Population fractions	
p_i, i =0,...,13, onset of symptoms after day i	0.000792, 0.00198, 0.1056, 0.198, 0.2376, 0.0858, 0.0528, 0.0462, 0.0396, 0.0264, 0.0198, 0.0198, 0.0198, 0
$q_{s,i}$, i =0,...,4, symptomatic quarantine after day/stage i	C: 0.1, 0.4, 0.8, 0.9, 0.99; H: 0.2, 0.5, 0.9, 0.98, 0.99
r, transition to next symptomatic day/stage	0.2
ν, symptomatic hospitalization	0.075
ι, icu admission rate of hospitalized patients	0.2

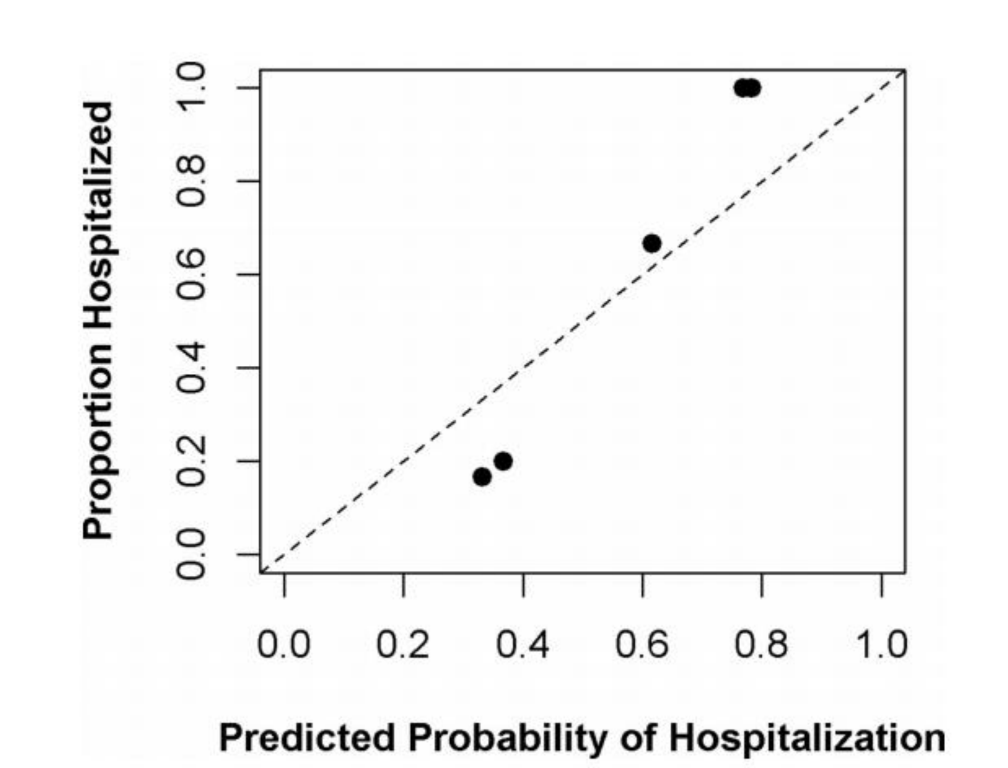
Modeling Terms & Glossary

- Parameter: an inherent characteristic of a system or population
  - Sometimes we use models to estimate parameters
  - Sometimes we already know parameters and put them into models

Table 1. Variable and parameters common for all geographic locations

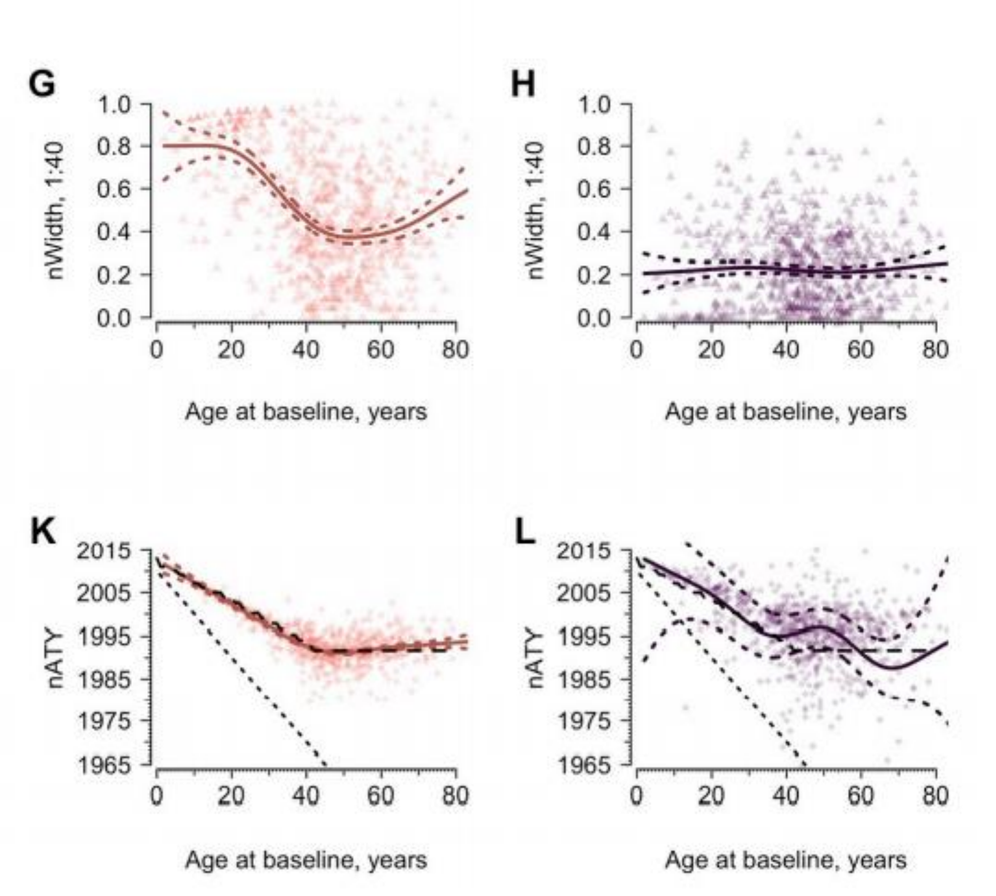
Parameter, meaning	Value
β, basal transmission rates	optimized to fit data
Factors modifying transmission rate	
ε, asymptomatic transmission	0.75
ρ, reduced healthcare worker interactions	0.8
ρ_v, reduced visitor-community interaction	0.5
γ, quarantine	0.2
γ_v, quarantine for visitor	0.3
κ, hospital precautions	0.5
η, healthcare worker precautions	0.2375
Population fractions	
p_i, i =0,...,13, onset of symptoms after day i	0.000792, 0.00198, 0.1056, 0.198, 0.2376, 0.0858, 0.0528, 0.0462, 0.0396, 0.0264, 0.0198, 0.0198, 0.0198, 0
$q_{s,i}$, i =0,...,4, symptomatic quarantine after day/stage i	C: 0.1, 0.4, 0.8, 0.9, 0.99; H: 0.2, 0.5, 0.9, 0.98, 0.99
r, transition to next symptomatic day/stage	0.2
ν, symptomatic hospitalization	0.075
ι, icu admission rate of hospitalized patients	0.2

Modeling Terms & Glossary



- Validation: a way to assess whether a model's predictions are correct and how correct they are
    - Usually we are comparing the model's results with some real data

Modeling Terms & Glossary



- Uncertainty: difference between model estimate/prediction and reality
  - Models are best representation of reality (hopefully)
  - “All models are wrong, some are useful” -Box

Questions & Clarifications

Workshop Schedule

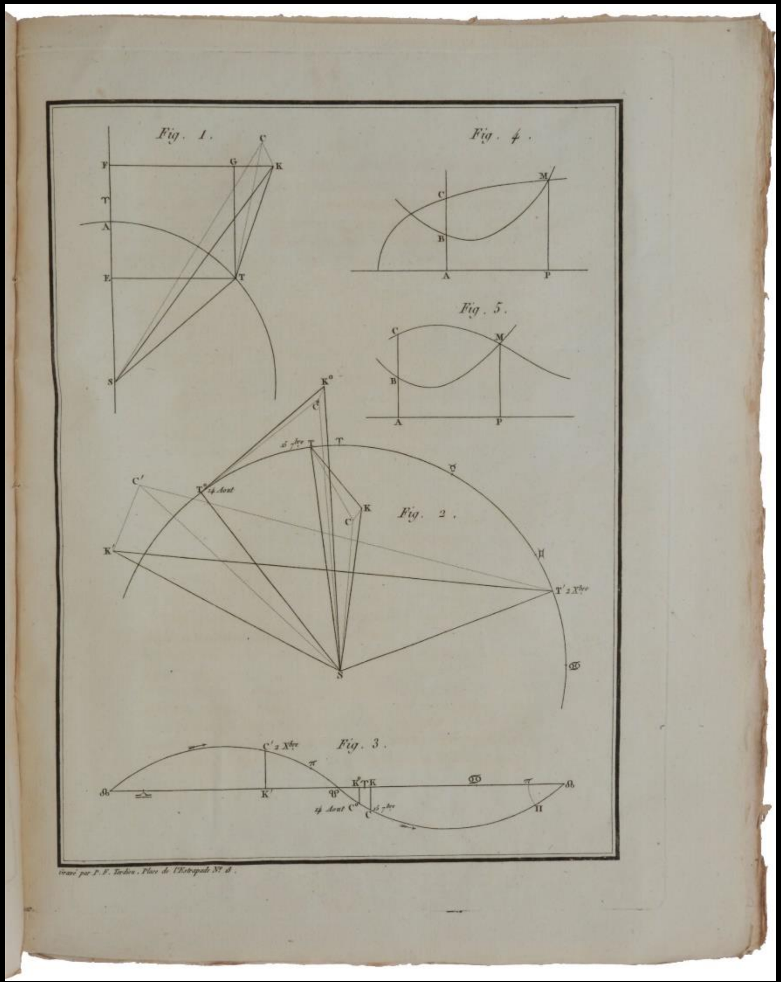
Time	Topics
~~2:00–2:10 pm~~	~~Outline & Introduction~~
~~2:10–3:00 pm~~	~~Defining Models & Modeling Terms~~
3:00–3:30 pm	Types of Models
3:30–3:40 pm	Break
3:40–4:00 pm	Modeling Goals & Questions
4:00–4:20 pm	Uncertainty in Models
4:20–4:40 pm	Interpreting & Evaluating Modeling Papers
4:40–5:00 pm	Questions & Discussion

Types of Models

And Modeling Approaches!

Many Models Exist!

- Models have been developed for multiple fields of study
  - Economics, epidemiology, agriculture, computer science, engineering, astronomy, physics
  - Epidemiology borrows lots of models!
  - Mixed terminology for models



Legendre, 1805

Modeling Types - Glossary

- An approach to modeling/analyzing data
- A way of thinking about data representation
- Underlying framework/philosophy of models

Three Major Model Types

(in epidemiology)
(general modeling approaches)

- Statistical models
- Mathematical models
- Machine learning

Three Major Model Types

(used in epidemiology)
(general modeling approaches)

- Statistical models
  - Regression (many!), time series
- Mathematical models
- Machine learning

Three Major Model Types

(used in epidemiology)
(general modeling approaches)

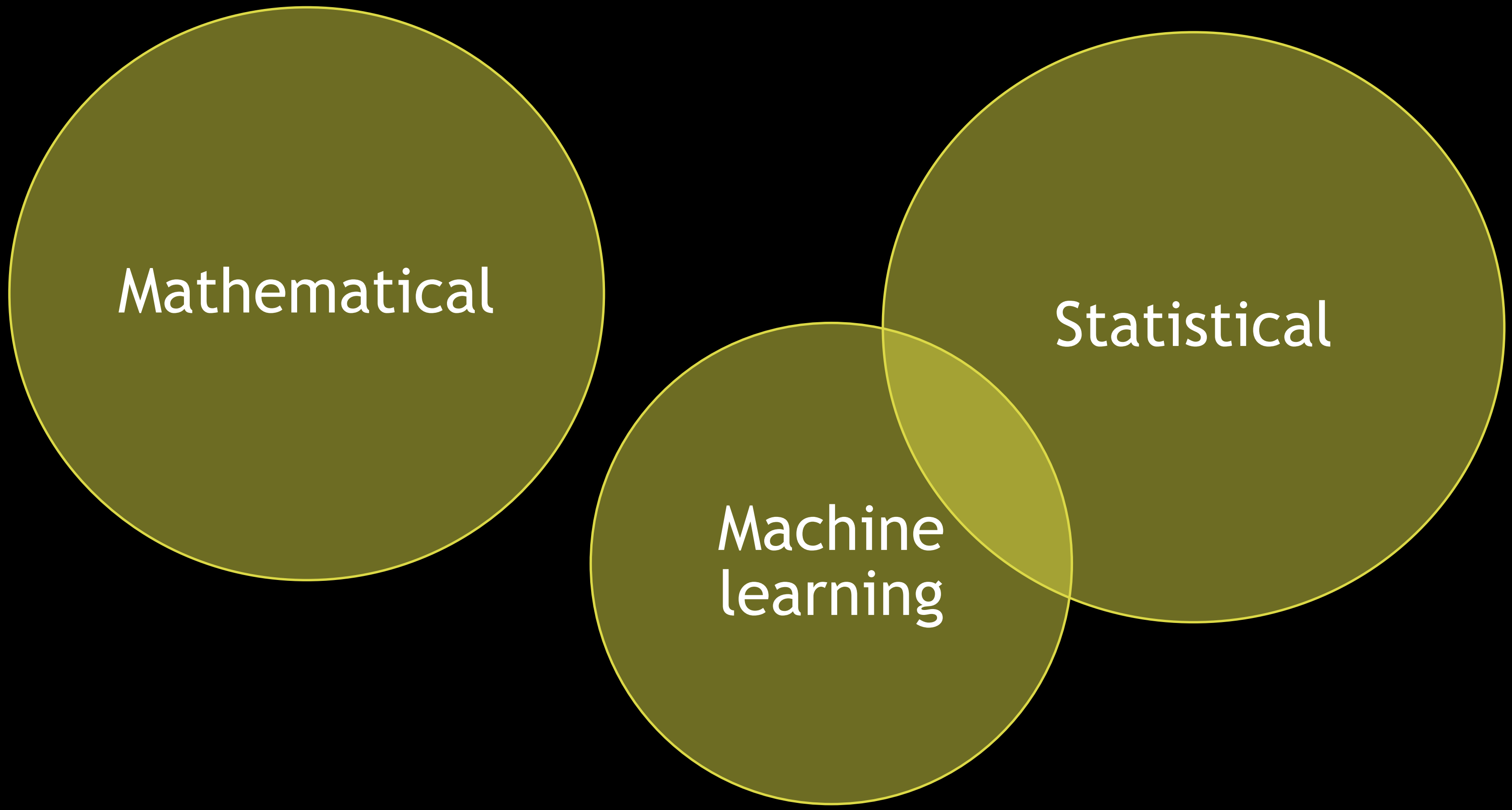
- Statistical models
  - Regression (many!), time series
- Mathematical models
  - Compartmental, mechanistic, agent-based
- Machine learning

- Statistical models
  - Regression (many!), time series
- Mathematical models
  - Compartmental, mechanistic, agent-based
- Machine learning
  - Uses algorithms and statistical models

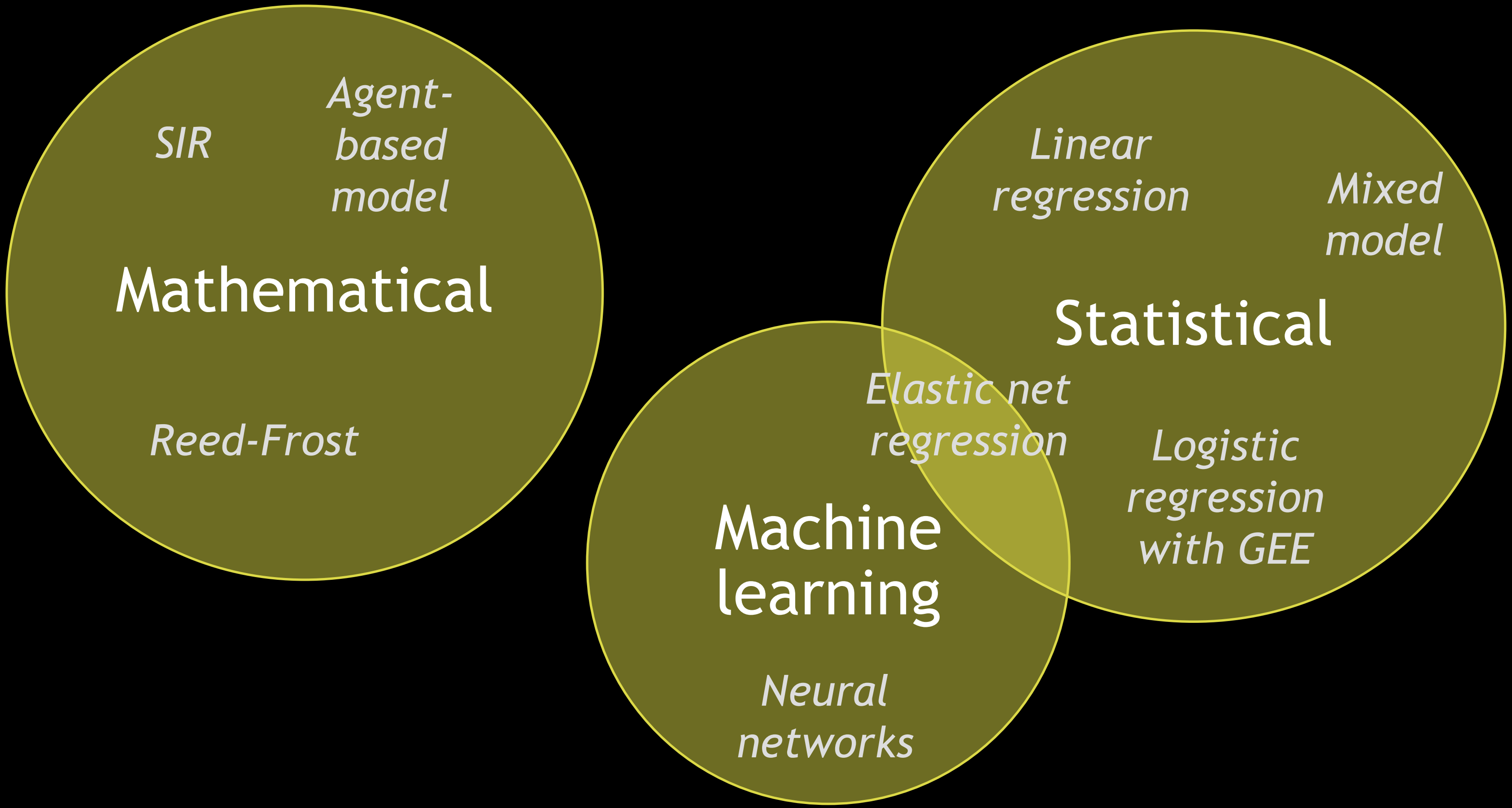
Three Major Model Types

(used in epidemiology)
(general modeling approaches)

Relationships

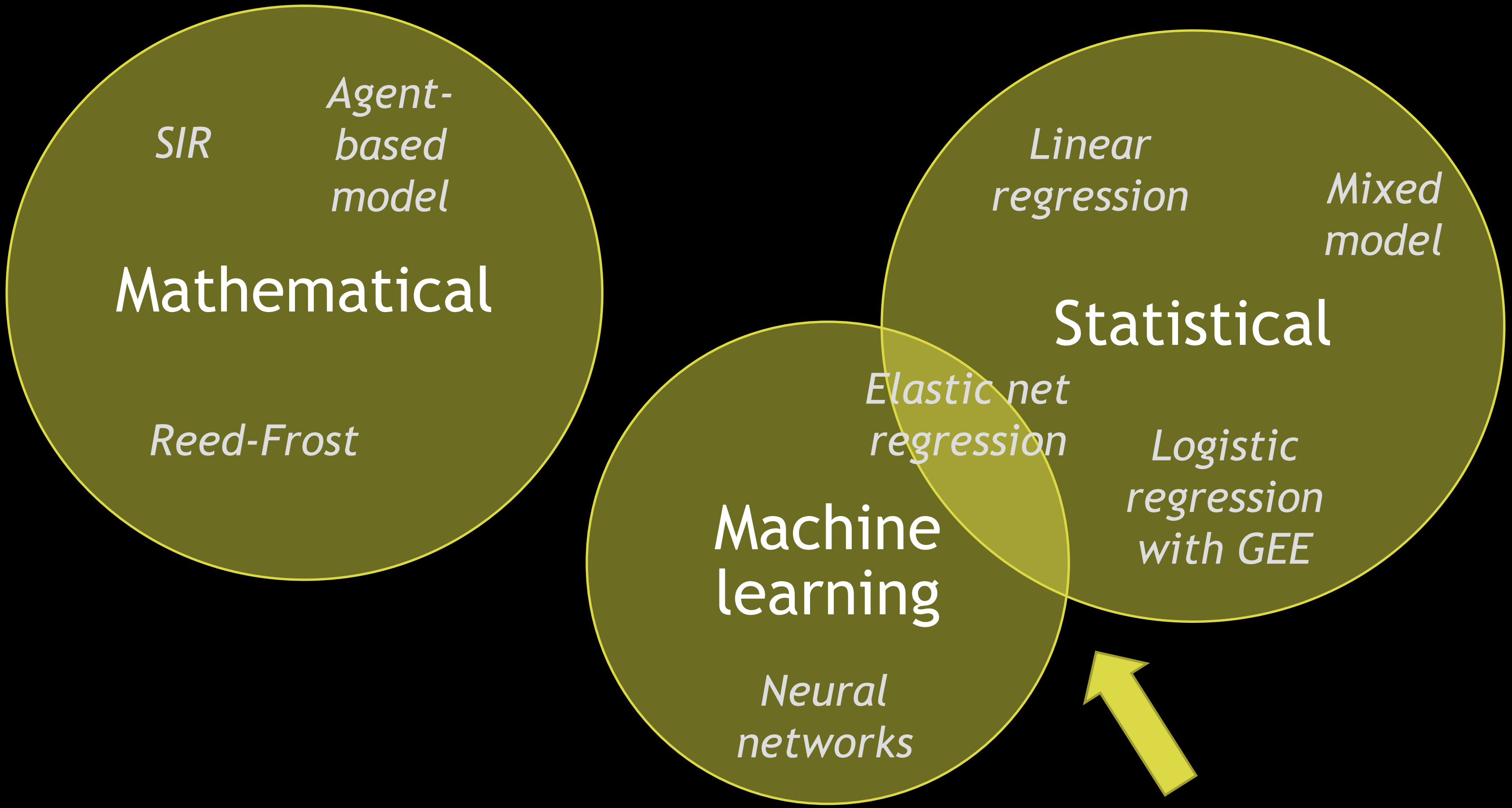


Relationships



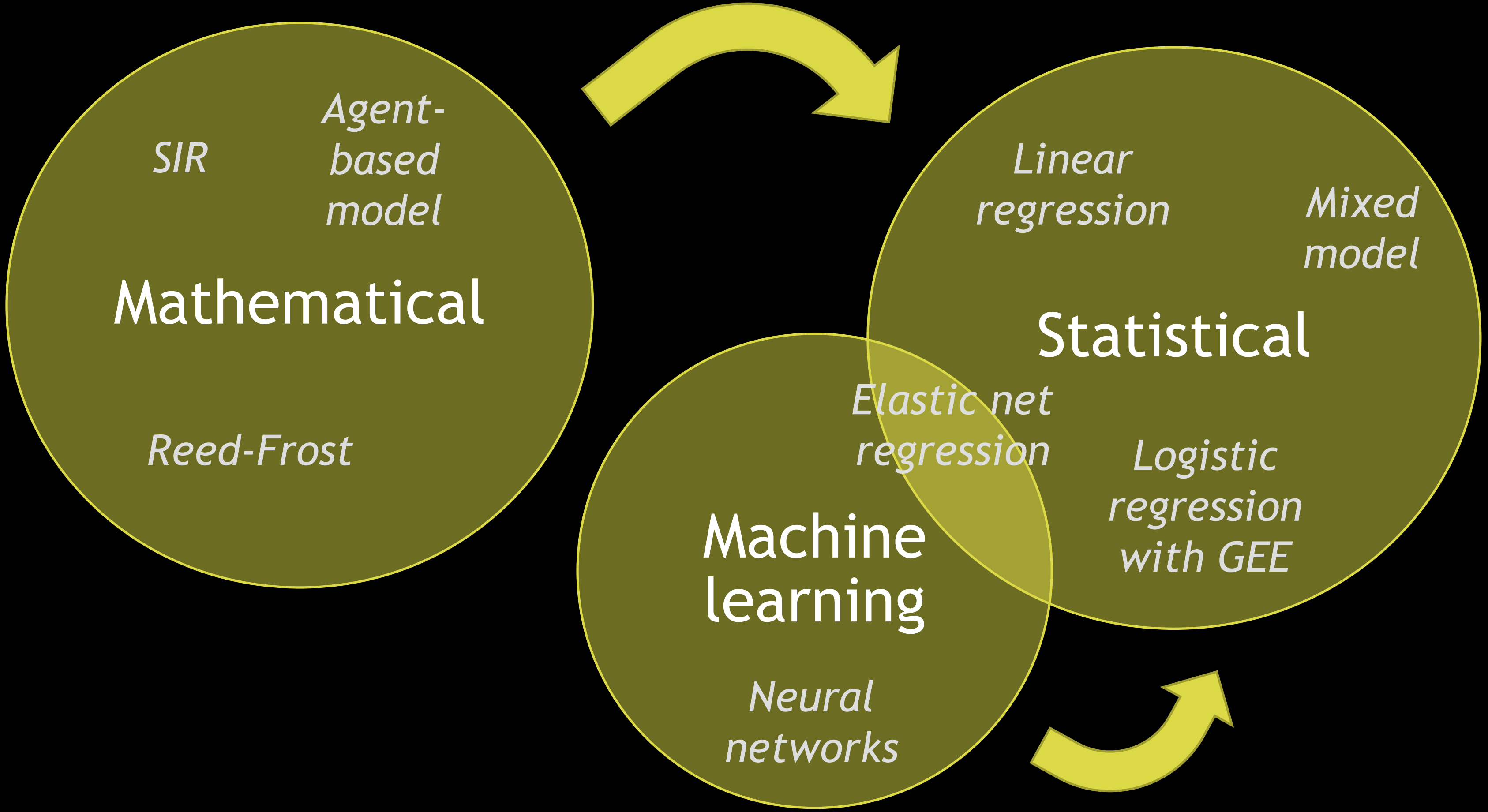
Relationships - Overlap

Same model, different context



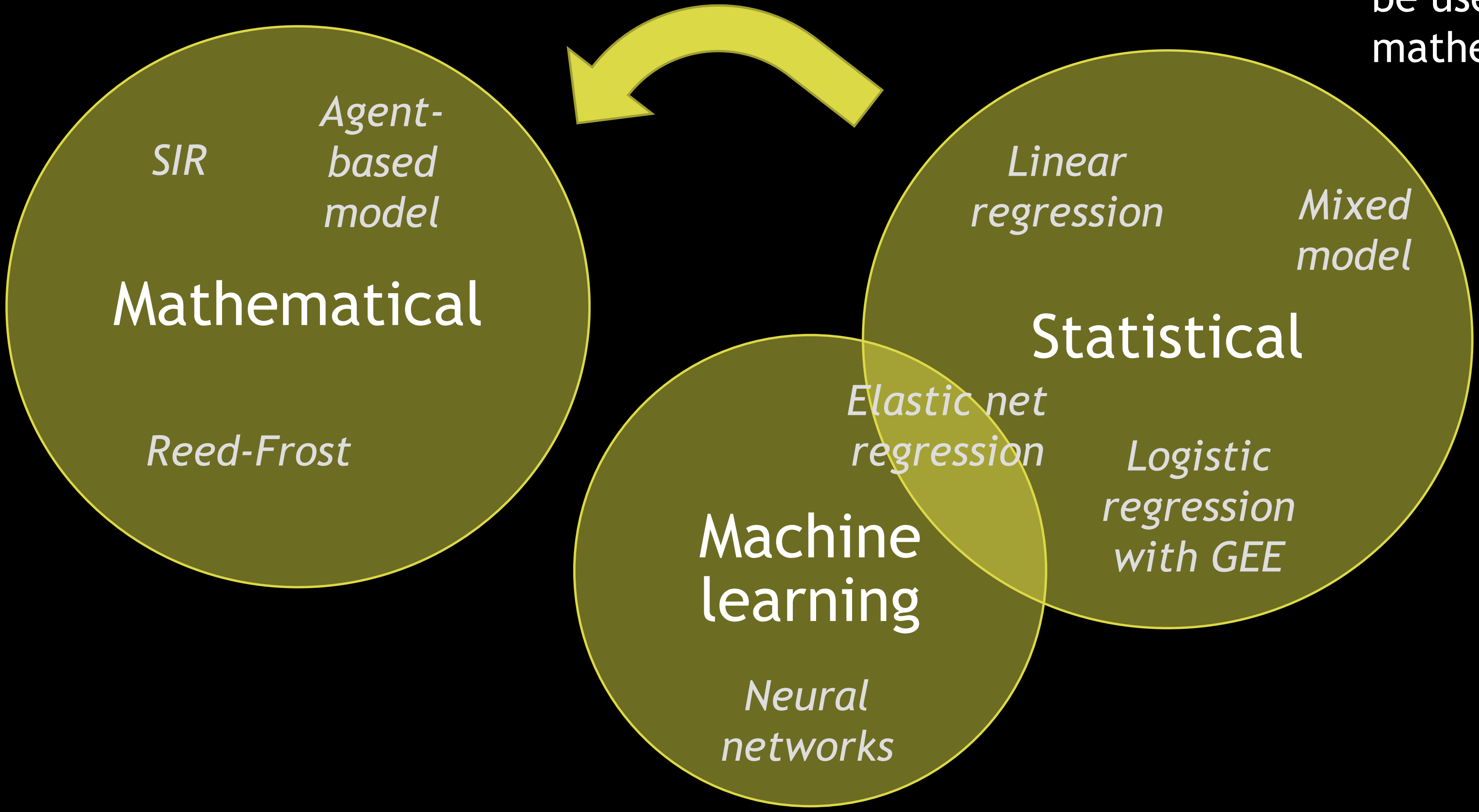
Relationships - Evaluation

Statistical methods often used to evaluate other models



Relationships – Parameter Estimates

Statistical models produce estimates to be used in mathematical models



Statistical Modeling

- Uses methods based in theory (frequentist, Bayesian, likelihood)
- Way to understand data and sources of its variability
- Concepts:
  - Our data are a sample from some larger population
  - Our data come from a distribution (parametric) or not (non-parametric)
- Generalization of statistical test

Statistical Modeling

Statistical Test	Statistical Model
t-test, ANOVA	multiple linear regression
Wilcoxon, Spearman, Kruskal-Wallis	proportional odds ordinal logistic model
Log-rank	Cox regression

Statistical Modeling

- We start with the data
- Our data and question of interest will determine the model we use
- For our model results, we want to have an unbiased estimate of the relationship between the inputs and the outcome
- A high quality statistical model will have limited biases and an appropriate model
  - Results should account for uncertainty
  - Hopefully uncertainty is limited

Statistical Modeling Questions

- What is the effect of temperature on mosquito abundance?
- Does vaccination affect disease risk?
- What will the life expectation be for people with certain characteristics?

Mathematical Modeling

- Formal description of a process we are interested in
- Model structure/form is critical first decision
  - Often depends on type of disease being studied
- Parameters/inputs for the model often come from previous knowledge or other literature
- Models commonly formulated with differential equations
- A high-quality mathematical model should examine multiple scenarios
  - Validated if possible
  - Results should account for uncertainty
  - Hopefully uncertainty is limited

Mathematical Modeling

- We start with our model: based on system processes
- Often, we use equations to create a scenario (set of conditions)
  - Example: population of city, level of interaction, certain disease, susceptibility of population
- If we can create a realistic model, then we can make changes to model and see what happens
  - Example: what happens when we vaccinate some people?
- This is similar to running many experiments and observing what happens
- Results from these observations can be compared to real data
  - Example: do model results match what happened in real life?

Mathematical Modeling Questions

- If we vaccinate the population, how many cases will occur?
- Why does the disease have a biennial pattern?
- Given the current population and disease conditions, how many deaths will occur?

Machine Learning

- Subset of artificial intelligence developed in computer science
- Using computers to develop predictions from data
- Need to make predictions in scenarios where statistical models were inappropriate
  - Image recognition
  - Handwriting recognition
  - Financial markets

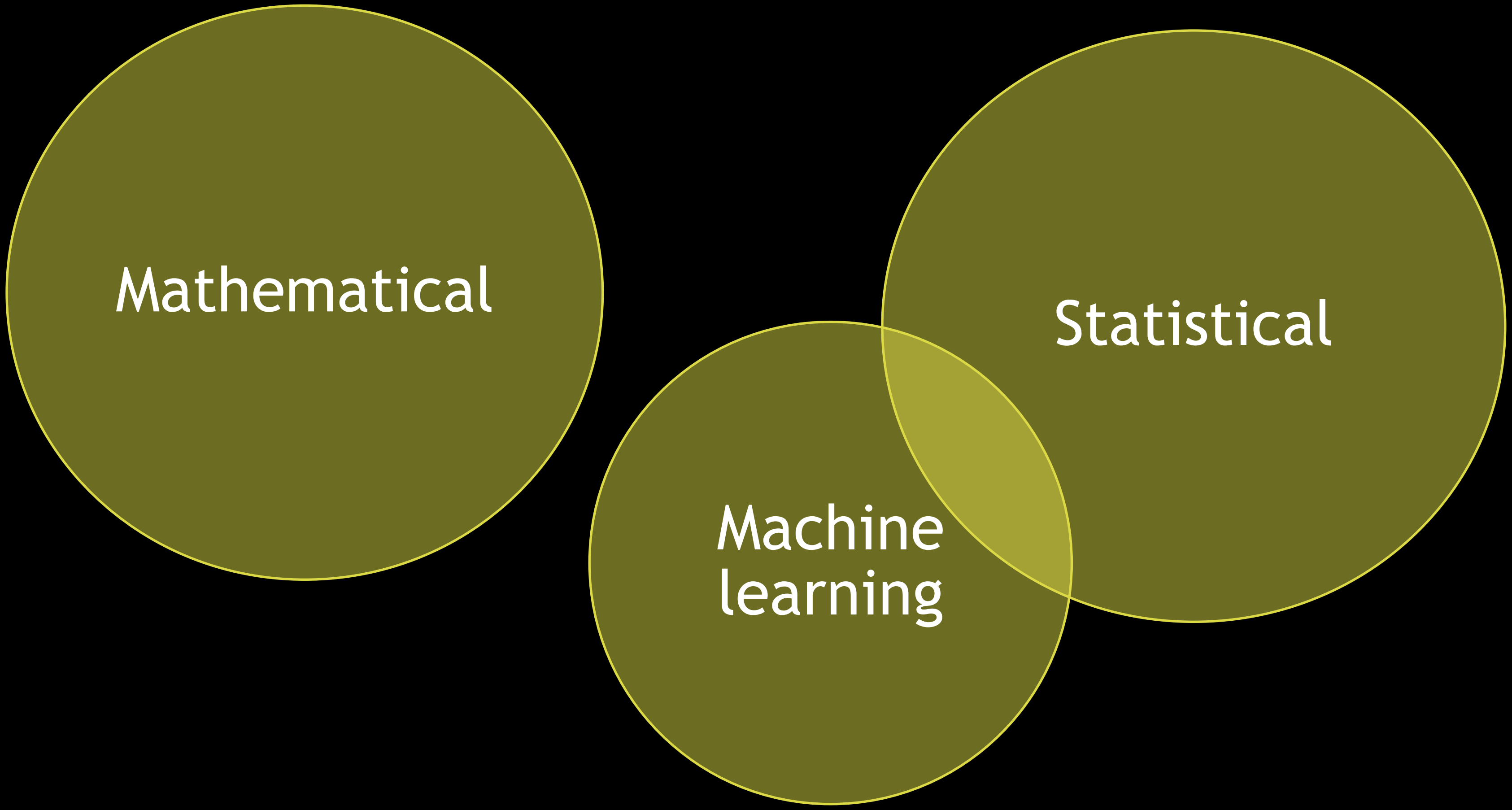
Machine Learning

- Methods may be very complex and difficult to understand, usually not interpretable
- Many models run repeatedly
- Focus on prediction results

Machine Learning Questions

- Will people with certain characteristics get the disease?
- What is the temperature in this location?
- What are the types of houses in this village?

Similarities & Differences



Similarities Between Modeling Types

Feature	SM	MM	ML
Model can be represented by an equation	Y	Y	S
Model uncertainty is important	Y	Y	Y
We are estimating parameters	Y	S	S
We care about the parameter values	Y	Y	N
We are making predictions	S	S	S
SM=Statistical Model, MM=Mathematical Model, ML=Machine Learning Y=Yes, S=Sometimes, N=No			

Differences Between Modeling Types

Feature	SM	MM	ML
Uses differential (or stochastic) equations	N	Y	N
Good for predictions	S	Y	Y
Single equation represents model	Y	S	S
Multiple equations represent model	N	Y	S
SM=Statistical Model, MM=Mathematical Model, ML=Machine Learning Y=Yes, S=Sometimes, N=No			

Statistical vs. Mathematical Models

- Statistical: we are finding the relationship between the variables that is the most likely explanation for the data we have (or the relationship with least uncertainty)
  - We are estimating the <u>size</u> and <u>direction</u> of the relationships
  - These are our parameters
- Mathematical: we assume relationships/model structure, make changes and calculations, then compare to real data (validation)
  - We are estimating or predicting outcome size
  - Our inputs are our parameters

Questions & Clarifications

Workshop Schedule

Time	Topics
~~2:00—2:10 pm~~	~~Outline & Introduction~~
~~2:10—3:00 pm~~	~~Defining Models & Modeling Terms~~
~~3:00—3:30 pm~~	~~Types of Models~~
3:30—3:40 pm	Break
3:40—4:00 pm	Modeling Goals & Questions
4:00—4:20 pm	Uncertainty in Models
4:20—4:40 pm	Interpreting & Evaluating Modeling Papers
4:40—5:00 pm	Questions & Discussion

What Can We Do with Models?

(and what can’t we do)

What is the Goal of Modeling?

- Inference
- Prediction

Model Use: Inference

- Estimate relationships between outcome and input data
  - Are inputs associated with the outcome?
  - How are inputs related to outcome?
- Understand how data were generated
  - Why are we observing a particular pattern of data?

Model Use: Inference

- Estimate relationships between outcome and input data
  - Are inputs associated with the outcome?
    - Are climate variables associated with COVID-19 rates?
  - How are inputs related to outcome?
    - How is population size related to measles incidence?
- Understand how data were generated
  - Why are we observing a particular pattern of data?
    - Why are gonorrhea levels staying at a low level?

Model Use: Prediction

- Predicting outcomes
  - Predict general outcomes
    -
  - Predicting outcomes for an individual
    -
  - Predict impact of controls
    -
- Forecasting
  - Predicting what will happen in the future
    -

Model Use: Prediction

- Predicting outcomes
  - Predict general outcomes
    - What is the average effect of treatment on population?
  - Predicting outcomes for an individual
    - Will a person with certain characteristics get the disease?
  - Predict impact of controls
    - How will vaccination affect the number of cases?
- Forecasting
  - Predicting what will happen in the future
    - How many cases will there be in total?

Comparison

Inference Model

- Carefully consider variables
- Interpretable
- Previous knowledge to inform relationships
- Generalizable

Prediction Model

- Can include many variables
- Accurate predictions
- Model selection
- Generalizable
- Validation critical

Inference or Prediction?

- Select your responses to the research questions in the Google Form
- Is the goal inference or prediction?

Inference or Prediction?

- Select your responses to the research questions in the Google Form
- Is the goal inference or prediction?

Question	Goal	Note
What is the effect of temperature on mosquito abundance?	Inference	Estimating effects are a common inferential question
If we vaccinate the population, how many cases will occur?	Prediction	“What-if” scenarios are common prediction goals.
What is the temperature in this new location?	Prediction	
Given the current population and disease conditions, how many deaths will occur?	Prediction	

Inference or Prediction?

- Select your responses to the research questions in the Google Form
- Is the goal inference or prediction?

Question	Goal	Note
What is the life expectancy for people with certain characteristics?	Prediction or Inference	If we estimate the effect of characteristics on life expectancy, that would be inference. But we could also build a prediction model for life expectancy based on characteristics.
Does vaccination affect disease risk?	Inference	Estimating an effect!

Inference or Prediction?

- Select your responses to the research questions in the Google Form
- Is the goal inference or prediction?

Question	Goal	Note
Why does the disease have a biennial pattern?	Inference	
What are the types of houses in this village?	Prediction	
Will people with certain characteristics get the disease?	Prediction	

Questions & Clarifications

Workshop Schedule

Time	Topics
~~2:00—2:10 pm~~	~~Outline & Introduction~~
~~2:10—3:00 pm~~	~~Defining Models & Modeling Terms~~
~~3:00—3:30 pm~~	~~Types of Models~~
3:30—3:40 pm	~~Break~~
3:40—4:00 pm	~~Modeling Goals & Questions~~
4:00—4:20 pm	Uncertainty in Models
4:20—4:40 pm	Interpreting & Evaluating Modeling Papers
4:40—5:00 pm	Questions & Discussion

Uncertainty in Models

- Almost all models will include a measure of uncertainty for their estimates or predictions

Where Does Uncertainty Come From?

- Systematic errors
  - Inaccuracies from limits in our ability to take measurements
  - Also called “observational stochasticity”
  - Measurement error, selection bias
- Random errors
  - Many parts of life are random
  - This is a natural part of the world
  - Also called “process stochasticity”
  - Unpredictable by definition

Where Does Uncertainty Come From?

Systematic Errors

- Data entry error
- Using poor-quality test
- Having an interviewer who is rude to subjects

Random Errors

- Whether a suspected contact answers the phone
- Viral transmission
- Meeting 50 people instead of 10

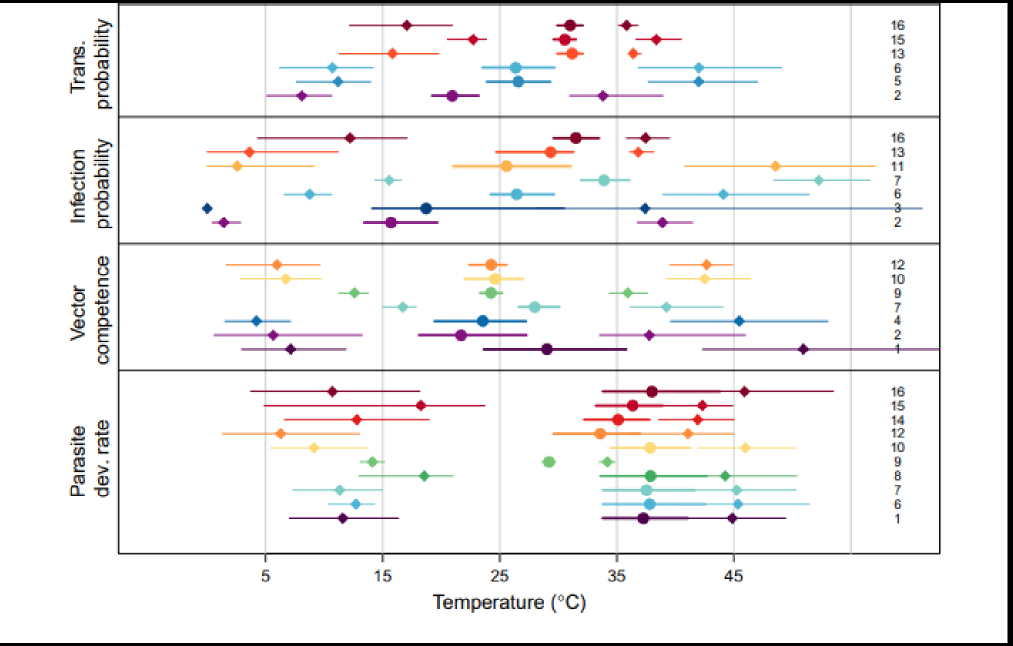
Uncertainty in Models

- Usually both are present
- We try to minimize uncertainty (usually cannot eliminate)
    - Minimizing systematic error
    - Minimizing random error

Uncertainty in Models

- Usually both are present
- We try to minimize uncertainty (usually cannot eliminate)
  - Minimizing systematic error
    - Careful measurements: good tools, capturing all cases
    - Appropriate study design, correct model
  - Minimizing random error
    - Increase sample size: more data points, more simulations of model
    - Repeating measurements: more measures more likely to be correct

Representing Uncertainty



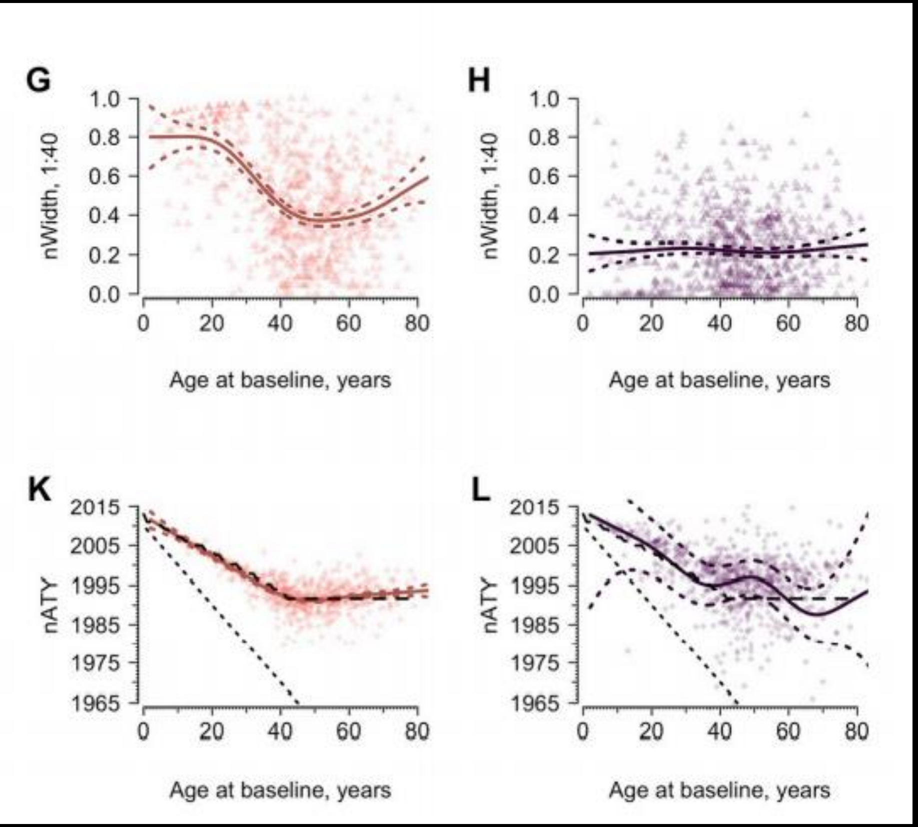
- Model results should include estimate/prediction and an interval
  - Confidence interval
  - Credible interval

Representing Uncertainty

Source	Location	Participants, No.	Participants with SARS-CoV-2 infection, No.	SAR (95% CI)
Household contacts				
Boscolo-Rizzo et al,[24] 2020	Treviso Province, Italy	121	54	0.45 (0.36-0.54)
Patel et al,[53] 2020	London, UK	185	79	0.43 (0.36-0.50)
Rosenberg et al,[55] 2020	New York, US	343	131	0.38 (0.33-0.43)
Dattner et al,[29] 2020	Bnei Brak, Israel	2824	981	0.35 (0.33-0.37)
Lopez Bernal et al,[47] 2020	UK	472	161	0.34 (0.30-0.38)
Wu et al,[65] 2020	Zhuhai, China	148	48	0.32 (0.25-0.40)
Wang et al,[61] 2020	Wuhan, China	155	47	0.30 (0.23-0.38)
Teherani et al,[59] 2020	Atlanta, US	108	31	0.29 (0.21-0.38)
Lewis et al,[44] 2020	Utah and Wisconsin, US	188	52	0.28 (0.21-0.34)
Dawson et al,[30] 2020	Wisconsin, US	64	16	0.25 (0.15-0.36)
Wang et al,[63] 2020	Beijing, China	335	77	0.23 (0.19-0.28)
Han,[35] 2020	South Korea	14	3	0.21 (0.03-0.47)
Böhmer et al,[23] 2020	Bavaria, Germany	24	5	0.21 (0.07-0.40)

- Model results should include estimate/prediction and an interval
  - Confidence interval
  - Credible interval
  - Usually at 95%

Representing Uncertainty



- Model results should include estimate/prediction and an interval
  - Confidence interval
  - Credible interval
  - Usually at 95%
- Narrower interval indicates higher certainty about model estimate/prediction

Questions & Clarifications

Workshop Schedule

Time	Topics
~~2:00—2:10 pm~~	~~Outline & Introduction~~
~~2:10—3:00 pm~~	~~Defining Models & Modeling Terms~~
~~3:00—3:30 pm~~	~~Types of Models~~
3:30—3:40 pm	~~Break~~
3:40—4:00 pm	~~Modeling Goals & Questions~~
~~4:00—4:20 pm~~	~~Uncertainty in Models~~
4:20—4:40 pm	Interpreting & Evaluating Modeling Papers
4:40—5:00 pm	Questions & Discussion

Why Are We Reading Papers?

Interpreting & Evaluating Papers

- What is the research question?
  - Predict, describe, determine, estimate

Here, we describe paired antibody profiles measured at two time points (baseline from 2009 to 2011 and follow-up from 2014 to 2015), roughly four years apart, in a large sample of individuals from an ongoing cohort study in Guangzhou, Guangdong Province, China [18]. We measured immune responses to multiple chronologically ordered H3N2 influenza strains (referred to as antibody profiles) that represent the history of H3N2 circulation in humans since its emergence in 1968. We aim to determine how those profiles vary across individuals and between study visits, and to test if there exist features of these antibody profiles that are more predictive of the odds of seroconversion to recently circulating strains than homologous titers only.

Interpreting & Evaluating Papers

Here, we describe paired antibody profiles measured at two time points (baseline from 2009 to 2011 and follow-up from 2014 to 2015), roughly four years apart, in a large sample of individuals from an ongoing cohort study in Guangzhou, Guangdong Province, China [18]. We measured immune responses to multiple chronologically ordered H3N2 influenza strains (referred to as antibody profiles) that represent the history of H3N2 circulation in humans since its emergence in 1968. We aim to determine how those profiles vary across individuals and between study visits, and to test if there exist features of these antibody profiles that are more predictive of the odds of seroconversion to recently circulating strains than homologous titers only.

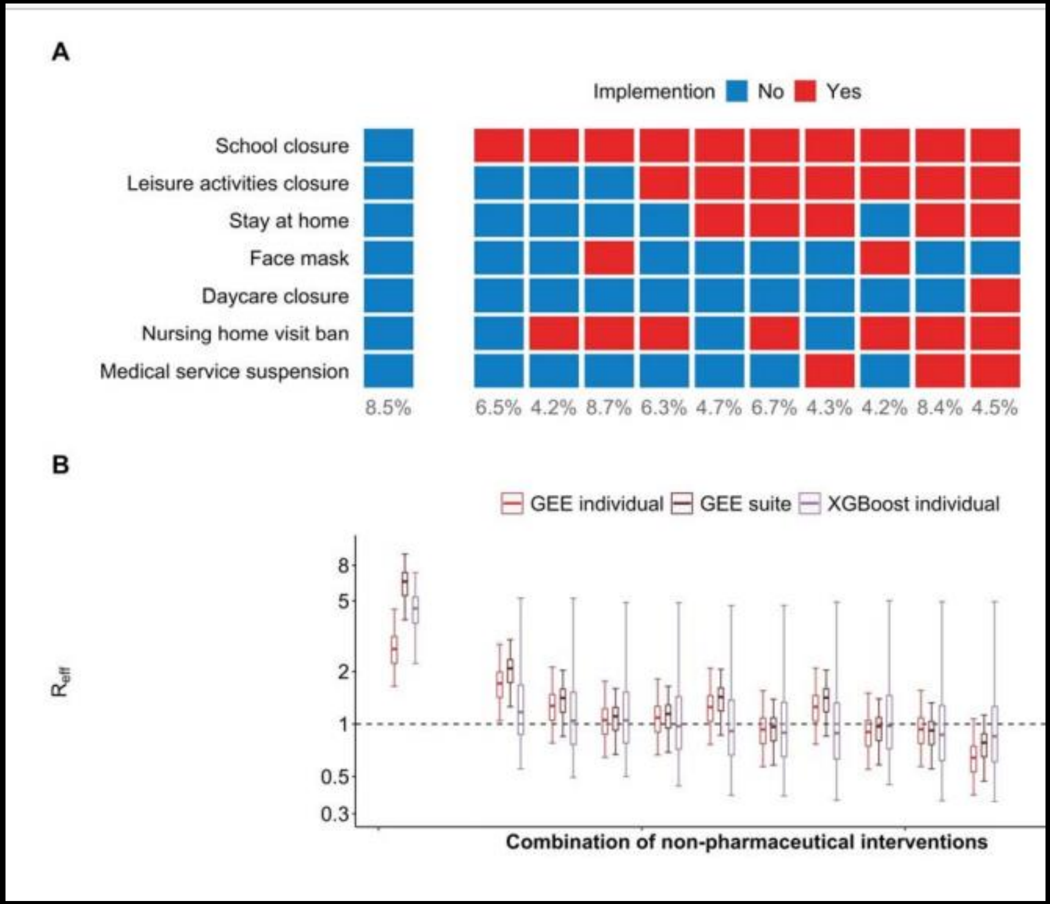
- What is the research question?
  - Predict, describe, determine, estimate
- Three questions/goals
  - Describe antibody profiles
  - Determine profile variability by individual
  - Can we use profiles to predict infection risk?

Interpreting & Evaluating Papers

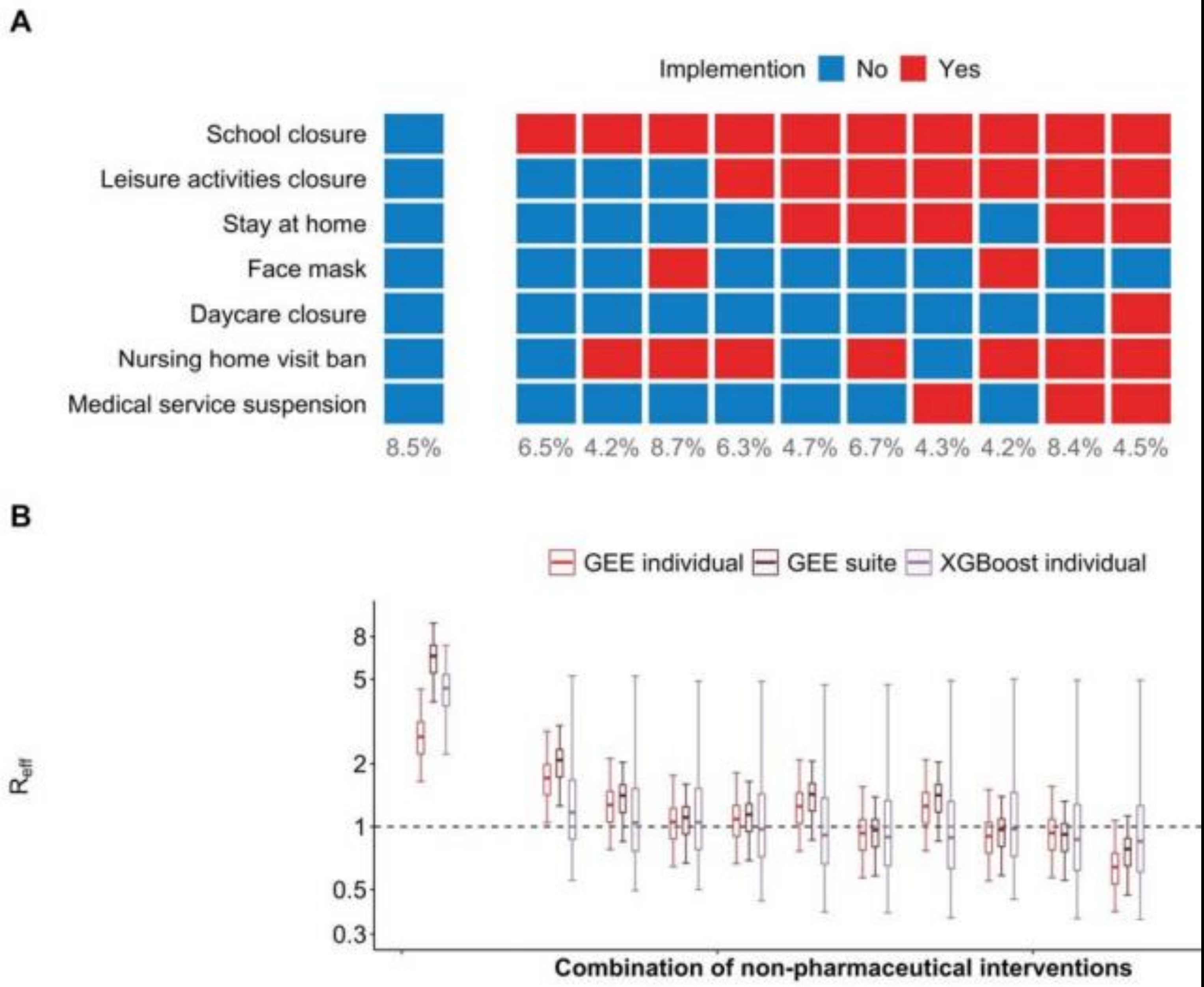
- What is the research question?
- What is the method?
  - (Optional)
  - Fitting, built, using
  - Model comparison

Effects of pre-existing immunity on seroconversion to recent strains. We predicted the odds of seroconversion ($c_{i,j}$) to one of four recent strains i (i.e. A/HongKong/2014, A/Texas/2012, A/Victoria/2009 or A/Perth/2009) by fitting logistic regression with predictors that

Interpreting & Evaluating Papers



- What is the research question?
- What is the method?
- What are the results?


A
Implemention No Yes
School closure
Leisure activities closure
Stay at home
Face mask
Daycare closure
Nursing home visit ban
Medical service suspension
8.5%
6.5% 4.2% 8.7% 6.3% 4.7% 6.7% 4.3% 4.2% 8.4% 4.5%
B
GEE individual GEE suite XGBoost individual
R_{eff}
8
5
2
1
0.5
0.3
Combination of non-pharmaceutical interventions

Interpreting & Evaluating Papers

- What is the research question?
- What is the method?
- What are the results?
- Do the results seem reasonable?
  - Too good to be true?
  - Reputable journal?
  - Re-examine methods
  - Discussion section

HOW TO READ SCIENTIFIC PAPERS

Much of a scientist's work involves reading research papers. Because scientific articles are different from other texts, like novels or newspaper stories, they should be read differently. Here are some tips to be able to read and understand them.

1 SKIM

First get the "big picture" by reading the title, key words and abstract carefully; this will tell you the major findings and why they matter.

- Quickly scan the article without taking notes; focus on headings and subheadings.
- Note the publishing date; for many areas, current research is more relevant.
- Note any terms and parts you don't understand for further reading.

RE-READ 2

Read the article again, asking yourself questions such as:

- What problem is the study trying to solve?
- Are the findings well supported by evidence?
- Are the findings unique and supported by other work in the field?
- What was the sample size? Is it representative of the larger population?
- Is the study repeatable?
- What factors might affect the results?

If you are unfamiliar with key concepts, look for them in the literature.

3 INTERPRET

- Examine graphs and tables carefully.
- Try to interpret data first before looking at captions.
- When reading the discussion and results, look for key issues and new findings.
- Make sure you have distinguished the main points. If not, go over the text again.

SUMMARIZE 4

- Take notes; it improves reading comprehension and helps you remember key points.
- If you have a printed version, highlight key points and write on the article. If it's on screen, make use of markers and comments.

Practice practice practice!

Read what interests you, read with colleagues!

We will practice looking at papers throughout the workshop!

If you find a paper, send it to me and we can go over it together or in class!

Workshop Schedule

Time	Topics
~~2:00—2:10 pm~~	~~Outline & Introduction~~
~~2:10—3:00 pm~~	~~Defining Models & Modeling Terms~~
~~3:00—3:30 pm~~	~~Types of Models~~
3:30—3:40 pm	~~Break~~
3:40—4:00 pm	~~Modeling Goals & Questions~~
~~4:00—4:20 pm~~	~~Uncertainty in Models~~
~~4:20—4:40 pm~~	~~Interpreting & Evaluating Modeling Papers~~
4:40—5:00 pm	Questions & Discussion

Questions & Discussion

Additional Materials

- Review with readings I sent (Statistics Explained & MathMod Explained)
  - Send email/WhatsApp if you have questions!
- To be sent:
  - Instructions to install R and Rstudio
  - Another paper to help with review (Leek & Peng)
  - Updated syllabus (when dates are set)