Model evaluation and comparison

Jonathan Dushoff, McMaster University

MMED 2019

Goals

- Discuss model types and model goals

Goals

- Discuss model types and model goals
- Discuss the value of simulation for validating models

Goals

- Discuss model types and model goals
- Discuss the value of simulation for validating models
- Discuss metrics for evaluating fit

Goals

- Discuss model types and model goals
- Discuss the value of simulation for validating models
- Discuss metrics for evaluating fit
  - Put the Goodness of fit test in its place

Goals

- Discuss model types and model goals
- Discuss the value of simulation for validating models
- Discuss metrics for evaluating fit
  - Put the Goodness of fit test in its place

Do I have a good model?

- What is my model trying to accomplish?

Do I have a good model?

- What is my model trying to accomplish?
  - Generating hypotheses

Do I have a good model?

- What is my model trying to accomplish?
  - Generating hypotheses
  - Evaluating plausibility

Do I have a good model?

- What is my model trying to accomplish?
  - Generating hypotheses
  - Evaluating plausibility
  - Prediction

Do I have a good model?

- What is my model trying to accomplish?
  - Generating hypotheses
  - Evaluating plausibility
  - Prediction
  - Extrapolation

Do I have a good model?

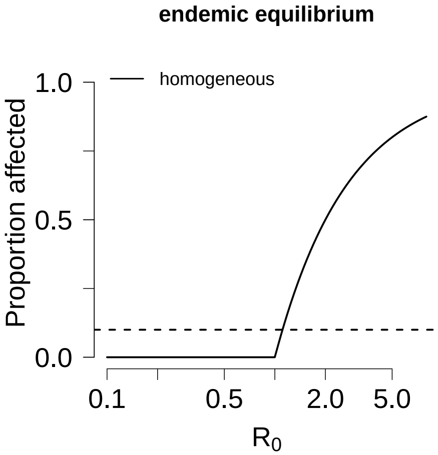
- What is my model trying to accomplish?
  - Generating hypotheses
  - Evaluating plausibility
  - Prediction
  - Extrapolation
  - Mechanistic understanding

Do I have a good model?

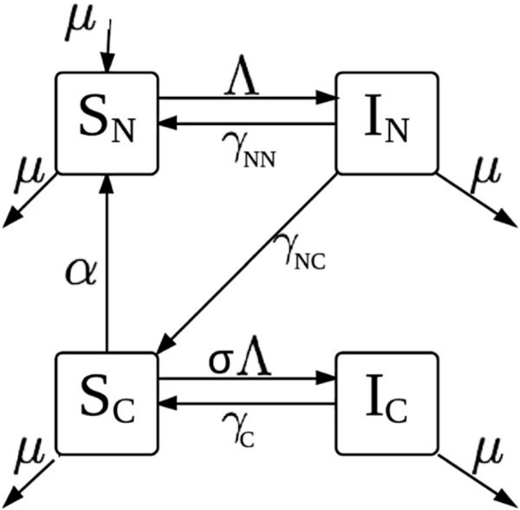
- What is my model trying to accomplish?
  - Generating hypotheses
  - Evaluating plausibility
  - Prediction
  - Extrapolation
  - Mechanistic understanding

Outline

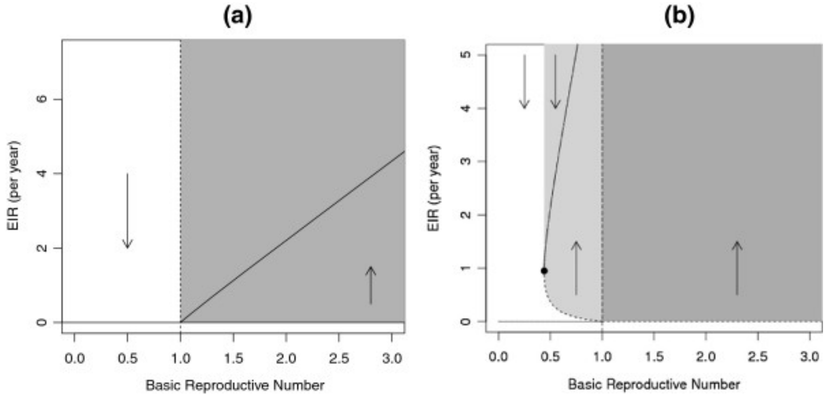
Disease thresholds



Effects of clinical immunity

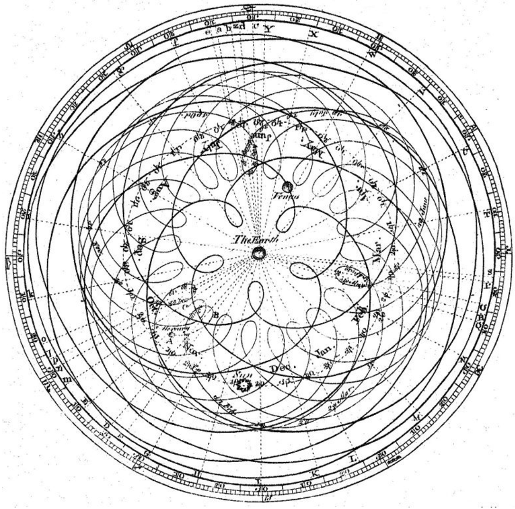


Bistability

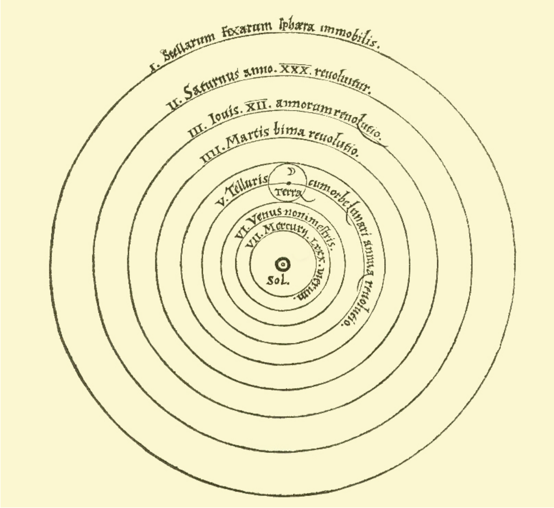


Outline

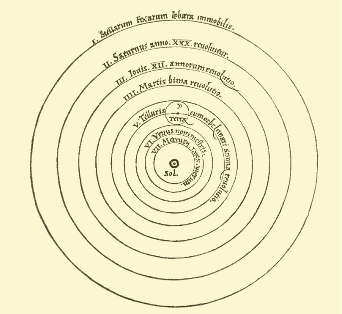
Ptolemy v. Copernicus

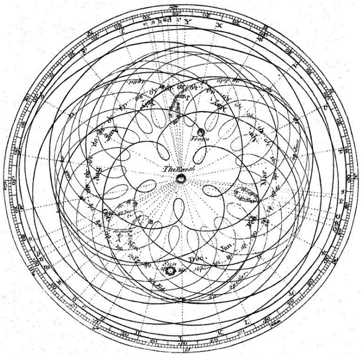


Ptolemy v. Copernicus

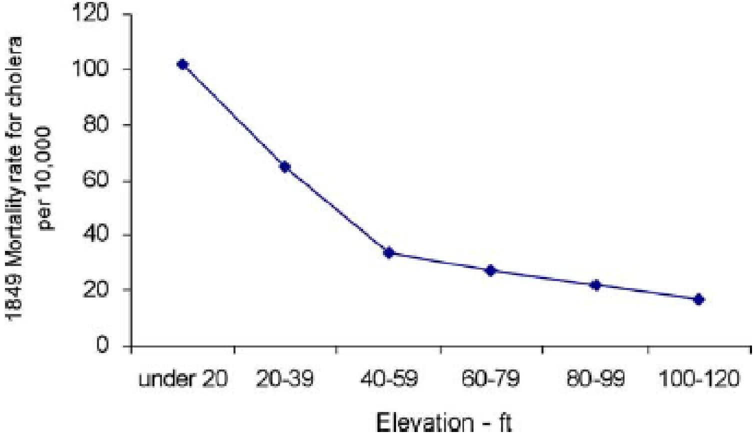


Ptolemy v. Copernicus





Where will we see cholera cases?

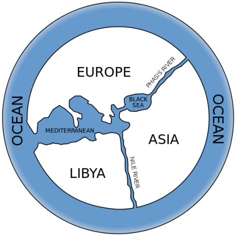


Where will we see cholera cases?

Outline

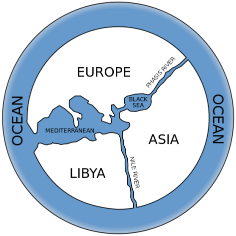
Model Validation

- Does your fitting algorithm match your *model world*?



Model Validation

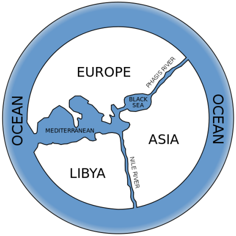
- Does your fitting algorithm match your *model world*?



- If you use your fitting algorithm on simulations from your model world, then you *know the right answer*!

Model Validation

- Does your fitting algorithm match your *model world*?



- If you use your fitting algorithm on simulations from your model world, then you *know the right answer*!

Validation measures

- Coverage

Validation measures

- Coverage
- Precision

Validation measures

- Coverage
- Precision
- Bias?

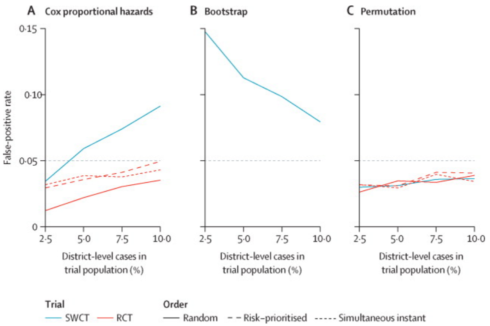
Validation measures

- Coverage
- Precision
- Bias?
- Accuracy?

Validation measures

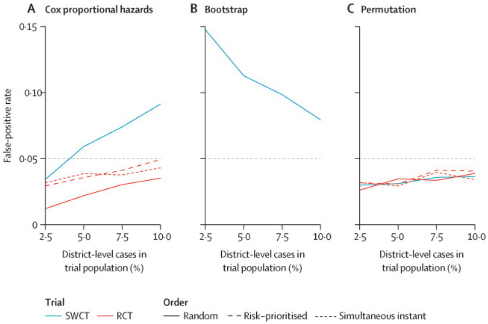
- Coverage
- Precision
- Bias?
- Accuracy?

Coverage



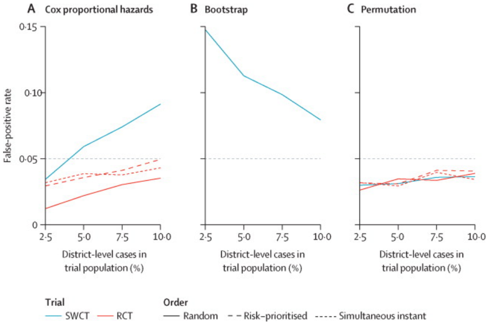
- The right answer should be inside your 95% confidence interval 95% of the time

Coverage



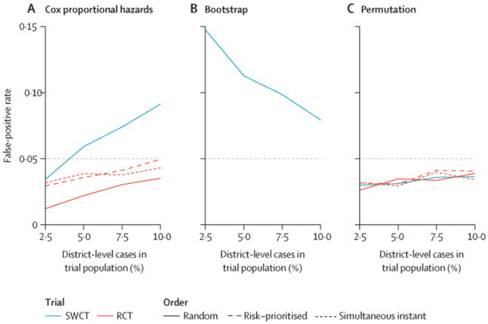
- The right answer should be inside your 95% confidence interval 95% of the time
  - If more, your model is *too conservative*

Coverage



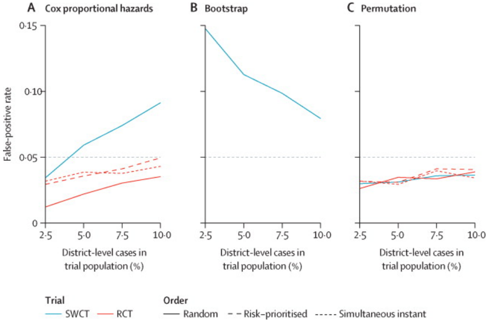
- The right answer should be inside your 95% confidence interval 95% of the time
  - If more, your model is *too conservative*
  - If less, your model is *invalid*

Coverage



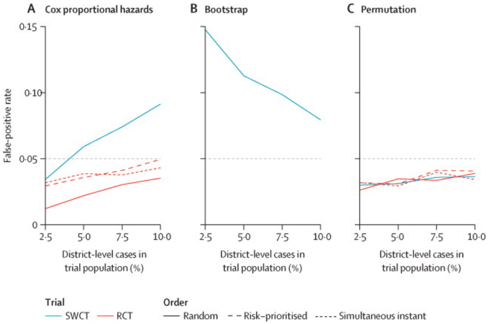
- The right answer should be inside your 95% confidence interval 95% of the time
  - If more, your model is *too conservative*
  - If less, your model is *invalid*
- In many cases it's good to look at the two tails separately:

Coverage



- The right answer should be inside your 95% confidence interval 95% of the time
  - If more, your model is *too conservative*
  - If less, your model is *invalid*
- In many cases it's good to look at the two tails separately:
  - How often do you overestimate? Underestimate?

Coverage



- The right answer should be inside your 95% confidence interval 95% of the time
  - If more, your model is *too conservative*
  - If less, your model is *invalid*
- In many cases it's good to look at the two tails separately:
  - How often do you overestimate? Underestimate?

Precision

- You should aim to make your confidence intervals as narrow as possible

Precision

- You should aim to make your confidence intervals as narrow as possible
  - Provide as much information as possible

Precision

- You should aim to make your confidence intervals as narrow as possible
  - Provide as much information as possible
- As data increases, your precision should increase

Precision

- You should aim to make your confidence intervals as narrow as possible
  - Provide as much information as possible
- As data increases, your precision should increase
  - CIs should approach zero width

Precision

- You should aim to make your confidence intervals as narrow as possible
  - Provide as much information as possible
- As data increases, your precision should increase
  - CIs should approach zero width

Bias?

- Nobody wants to be biased

Bias?

- Nobody wants to be biased
  - You shouldn't ignore sources of bias (confounders)

Bias?

- Nobody wants to be biased
  - You shouldn't ignore sources of bias (confounders)
  - Estimators with good coverage and good precision are *asymptotically* unbiased

Bias?

- Nobody wants to be biased
  - You shouldn't ignore sources of bias (confounders)
  - Estimators with good coverage and good precision are *asymptotically* unbiased
- Not so clear you need to be *absolutely* unbiased

Bias?

- Nobody wants to be biased
  - You shouldn't ignore sources of bias (confounders)
  - Estimators with good coverage and good precision are *asymptotically* unbiased
- Not so clear you need to be *absolutely* unbiased
  - Bias is the difference between the *mean* expected prediction and the true value

Bias?

- Nobody wants to be biased
  - You shouldn't ignore sources of bias (confounders)
  - Estimators with good coverage and good precision are *asymptotically* unbiased
- Not so clear you need to be *absolutely* unbiased
  - Bias is the difference between the *mean* expected prediction and the true value
- Scale dependence

Bias?

- Nobody wants to be biased
  - You shouldn't ignore sources of bias (confounders)
  - Estimators with good coverage and good precision are *asymptotically* unbiased
- Not so clear you need to be *absolutely* unbiased
  - Bias is the difference between the *mean* expected prediction and the true value
- Scale dependence
  - an unbiased estimate of γ is automatically a biased estimate of D (but not asymptotically biased)

Bias?

- Nobody wants to be biased
  - You shouldn't ignore sources of bias (confounders)
  - Estimators with good coverage and good precision are *asymptotically* unbiased
- Not so clear you need to be *absolutely* unbiased
  - Bias is the difference between the *mean* expected prediction and the true value
- Scale dependence
  - an unbiased estimate of γ is automatically a biased estimate of D (but not asymptotically biased)
  - Medians (instead of means) can get avoid scale problems

Bias?

- Nobody wants to be biased
  - You shouldn't ignore sources of bias (confounders)
  - Estimators with good coverage and good precision are *asymptotically* unbiased
- Not so clear you need to be *absolutely* unbiased
  - Bias is the difference between the *mean* expected prediction and the true value
- Scale dependence
  - an unbiased estimate of γ is automatically a biased estimate of D (but not asymptotically biased)
  - Medians (instead of means) can get avoid scale problems

Accuracy?

- Nobody wants to be inaccurate

Accuracy?

- Nobody wants to be inaccurate
- Good coverage and good precision should guarantee good accuracy

Accuracy?

- Nobody wants to be inaccurate
- Good coverage and good precision should guarantee good accuracy

Outline

Conceptual models

Prediction

Model Validation

Model Evaluation
- Goodness of fit
- Capturing patterns
- Going beyond

Conclusion

Model Evaluation

- Does your model match the *real world*?

Model Evaluation

- Does your model match the *real world*?
  - *

Model Evaluation

- Does your model match the *real world*?
  - * No!

Model Evaluation

- Does your model match the *real world*?
  - * No!
- How well does your model match the real world?

ICI3D

Model Evaluation

- Does your model match the *real world*?
  - * No!
- How well does your model match the real world?

Outline

Goodness of fit

- Goodness of fit *statistics* describe how well a model prediction matches observed data

Goodness of fit

- Goodness of fit *statistics* describe how well a model prediction matches observed data
- Goodness of fit *tests* attempt to determine whether the observed difference between model and data is statistically significant

Goodness of fit

- Goodness of fit *statistics* describe how well a model prediction matches observed data
- Goodness of fit *tests* attempt to determine whether the observed difference between model and data is statistically significant

Your model is false!

- A goodness of fit test won't make it true

Your model is false!

- A goodness of fit test won't make it true
- You can "pass" a goodness of fit test by:

Your model is false!

- A goodness of fit test won't make it true
- You can "pass" a goodness of fit test by:
  - having a good model

Your model is false!

- A goodness of fit test won't make it true
- You can "pass" a goodness of fit test by:
  - having a good model
  - making very broad predictions

Your model is false!

- A goodness of fit test won't make it true
- You can "pass" a goodness of fit test by:
  - having a good model
  - making very broad predictions
  - having bad data

Your model is false!

- A goodness of fit test won't make it true
- You can "pass" a goodness of fit test by:
  - having a good model
  - making very broad predictions
  - having bad data
  - choosing an inappropriate way to compare

Your model is false!

- A goodness of fit test won't make it true
- You can "pass" a goodness of fit test by:
  - having a good model
  - making very broad predictions
  - having bad data
  - choosing an inappropriate way to compare
- So why would we do this?

Your model is false!

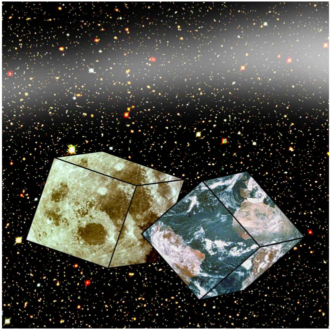
- A goodness of fit test won't make it true
- You can "pass" a goodness of fit test by:
  - having a good model
  - making very broad predictions
  - having bad data
  - choosing an inappropriate way to compare
- So why would we do this?
- For that matter, why do we use P values at all in biology?

Your model is false!

- A goodness of fit test won't make it true
- You can "pass" a goodness of fit test by:
  - having a good model
  - making very broad predictions
  - having bad data
  - choosing an inappropriate way to compare
- So why would we do this?
- For that matter, why do we use P values at all in biology?

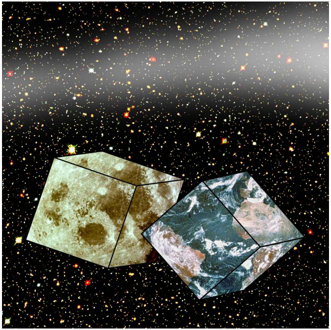
Passing goodness of fit tests

- I can make any model pass a goodness of fit test by broadening the uncertainty



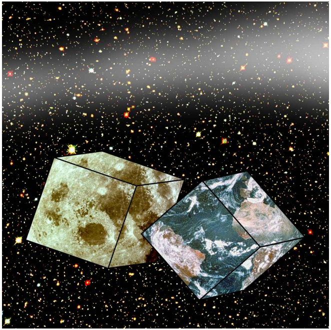
Passing goodness of fit tests

- I can make any model pass a goodness of fit test by broadening the uncertainty
- That doesn't make it a good model

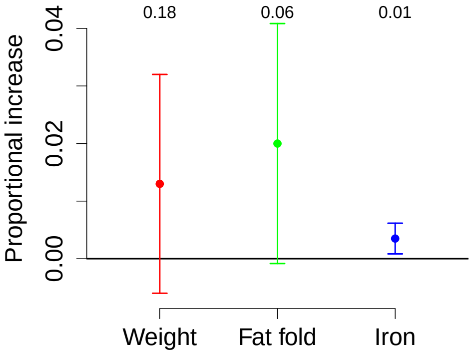


Passing goodness of fit tests

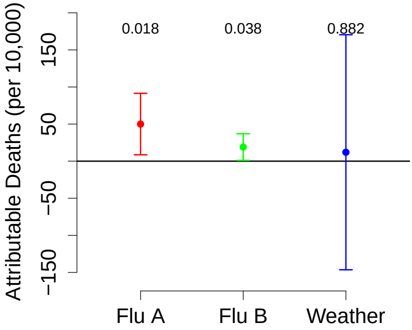
- I can make any model pass a goodness of fit test by broadening the uncertainty
- That doesn't make it a good model



Vitamin study

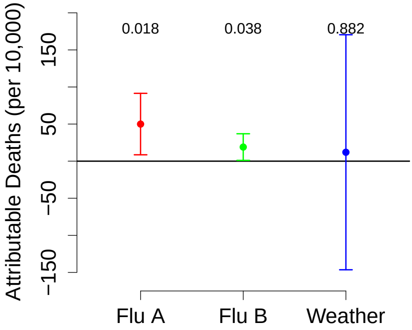


... with confidence intervals



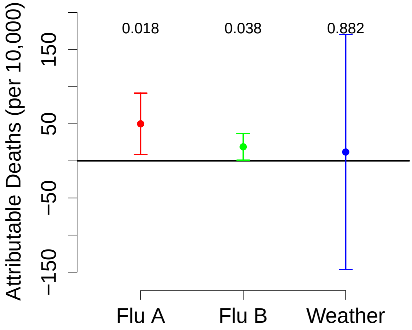
- **Never** say: A is significant and B isn't, so $A > B$

... with confidence intervals



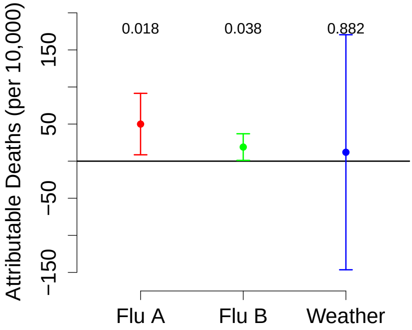
- **Never** say: A is significant and B isn't, so $A > B$
- **Instead:** Construct a statistic for the hypothesis $A > B$

... with confidence intervals



- **Never** say: A is significant and B isn't, so $A > B$
- **Instead:** Construct a statistic for the hypothesis $A > B$
  - May be difficult

... with confidence intervals



- **Never** say: A is significant and B isn't, so $A > B$
- **Instead:** Construct a statistic for the hypothesis $A > B$
  - May be difficult

Low P values

High P values

What does the P value mean?

- Low: you are seeing something clearly

What does the P value mean?

- Low: you are seeing something clearly
- High: you are seeing something unclearly

What does the P value mean?

- Low: you are seeing something clearly
- High: you are seeing something unclearly

Goodness of fit test

- Your model is *not* reality (null hypothesis is false)

Goodness of fit test

- Your model is *not* reality (null hypothesis is false)
- Can we see the difference clearly?

Goodness of fit test

- Your model is *not* reality (null hypothesis is false)
- Can we see the difference clearly?
  - If *no*, model may be *good* or *bad*.

Goodness of fit test

- Your model is *not* reality (null hypothesis is false)

- Can we see the difference clearly?
  - If *no*, model may be *good* or *bad*.
    - We probably can't add any more complexity based on current data

Goodness of fit test

- Your model is *not* reality (null hypothesis is false)
- Can we see the difference clearly?
  - If *no*, model may be *good* or *bad*.
    - We probably can't add any more complexity based on current data
  - If *yes*, model may be *good* or *bad*.

Goodness of fit test

- Your model is *not* reality (null hypothesis is false)
- Can we see the difference clearly?
  - If *no*, model may be *good* or *bad*.
    - We probably can't add any more complexity based on current data
  - If *yes*, model may be *good* or *bad*.
    - We *may* be able to add more complexity based on current data

Goodness of fit test

- Your model is *not* reality (null hypothesis is false)
- Can we see the difference clearly?
  - If *no*, model may be *good* or *bad*.
    - We probably can't add any more complexity based on current data
  - If *yes*, model may be *good* or *bad*.
    - We *may* be able to add more complexity based on current data
    - But we may not need to

Goodness of fit test

- Your model is *not* reality (null hypothesis is false)
- Can we see the difference clearly?
  - If *no*, model may be *good* or *bad*.
    - We probably can't add any more complexity based on current data
  - If *yes*, model may be *good* or *bad*.
    - We *may* be able to add more complexity based on current data
    - But we may not need to
- Reward and punishment

Goodness of fit test

- Your model is *not* reality (null hypothesis is false)
- Can we see the difference clearly?
  - If *no*, model may be *good* or *bad*.
    - We probably can't add any more complexity based on current data
  - If *yes*, model may be *good* or *bad*.
    - We *may* be able to add more complexity based on current data
    - But we may not need to
- Reward and punishment

Outline

Capturing patterns

- You can ask:

Capturing patterns

- You can ask:
  - Does your model do a reasonable job of capturing the data?

Capturing patterns

- You can ask:
  - Does your model do a reasonable job of capturing the data?
    - You can use a goodness of fit *statistic* for this, and not worry about the P value

Capturing patterns

- You can ask:
  - Does your model do a reasonable job of capturing the data?
    - You can use a goodness of fit *statistic* for this, and not worry about the P value
  - Does your model capture patterns and relationships that you (or other experts) think are important?

Capturing patterns

- You can ask:
  - Does your model do a reasonable job of capturing the data?
    - You can use a goodness of fit *statistic* for this, and not worry about the P value
  - Does your model capture patterns and relationships that you (or other experts) think are important?

Outline

- Conceptual models
- Prediction
- Model Validation
- Model Evaluation
  - Goodness of fit
  - Capturing patterns
  - Going beyond
- Conclusion

Out-of-sample validation

- Does your model make predictions *outside* the range on which you calibrated it?

Out-of-sample validation

- Does your model make predictions *outside* the range on which you calibrated it?
  - Predicting gravitational shifts in star positions from measurements in Earth laboratories

Out-of-sample validation

- Does your model make predictions *outside* the range on which you calibrated it?
  - Predicting gravitational shifts in star positions from measurements in Earth laboratories
  - Predicting cholera outbreaks in Bangladesh from a model calibrated to Haiti

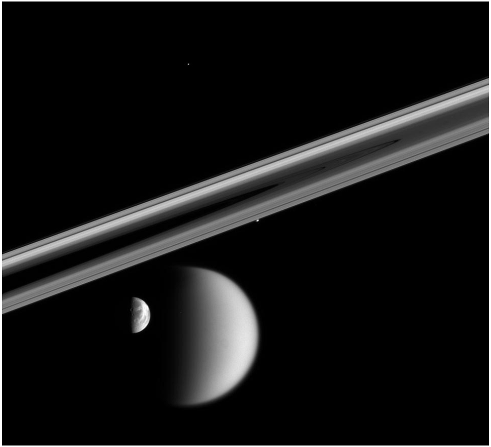
Out-of-sample validation

- Does your model make predictions *outside* the range on which you calibrated it?
  - Predicting gravitational shifts in star positions from measurements in Earth laboratories
  - Predicting cholera outbreaks in Bangladesh from a model calibrated to Haiti
  - Predicting influenza patterns in 2010 from a model calibrated from 2000–2009

Out-of-sample validation

- Does your model make predictions *outside* the range on which you calibrated it?
  - Predicting gravitational shifts in star positions from measurements in Earth laboratories
  - Predicting cholera outbreaks in Bangladesh from a model calibrated to Haiti
  - Predicting influenza patterns in 2010 from a model calibrated from 2000–2009

Predicting way out of sample



Essentially, all models are wrong, but some are useful.
– Box and Draper (1987), *Empirical Model Building . . .*

Test sets

- What is **test set** spelled backwards?

Test sets

- What is **test set** spelled backwards?
- Hold some data out while fitting your model

Test sets

- What is **test set** spelled backwards?
- Hold some data out while fitting your model
- Or just *pretend* to do this as an evaluation method

Test sets

- What is **test set** spelled backwards?
- Hold some data out while fitting your model
- Or just *pretend* to do this as an evaluation method
  - In other words, test what would happen under various withholding scenarios

Test sets

- What is **test set** spelled backwards?
- Hold some data out while fitting your model
- Or just *pretend* to do this as an evaluation method
  - In other words, test what would happen under various withholding scenarios
  - This can get very elaborate, and we should probably do it more

Test sets

- What is **test set** spelled backwards?
- Hold some data out while fitting your model
- Or just *pretend* to do this as an evaluation method
  - In other words, test what would happen under various withholding scenarios
  - This can get very elaborate, and we should probably do it more

Other model worlds

- The model you're *fitting* is probably pretty simple

Other model worlds

- The model you're *fitting* is probably pretty simple
- But you can *simulate* very complicated models, indeed

Other model worlds

- The model you're *fitting* is probably pretty simple
- But you can *simulate* very complicated models, indeed

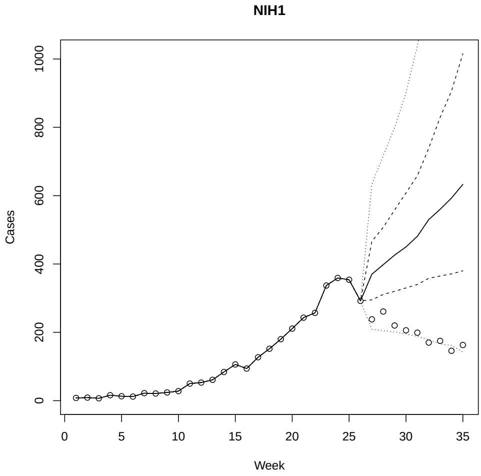
- How well can you do? Which details are important?

Other model worlds

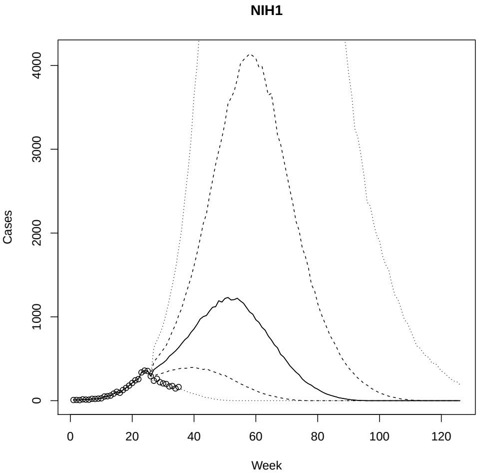
- The model you're *fitting* is probably pretty simple
- But you can *simulate* very complicated models, indeed

- How well can you do? Which details are important?

Other model worlds

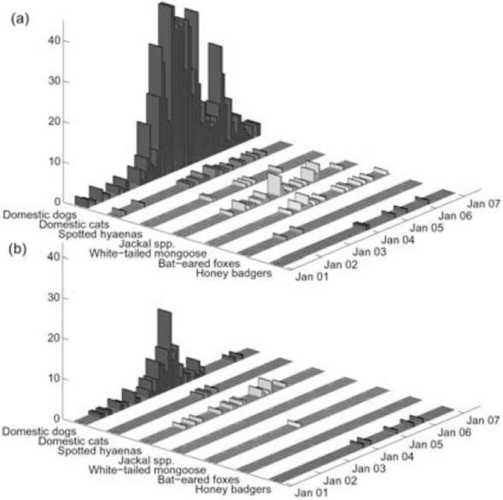


Other model worlds

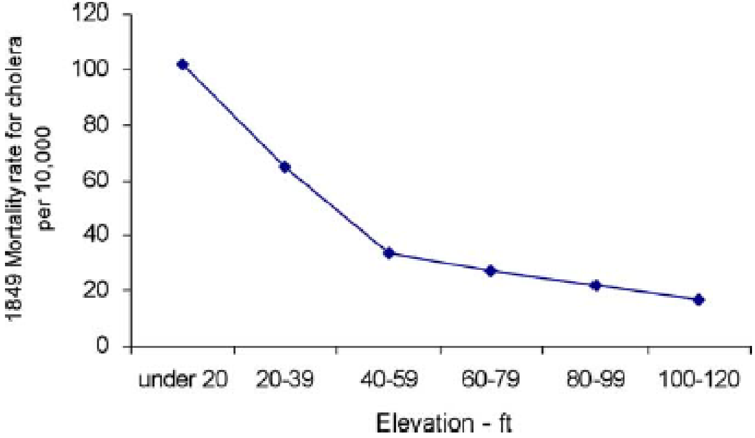


Generating hypotheses

Generating hypotheses



Testing hypotheses



Testing hypotheses

Testing hypotheses

Hard questions

Answers are not always easy

Outline

Summary

Dynamic models

- Clarify thinking

Summary

Dynamic models

- Clarify thinking
  - What are our assumptions, what else do we need to know?

Summary

Dynamic models

- Clarify thinking
  - What are our assumptions, what else do we need to know?
- Understand outcomes

Summary

Dynamic models

- Clarify thinking
  - What are our assumptions, what else do we need to know?
- Understand outcomes
  - Can heterogeneity explain the time course of HIV epidemics?

Summary

Dynamic models

- Clarify thinking
  - What are our assumptions, what else do we need to know?
- Understand outcomes
  - Can heterogeneity explain the time course of HIV epidemics?
  - Is it possible that MDA could break the cycle of malaria transmission in some areas?

Summary

Dynamic models

- Clarify thinking
  - What are our assumptions, what else do we need to know?
- Understand outcomes
  - Can heterogeneity explain the time course of HIV epidemics?
  - Is it possible that MDA could break the cycle of malaria transmission in some areas?
- Predict outcomes

Summary

Dynamic models

- Clarify thinking
  - What are our assumptions, what else do we need to know?
- Understand outcomes
  - Can heterogeneity explain the time course of HIV epidemics?
  - Is it possible that MDA could break the cycle of malaria transmission in some areas?
- Predict outcomes
  - What is the potential for a hepatitis A outbreak in Cape Town?

Summary

Dynamic models

- Clarify thinking
  - What are our assumptions, what else do we need to know?
- Understand outcomes
  - Can heterogeneity explain the time course of HIV epidemics?
  - Is it possible that MDA could break the cycle of malaria transmission in some areas?
- Predict outcomes
  - What is the potential for a hepatitis A outbreak in Cape Town?
  - What might happen if I improve testing-and-treatment outreach in Jamaica?

Summary

Dynamic models

- Clarify thinking
  - What are our assumptions, what else do we need to know?
- Understand outcomes
  - Can heterogeneity explain the time course of HIV epidemics?
  - Is it possible that MDA could break the cycle of malaria transmission in some areas?
- Predict outcomes
  - What is the potential for a hepatitis A outbreak in Cape Town?
  - What might happen if I improve testing-and-treatment outreach in Jamaica?
- Find new mechanisms

Summary

Dynamic models

- Clarify thinking
  - What are our assumptions, what else do we need to know?
- Understand outcomes
  - Can heterogeneity explain the time course of HIV epidemics?
  - Is it possible that MDA could break the cycle of malaria transmission in some areas?
- Predict outcomes
  - What is the potential for a hepatitis A outbreak in Cape Town?
  - What might happen if I improve testing-and-treatment outreach in Jamaica?
- Find new mechanisms
  - Why can't I explain my data? What haven't I thought of?

Summary

Dynamic models

- Clarify thinking
  - What are our assumptions, what else do we need to know?
- Understand outcomes
  - Can heterogeneity explain the time course of HIV epidemics?
  - Is it possible that MDA could break the cycle of malaria transmission in some areas?
- Predict outcomes
  - What is the potential for a hepatitis A outbreak in Cape Town?
  - What might happen if I improve testing-and-treatment outreach in Jamaica?
- Find new mechanisms
  - Why can't I explain my data? What haven't I thought of?

Summary

Evaluation

- Validation (inside your model world)

Summary

Evaluation

- Validation (inside your model world)
  - Does my fitting method work (assuming my model is right)?

Summary

Evaluation

- Validation (inside your model world)
  - Does my fitting method work (assuming my model is right)?
- Inspection (compare patterns)

Summary

Evaluation

- Validation (inside your model world)
  - Does my fitting method work (assuming my model is right)?
- Inspection (compare patterns)
- Prediction (and other out-of-sample comparison)

Summary

Evaluation

- Validation (inside your model world)
  - Does my fitting method work (assuming my model is right)?
- Inspection (compare patterns)
- Prediction (and other out-of-sample comparison)
  - Can my model predict things I haven't told it yet?

Summary

Evaluation

- Validation (inside your model world)
  - Does my fitting method work (assuming my model is right)?
- Inspection (compare patterns)
- Prediction (and other out-of-sample comparison)
  - Can my model predict things I haven't told it yet?
- Generate and test mechanistic hypotheses

Summary

Evaluation

- Validation (inside your model world)
  - Does my fitting method work (assuming my model is right)?
- Inspection (compare patterns)
- Prediction (and other out-of-sample comparison)
  - Can my model predict things I haven't told it yet?
- Generate and test mechanistic hypotheses

Title: Model evaluation and comparison
Attribution: Jonathan Dushoff, McMaster University, MMED 2019

Source URL: https://figshare.com/collections/International_Clinics_on_Infectious_Disease_Dynamics_and_Data/3788224
For further information please contact admin@ici3d.org.