DAIDD
Gainesville, Florida
December 2014

# Model Evaluation and Comparison

Jim Scott, Ph.D, M.A., M.P.H.

# Goals

- By the end of this talk, I hope you'll:
  - Have a good sense of what *model evaluation* is, why it's important, and how it's tied to your research question
  - Know some of the characteristics that are desirable in models

# Steps of Mathematical Modeling

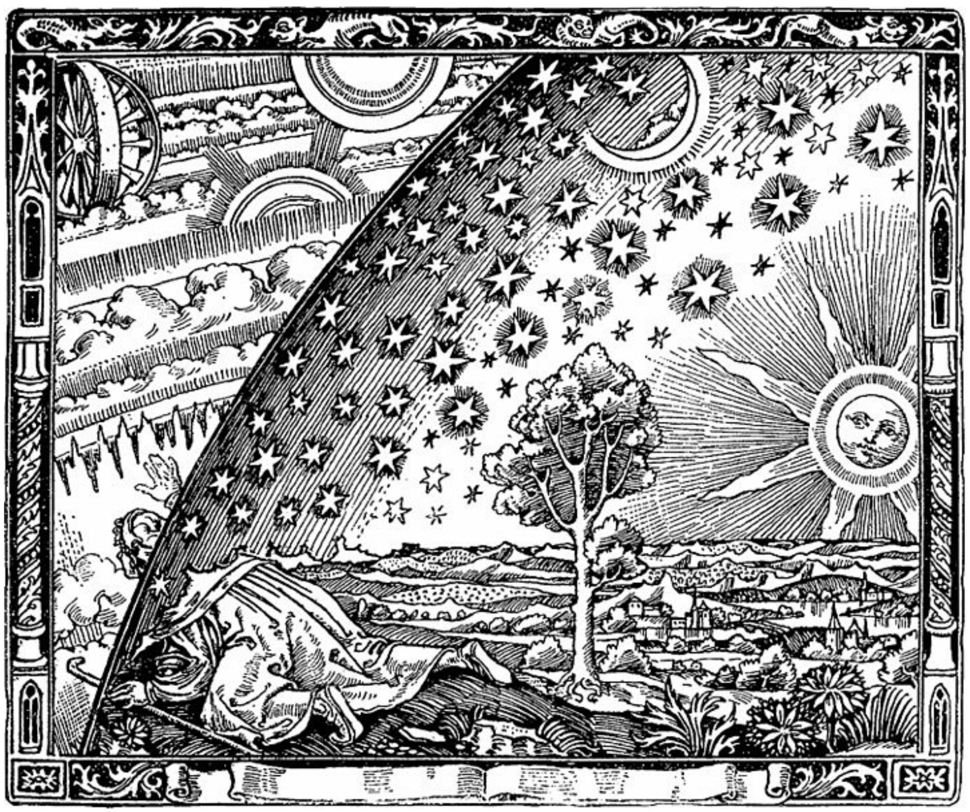
- Specific question
- Identify relevant factors and information
- Model formulation
- Mathematics
- Evaluation

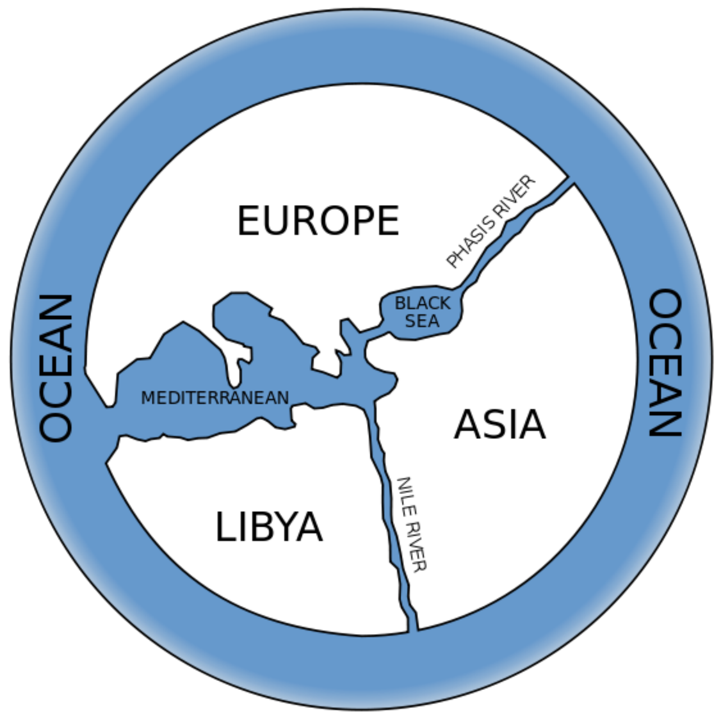
# Model evaluation

- "E" topics:
  - Edison, Thomas
  - Epicycle
  - Epidemiology
  - Elasticity
  - Eratosthenes
  - Euler
  - Existence
  - Evolution
  - Extrapolation
  - Eradication
- "M" topics:
  - Malthus, Thomas
  - Mars
  - Maternity
  - Maximum likelihood
  - Maxwell, James C.
  - Misery
  - Monte Carlo
  - Moon
  - Mortality
  - Mumps

# What makes a model "good"?

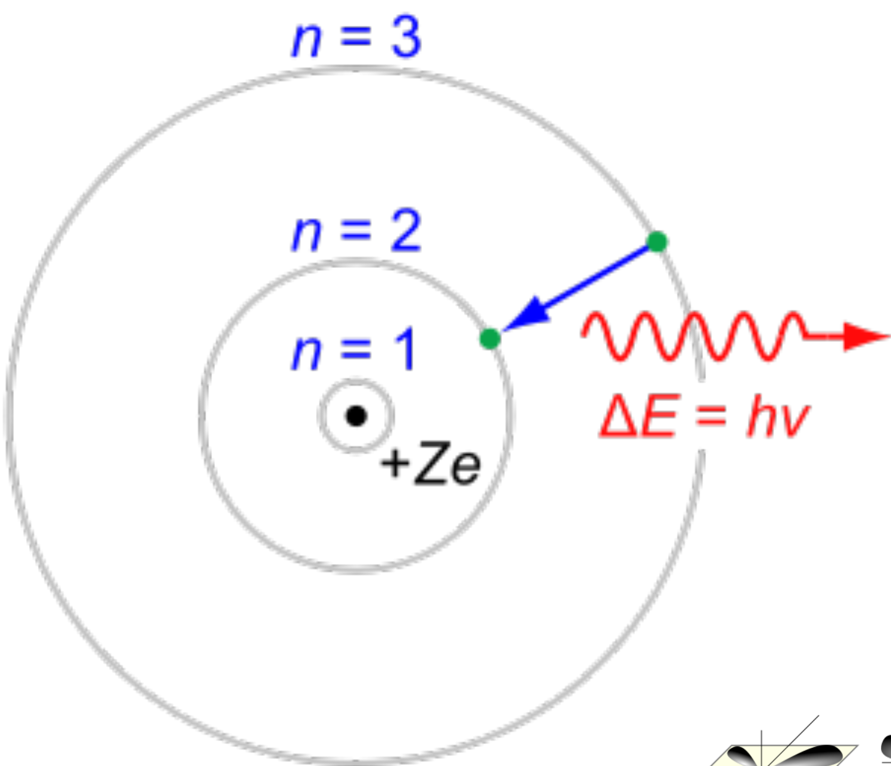
- Let's look at a few different types of models

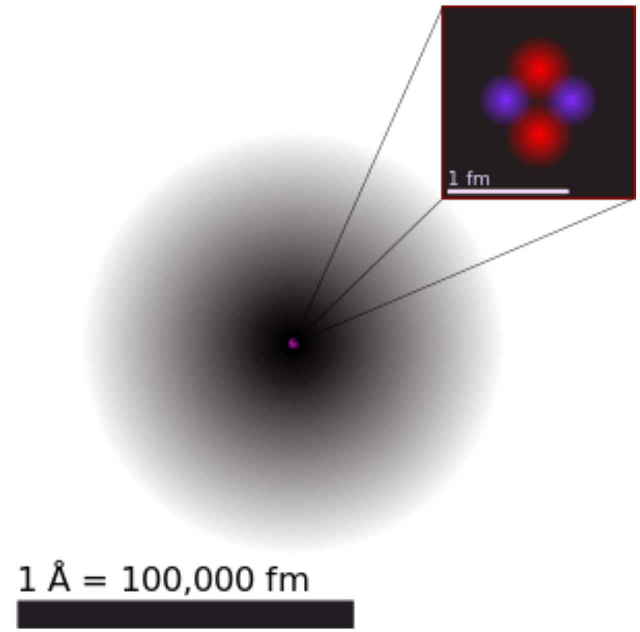


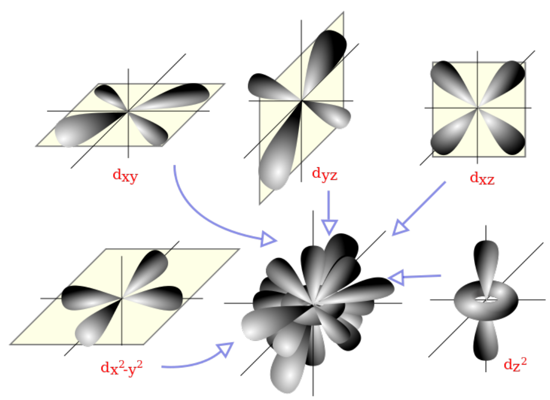



Source: Wikipedia

# What makes a model "good"?

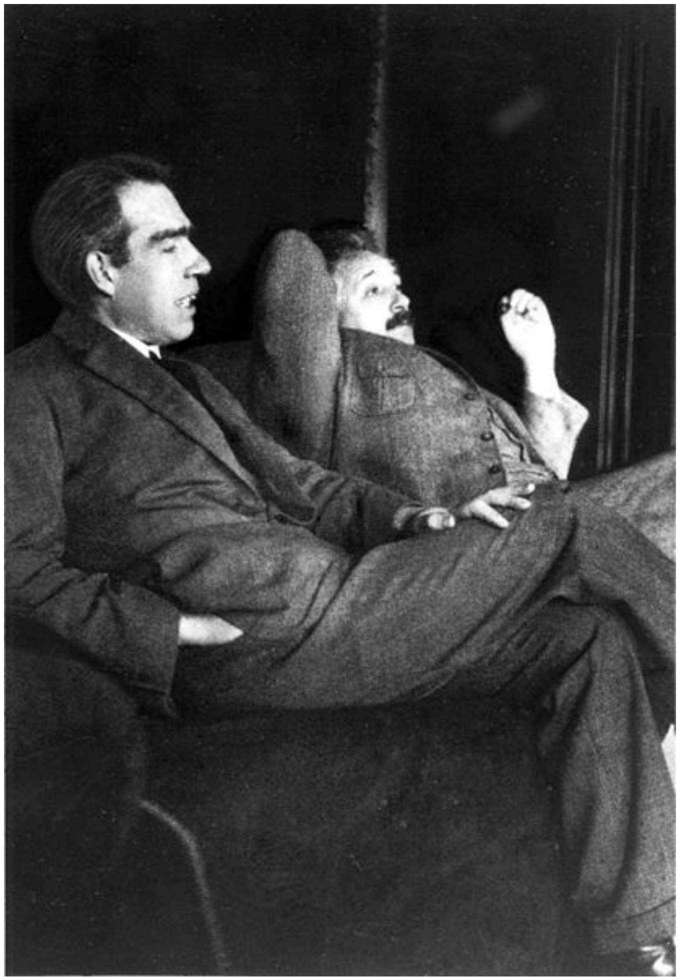


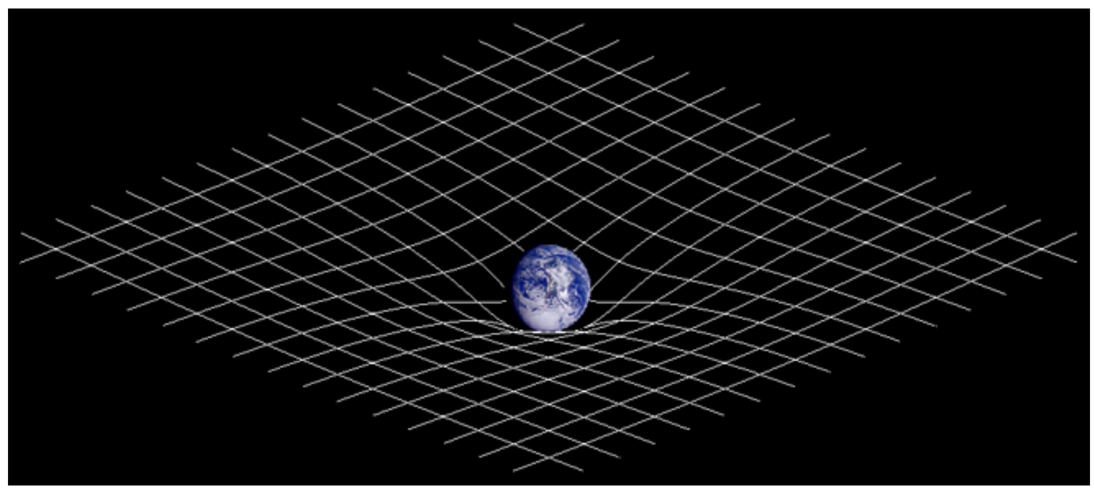




Source: Wikipedia

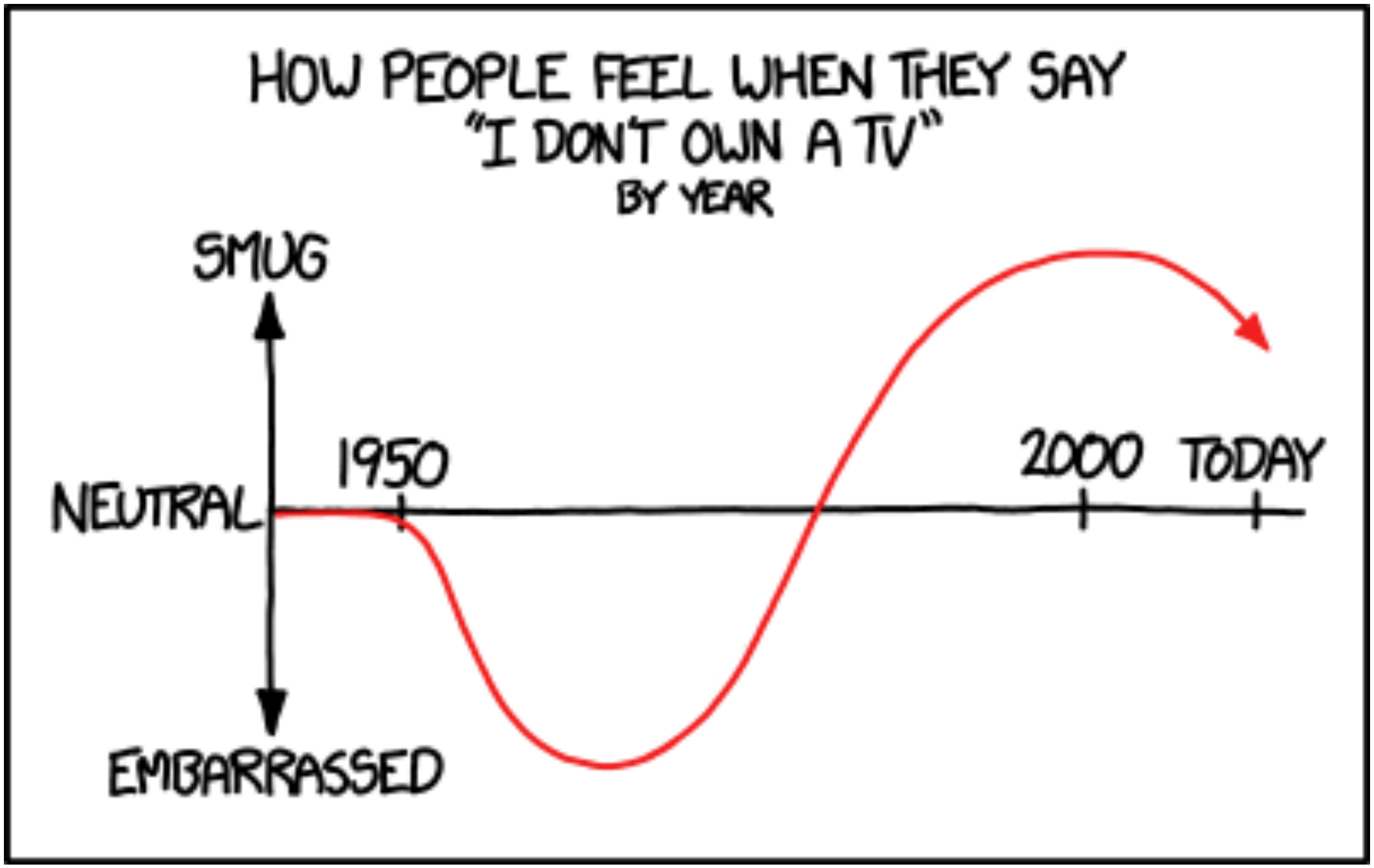
# What makes a model “good”?



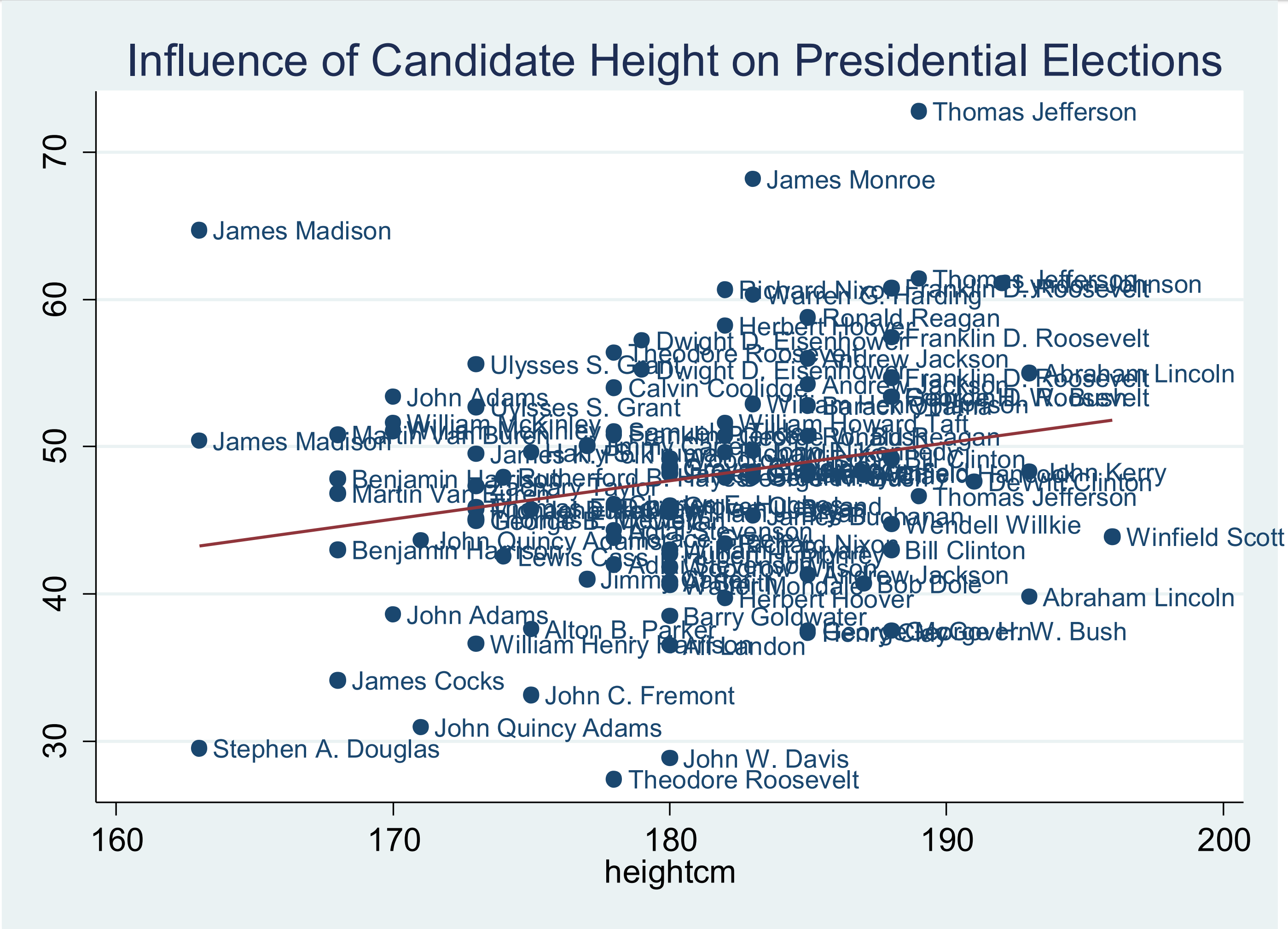
Source: Wikipedia

# What makes a model "good"?



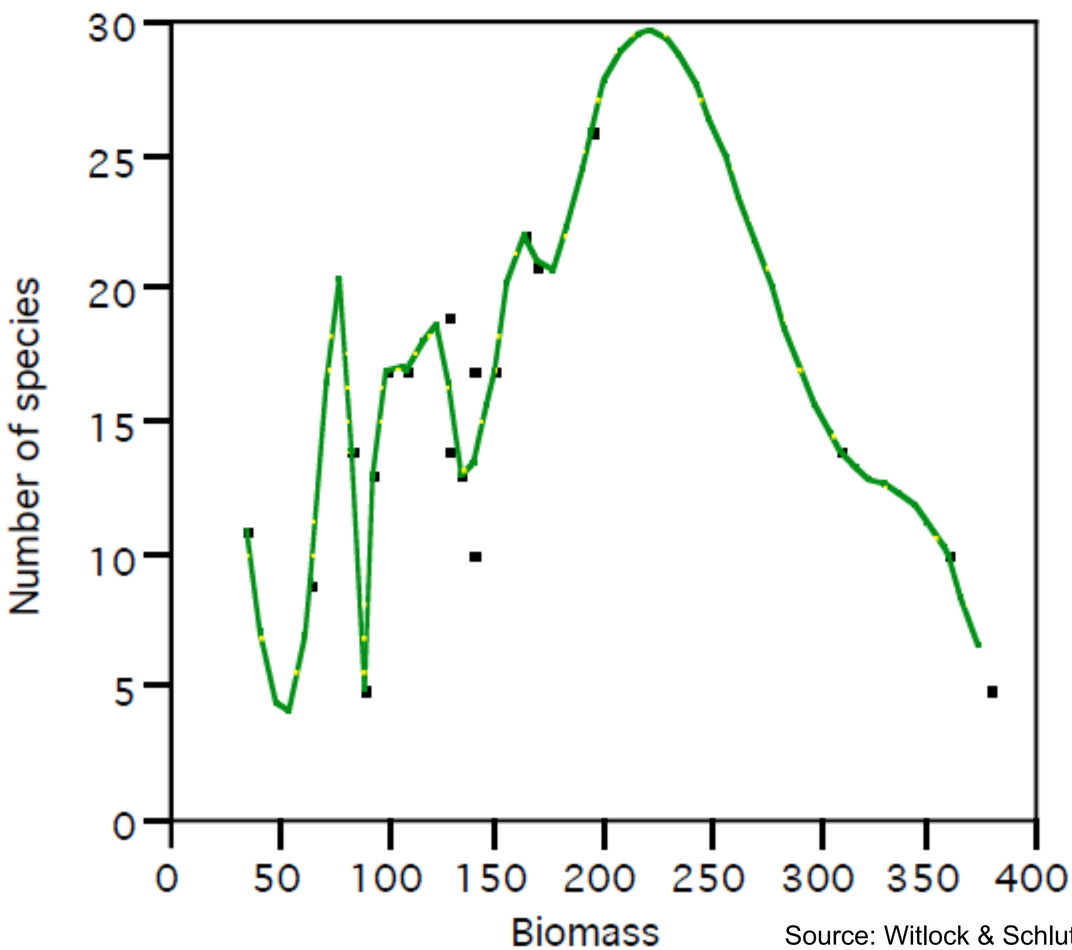

Source: XKCD

# What makes a model "good"?




**Estimated Percent Popular Vote = 1.07 + 0.26 * Height**

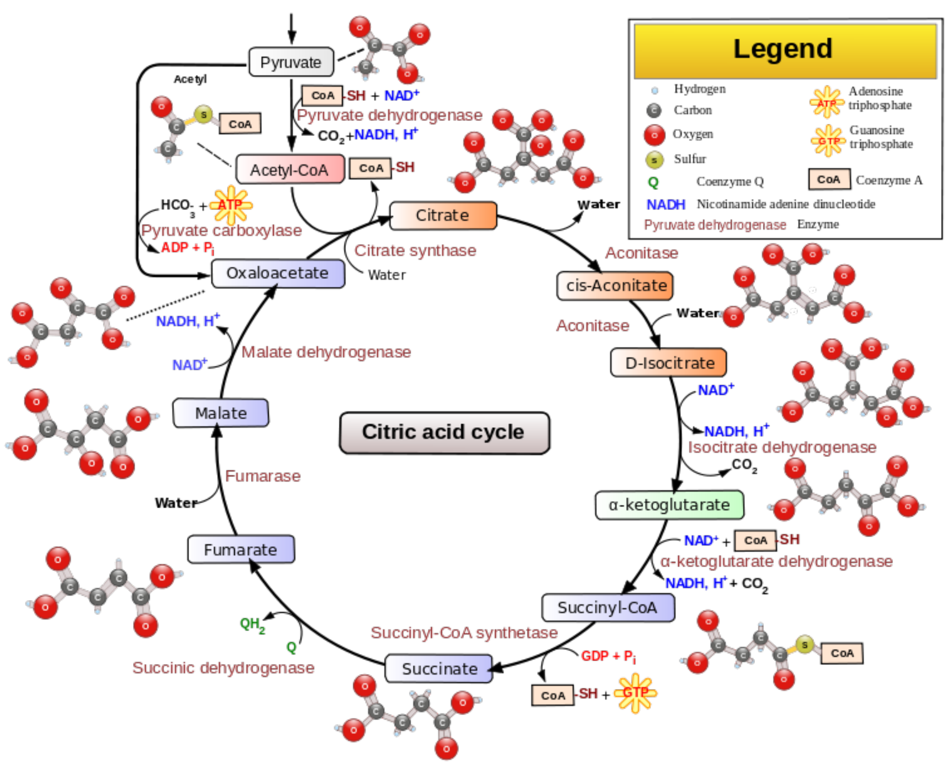
# What makes a model “good”?

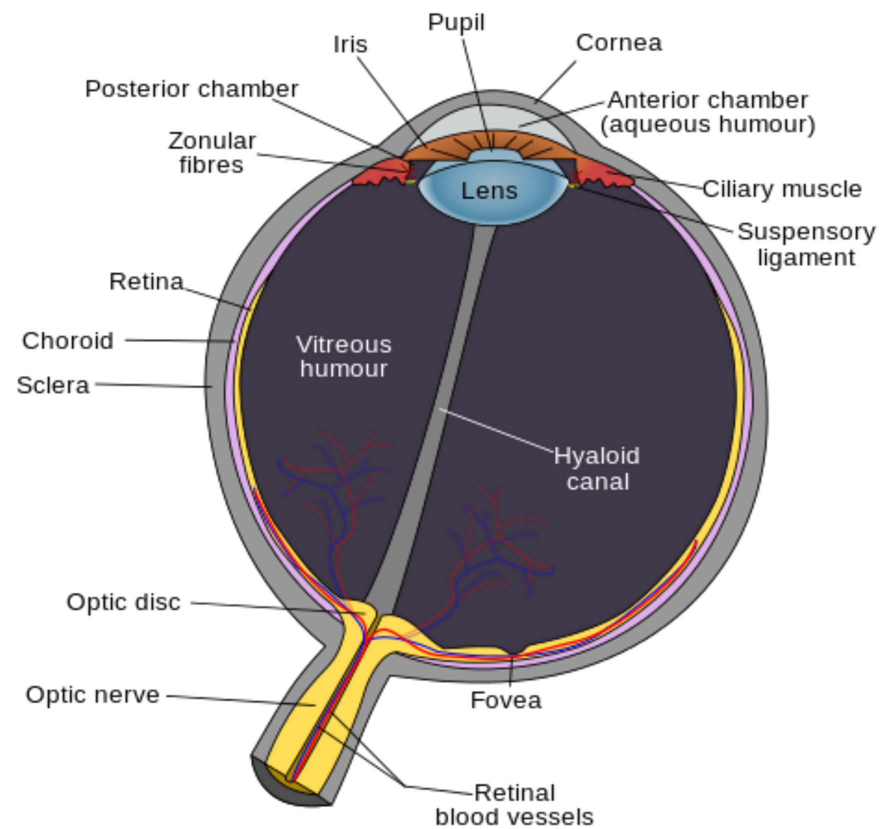



Source: Witlock & Schluter, Analysis of Biological Data

# What makes a model "good"?

## The Krebs Cycle




## The Eye

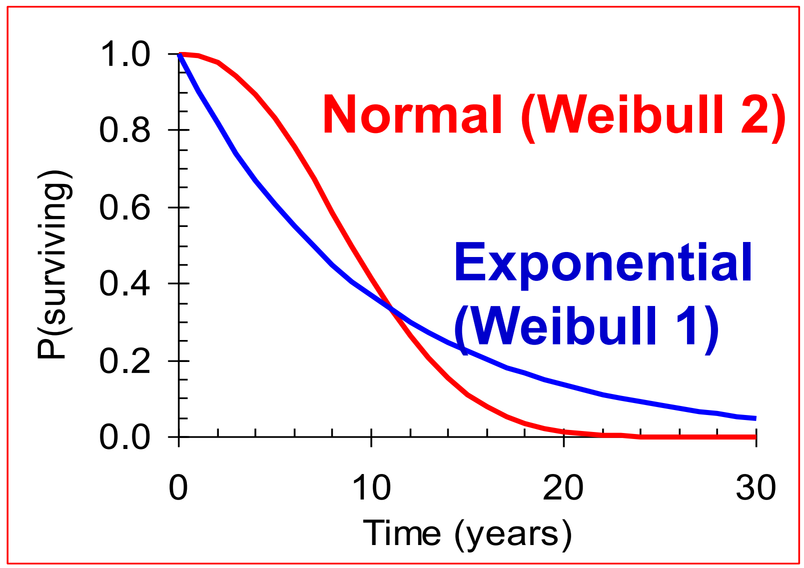
# What makes a model "good"?


$\beta$ = birth rate

$N$ = $S + I$

$\lambda$ = infection rate

$\overline{\delta I}$ = Weibull mortality




Slide credit: J. Hargrove/B. Williams

# That was the last one…remind them about the exercise….

- Consider the previous examples
- Talk with someone next to you
- Come up with a list of characteristics that a "good" model should have
- For example, you might say "simplicity"

# Examples of Models

- Flat world
- Atom
- Feelings about TV ownership
- Height and popular vote
- Species and biomass
- Krebs cycle
- The eye
- SIR
- $E=mc^2$

# Desirable Characteristics

- Accurate (i.e. low bias)
-
-
-
-
-
-
-
-

# Accuracy

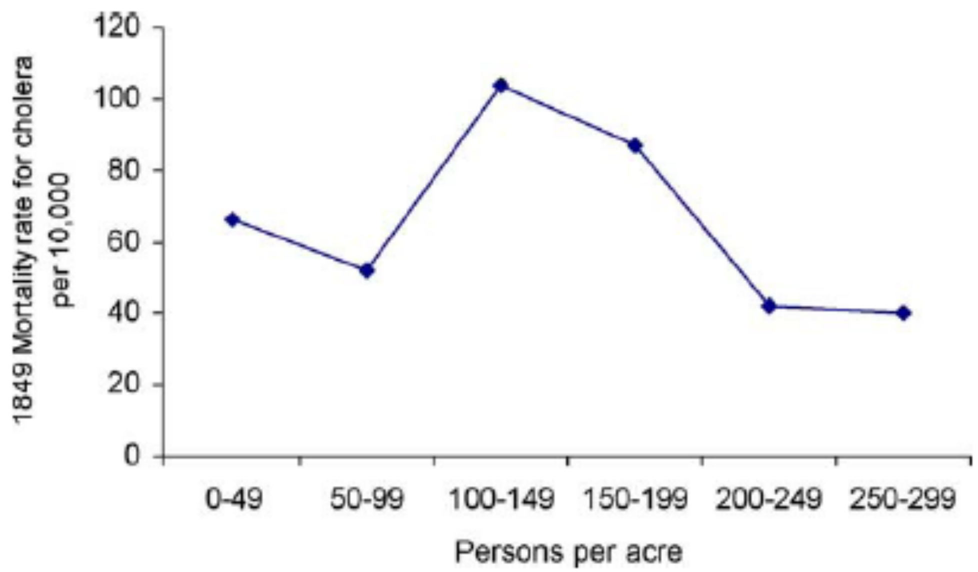
- A model is *accurate* if estimates based on the model match the truth
  - E.g. models that are used to predict the weather are reasonably accurate when predicting tomorrow's weather.  They are much less accurate at predicting the weather at times further into the future.
  - Does the model *fit* the data

# Farr and Cholera

John Snow, William Farr and the 1849 outbreak of cholera that affected London: a reworking of the data highlights the importance of the water supply

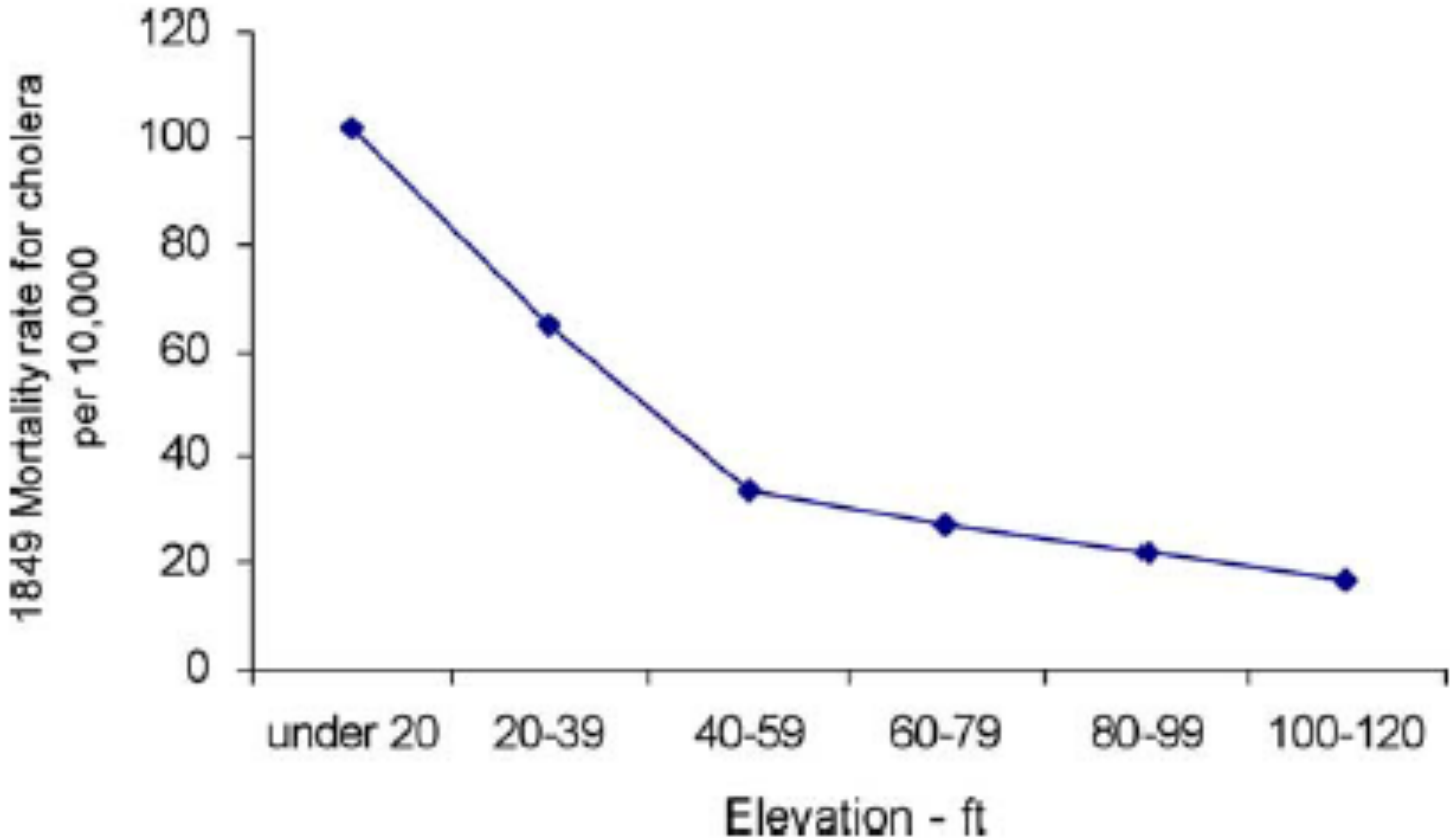
P. Bingham[a,*], N.Q. Verlander[b], M.J. Cheal[a]

Public Health (2004) **118**, 387-394




Source: Wikipedia

# Farr and Cholera



# Farr and Cholera

- Farr initially believed in the miasma theory of disease transmission – disease was propagated by "bad air"
- The higher the elevation, the better the air

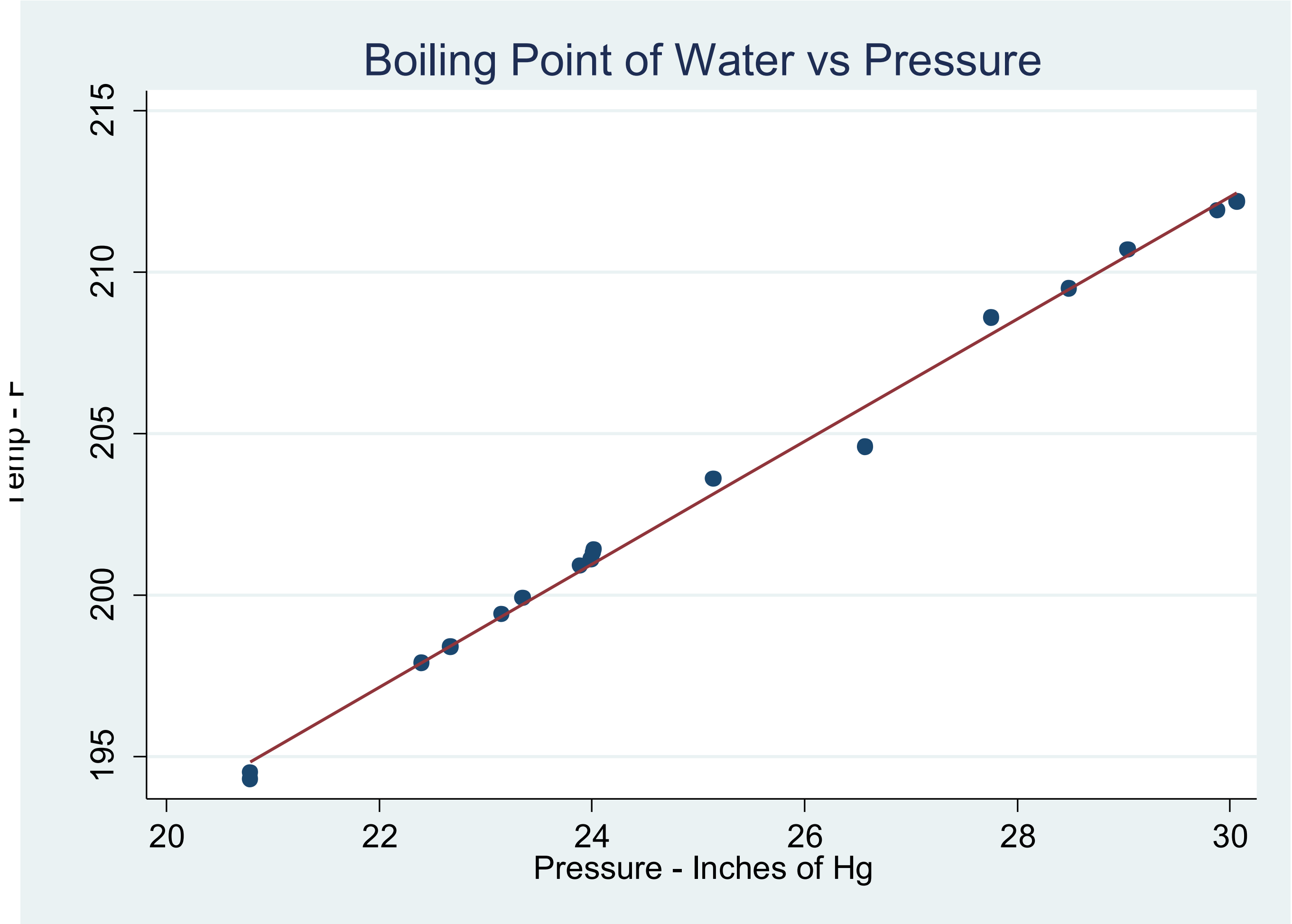
- Mortality at terrace level X = Mortality at terrace 1 / terrace level X

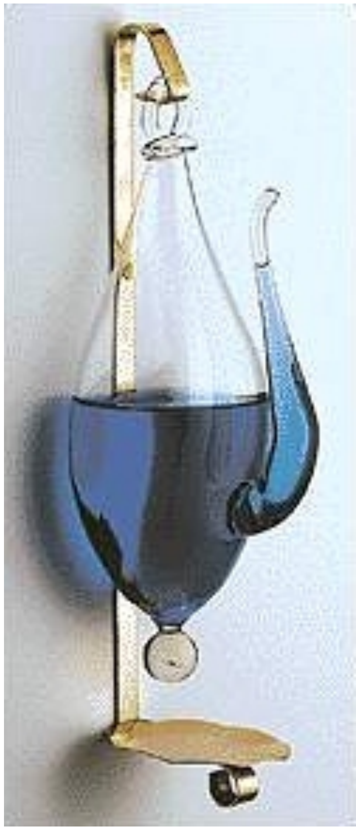
# Farr and Cholera

**Table 5** Observed mortality compared with expected for 'terraces of London'.

| Elevation of the 'terrace' above the Trinity high-water mark (feet) | Number of registration districts | Observed mortality from cholera per 10,000 inhabitants | 102 divided by 1, 2, 3, 4, 5 and 6 ('expected' mortality) |
|---|---|---|---|
| 0-19 | 16 | 102 | 102 |
| 20-39 | 7 | 65 | 51 |
| 40-59 | 8 | 34 | 34 |
| 60-79 | 3 | 27 | 26 |
| 80-99 | 2 | 22 | 20 |
| 100-120 | 1 | 17 | 17 |
| 340-359 | 1 | 8 | - |

# Boiling Point of $H_2O$ and Pressure



Source: Wikipedia

**r = 0.9972** **Est. Temp = 155.3 + 1.90 * Pressure**

# Desirable Characteristics

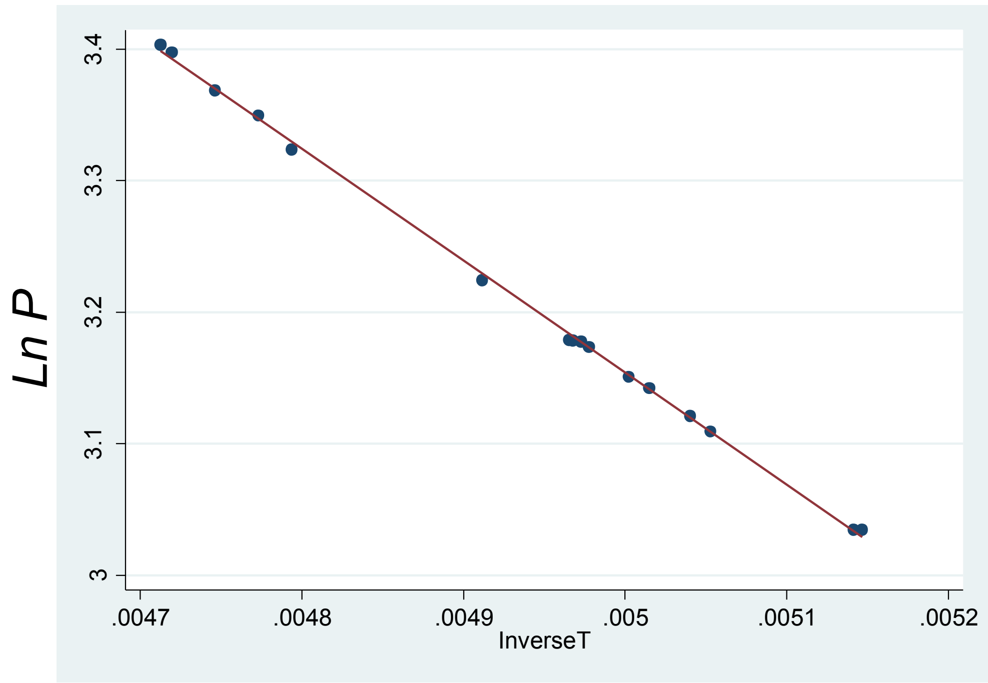
- Accurate (i.e. low bias)
- Descriptively realistic

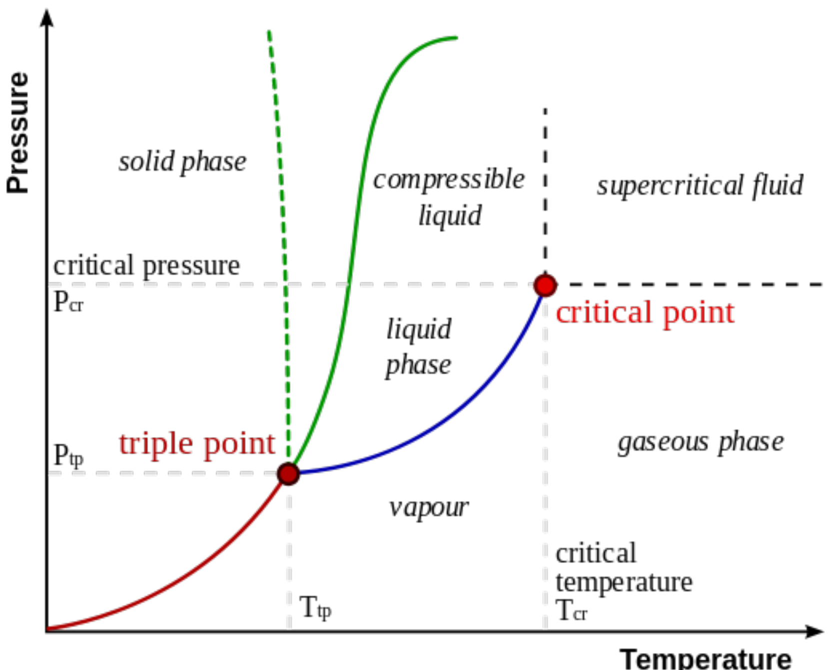
# Descriptive realism

- A model is *descriptively realistic* if it's derived from a correct description of the *mechanism* involved in whatever is being modeled
  - Corollary: underlying assumptions are correct
  - Statistical models are not descriptively realistic
    - For example, a linear equation only models a pattern in the data – there's nothing telling us what's going on behind the scene
  - An SIR model is more descriptively realistic
    - A mechanism for transmission is specified

# Clausius-Clapeyron Relation

$$Ln\ P = -\frac{L}{R}\left(\frac{1}{T}\right) + C$$




Source: Wikipedia

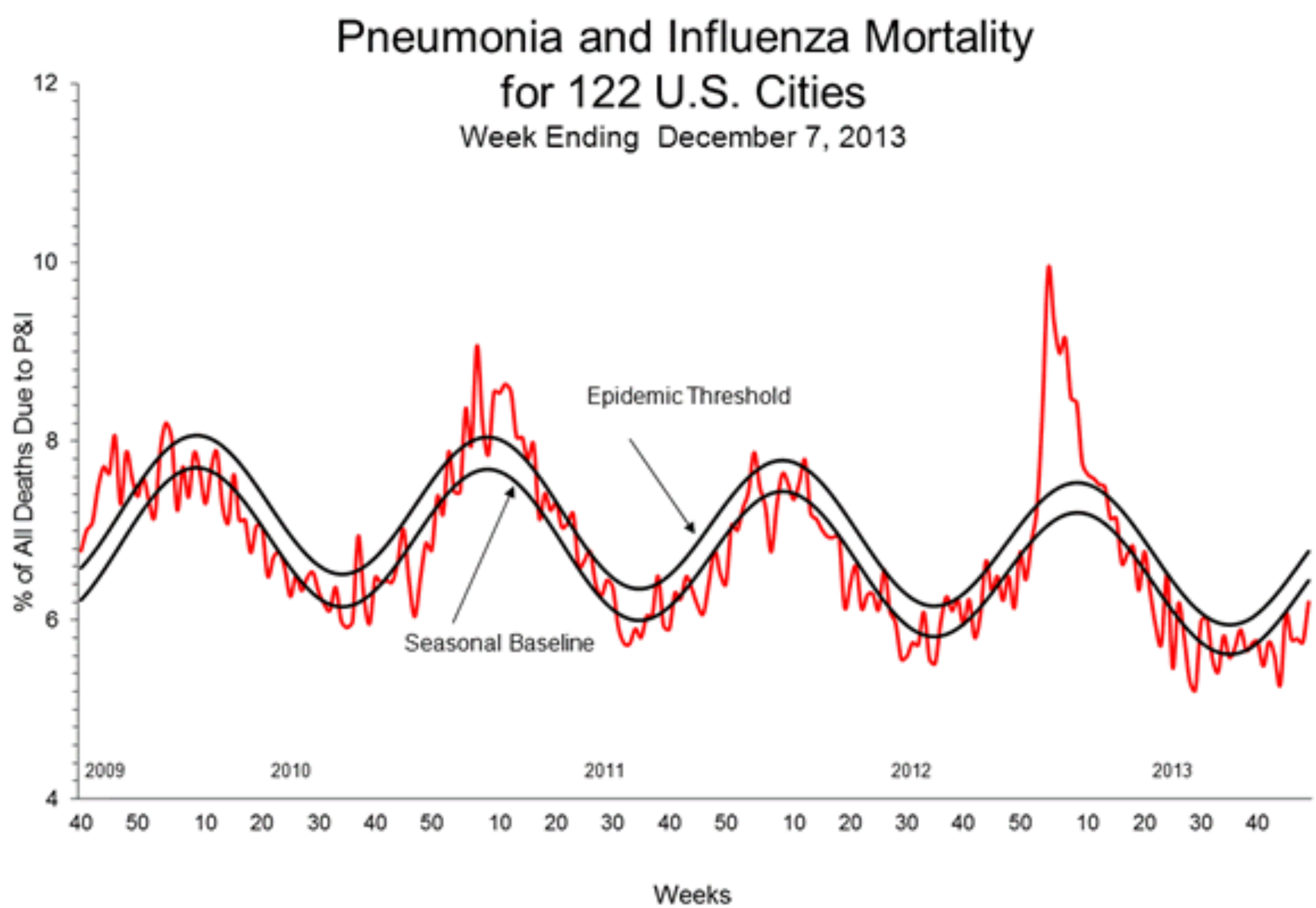
# Desirable Characteristics

- Accurate (i.e. low bias)
- Descriptively realistic
- Precise (i.e. low variability)

# Precision

- A model is *precise* if the estimates that the model produces have low variability
  - E.g. a model that estimates that it will start to rain in the next 3 – 6 *hours* is more precise than a model that estimates it will start to rain in the next 3 – 6 *days*

# Influenza




Weekly U.S. Influenza Surveillance Report, http://www.cdc.gov/flu/weekly/

# Estimating HIV Incidence

Estimating HIV incidence rates from age prevalence data in epidemic situations

Brian Williams[1,*,†], Eleanor Gouws[2], David Wilkinson[3] and Salim Abdool Karim[2]

STATISTICS IN MEDICINE
*Statist. Med.* 2001; **20**:2003–2016 (DOI: 10.1002/sim.840)

$$I(a,t) = P(a,t) - P(a-1,t)\frac{\bar{P}(t-1)}{\bar{P}(t)}e^{-\mu} \; *$$

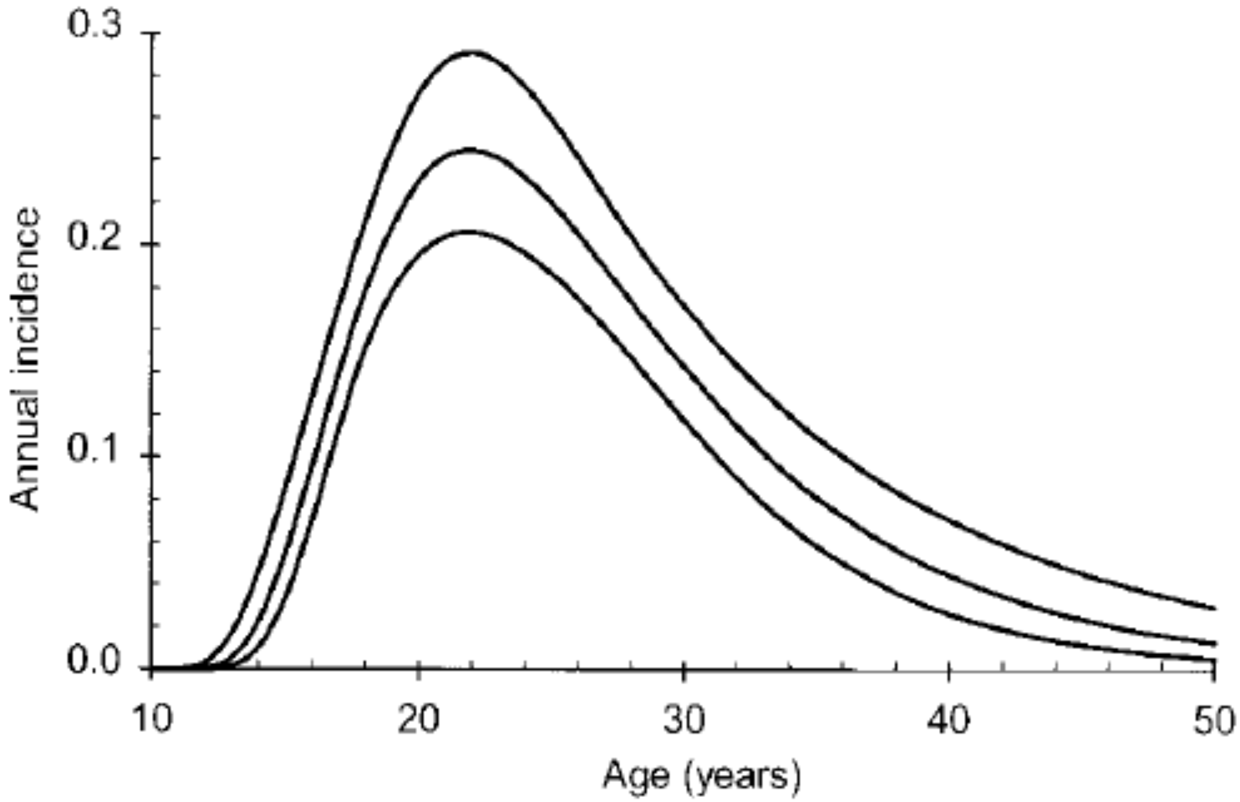
*I*(*a*,*t*) is the incidence at age *a* at time *t*

*P(a,t)* is the age-specific prevalence at age *a* at time *t*

* Simplified model

# Estimating HIV Incidence

The annual age-specific incidence per susceptible with 95 per cent confidence band

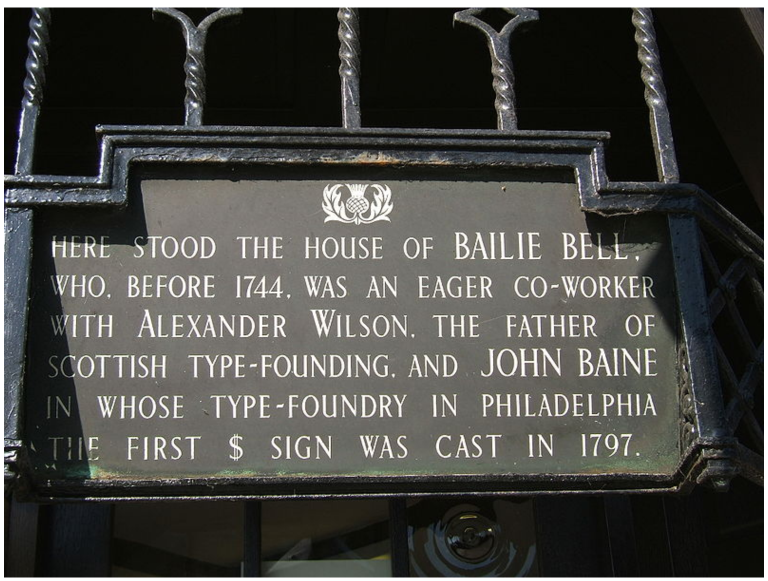


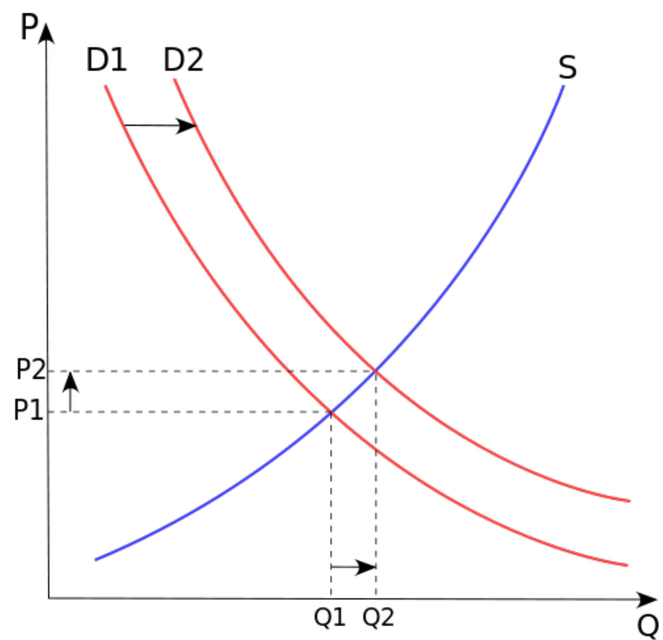
# Desirable Characteristics

- Accurate (i.e. low bias)
- Descriptively realistic
- Precise (i.e. low variability)
- General

# General

- A model is *general* if it applies to a wide variety of situations
  - e.g. the law of supply and demand




Source: Wikipedia

# Simple SIR Model

*A Contribution to the Mathematical Theory of Epidemics.*

By W. O. Kermack and A. G. McKendrick.

(Communicated by Sir Gilbert Walker, F.R.S.—Received May 13, 1927.)

(From the Laboratory of the Royal College of Physicians, Edinburgh.)

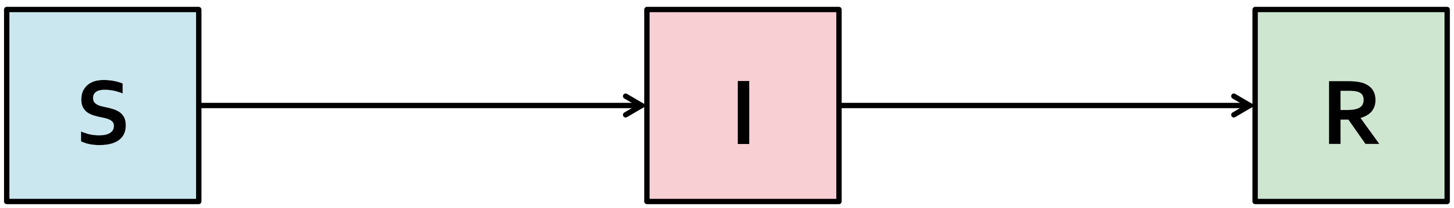
Source: *Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character*, Vol. 115, No. 772 (Aug. 1, 1927), pp. 700-721

In this case the equations are

$$\left.\begin{aligned} \frac{dx}{dt} &= -\kappa xy \\ \frac{dy}{dt} &= \kappa xy - ly \\ \frac{dz}{dt} &= ly \end{aligned}\right\}$$

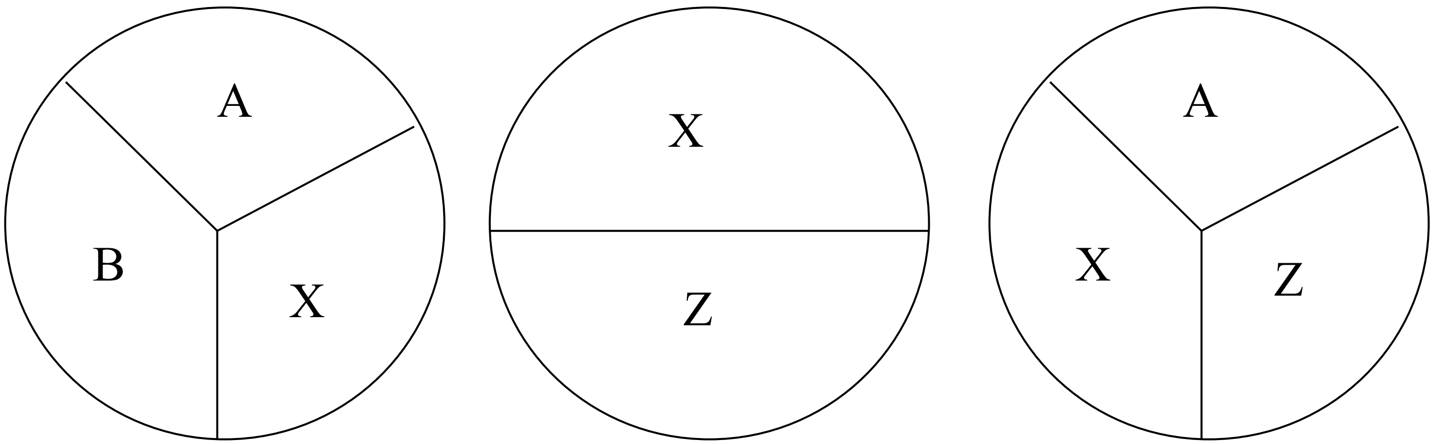
and as before $x + y + z = \mathrm{N}$.

# Simple SIR Model




$$\frac{dS}{dt} = \frac{-\beta SI}{N} \qquad \frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I \qquad \frac{dR}{dt} = \gamma I$$

# Sufficient – Component Cause Model (Rothman)




Each pie represents a *sufficient* cause for disease (i.e. disease is inevitable)

Each letter represents a *component* cause for a disease

The component cause X is a *necessary* cause (i.e. disease cannot occur without it)

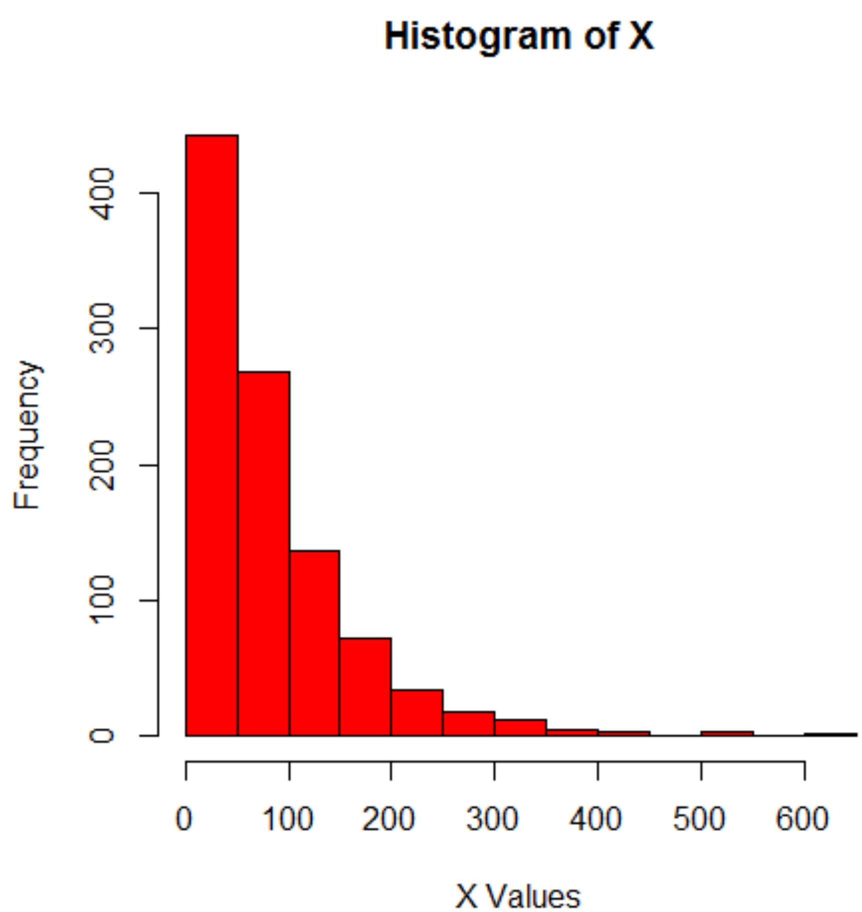
# Desirable Characteristics

- Accurate (i.e. low bias)
- Descriptively realistic
- Precise (i.e. low variability)
- General
- Robust
-
-
-

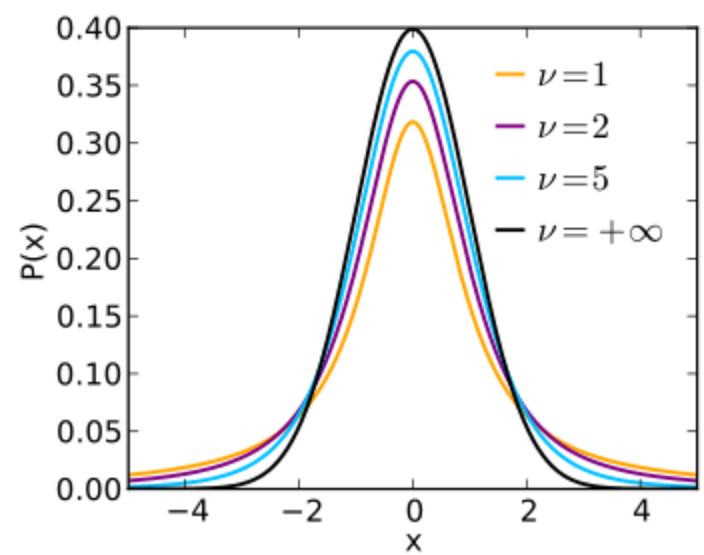
# Robust

- A model is *robust* if it is relatively immune to errors in the data and/or immune to small violations of model assumptions
  - Is the model very sensitive to relatively small changes in estimated input parameters?
    - Model is NOT robust
  - Do model predictions remain accurate even when some key assumptions do not strictly hold?
    - Model IS robust

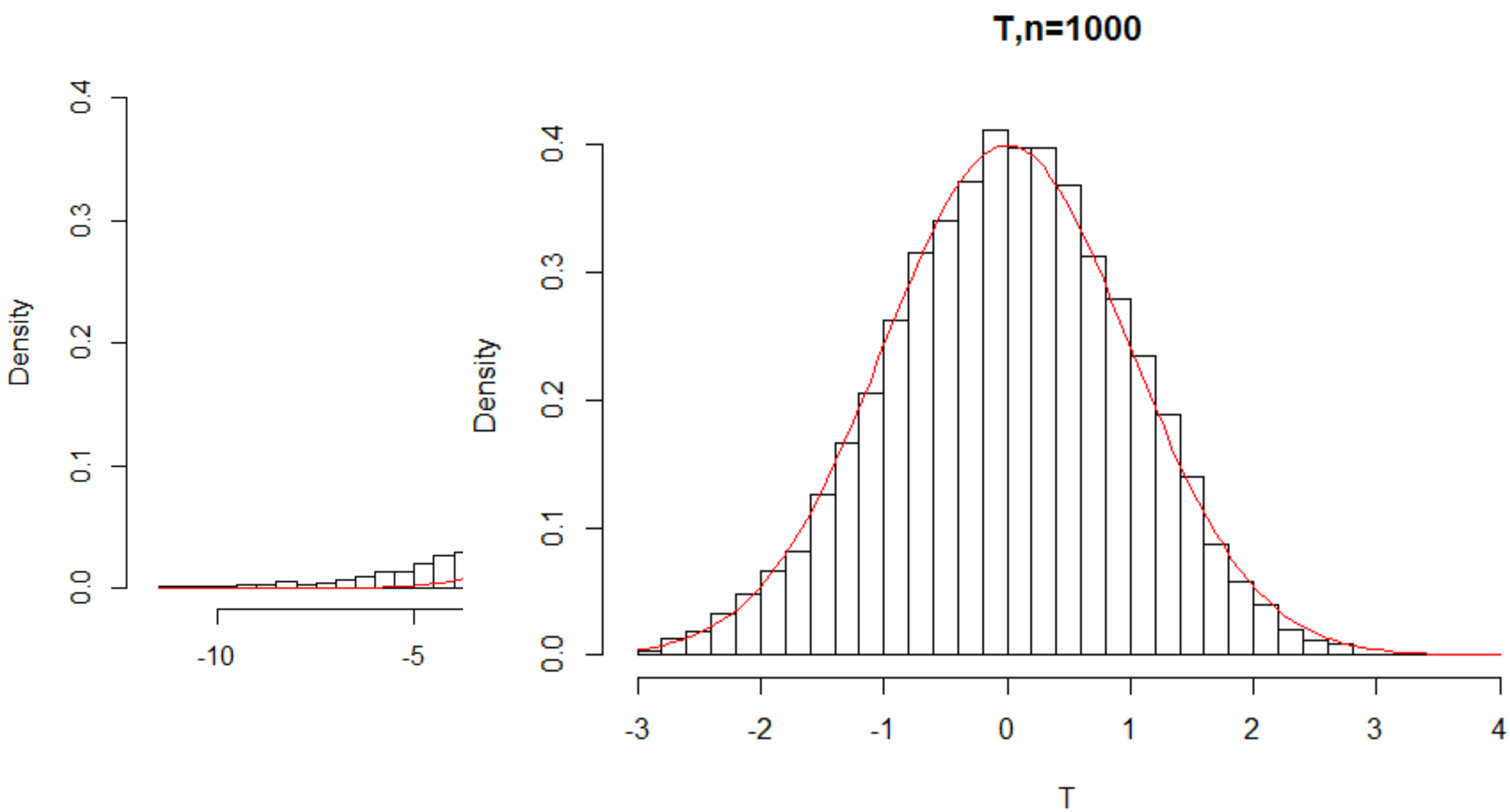
# T-tests




$$t_{n-1} = \frac{\bar{X} - \mu}{s/\sqrt{n}}$$



# T Simulation



# Water and Sanitation

## Integrating Disease Control Strategies: Balancing Water Sanitation and Hygiene Interventions to Reduce Diarrheal Disease Burden

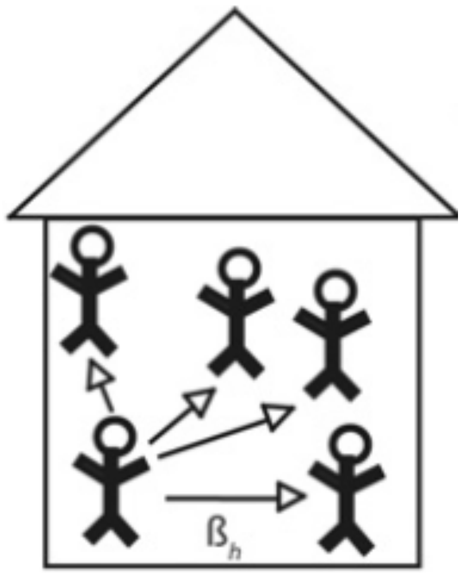
Joseph N. S. Eisenberg, PhD, MPH, James C. Scott, MPH, and Travis Porco, PhD, MPH

American Journal of Public Health | May 2007, Vol 97, No. 5

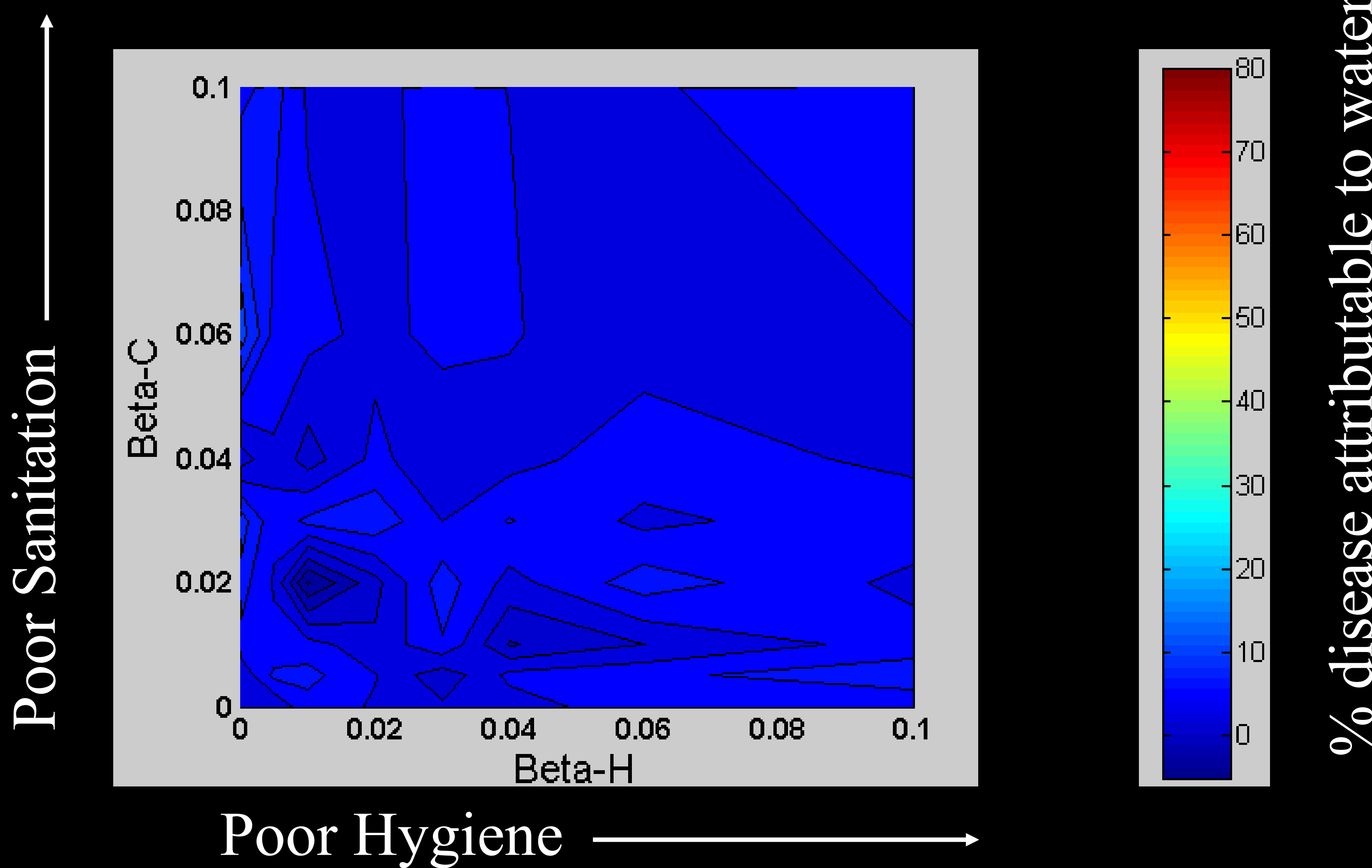
**TABLE 1—Parameter Values and Units Used in the Simulation Analysis**

| | Parameter Values | Units |
|---|---|---|
| $\rho$ | 10 | Days |
| $\phi$ | {0, 0.5, 1.0, 1.5, 2.0} | Pathogens/person/day |
| $\beta_c$ | {0, 0.005, 0.01, 0.02, 0.03, 0.04, 0.06, 0.1} | No. of transmission events/infected individual |
| $\beta_h$ | {0, 0.005, 0.01, 0.02, 0.03, 0.04, 0.06, 0.1} | No. of transmission events/infected individual |
| $\varepsilon$ | 30 | Pathogens |
| $\mu$ | 1 | Pathogens/day |
| $r$ | 0.000002 | Infections/pathogen |

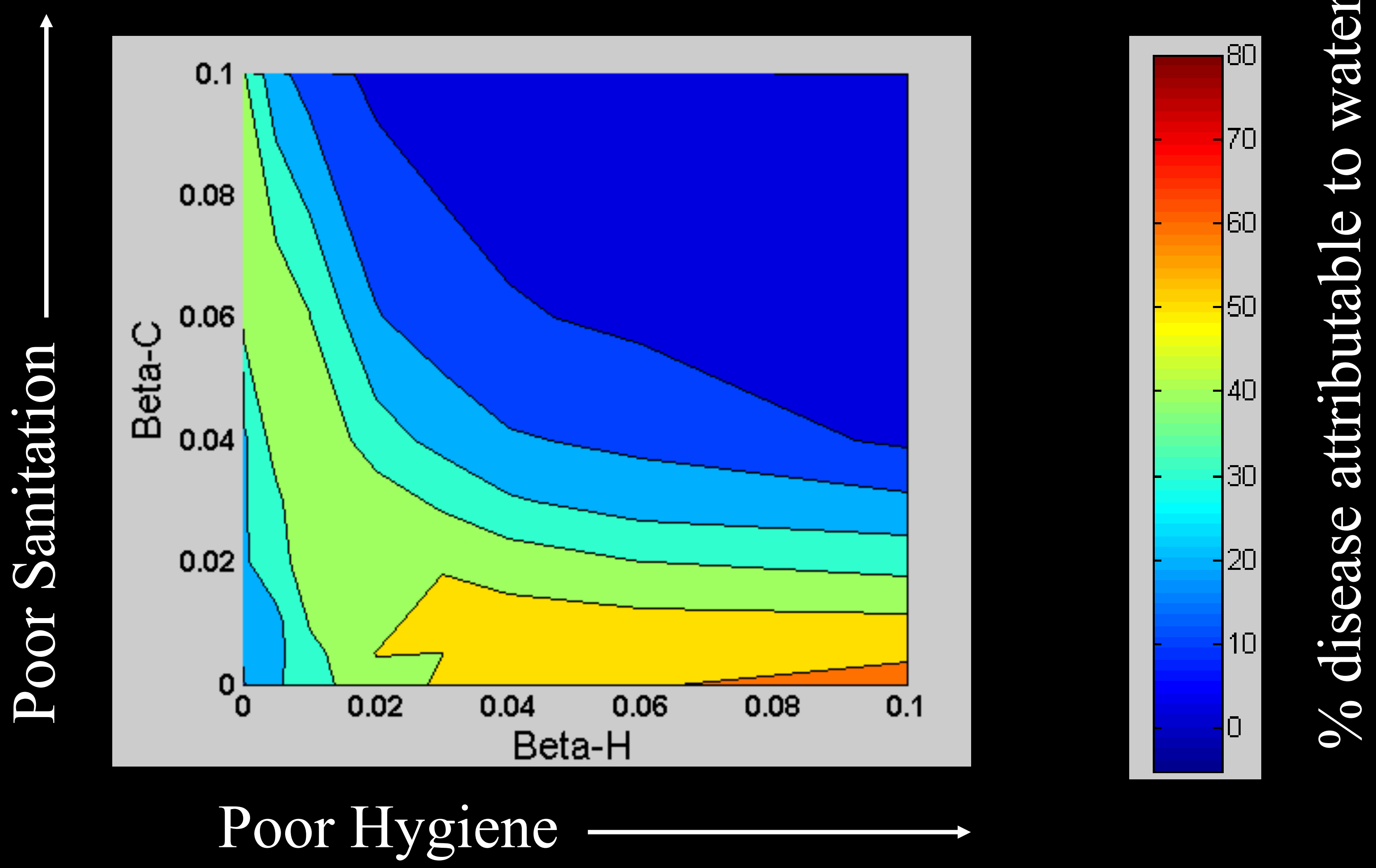
*Note.* $\rho$ = recovery rate; $\phi$ = rate at which infected individuals shed pathogens into the water supply; $\beta_c$ and $\beta_h$ = between- and within-household transmission rates; $\varepsilon$ = level of environmental contamination; $\mu$ = pathogen die-off rate in the water supply; $r$ = risk of infection per pathogen exposure.



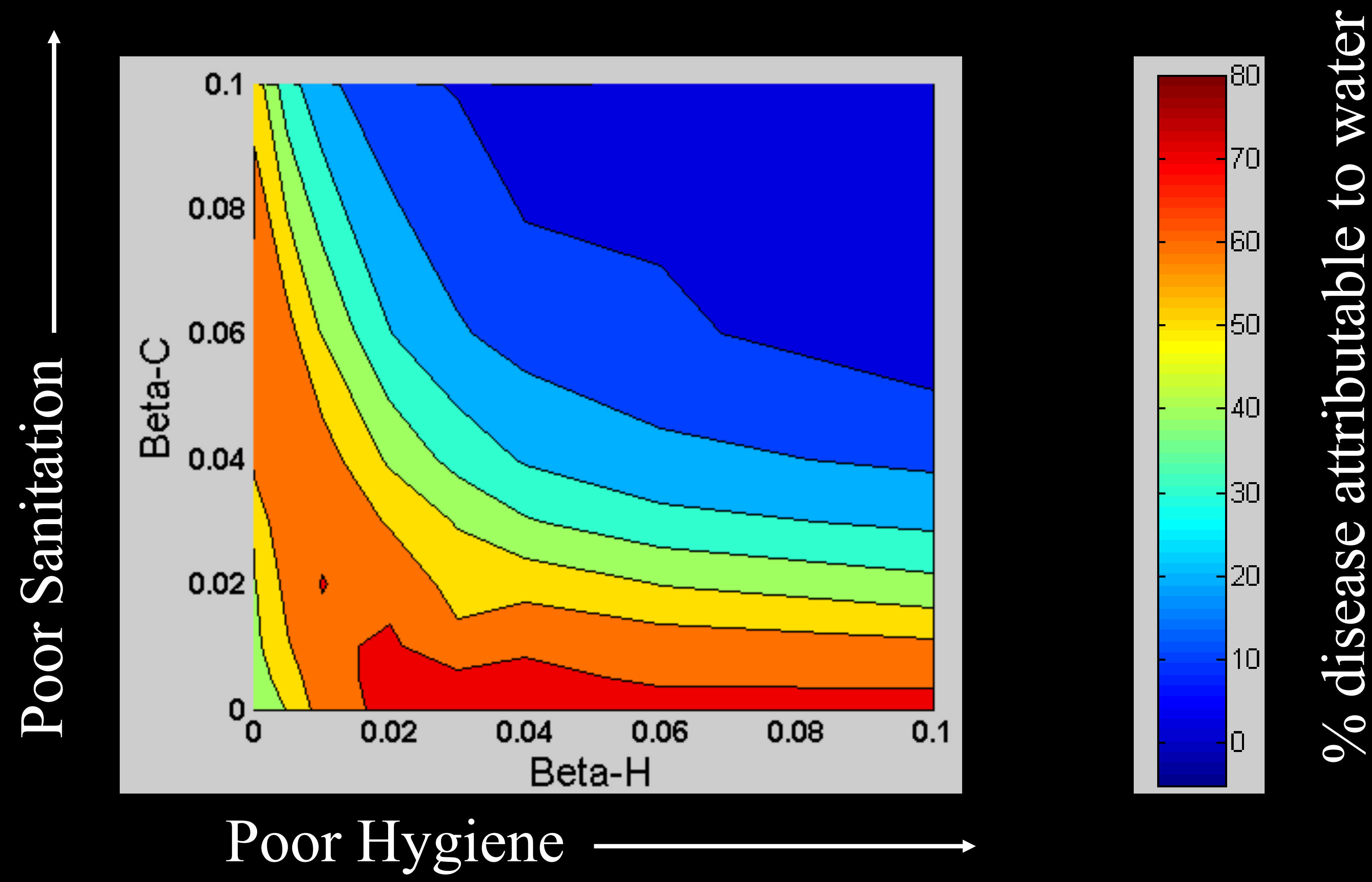
# No shedding of pathogens (contamination) into the water ($\phi = 0$)



# Moderate contamination ($\phi = 1.0$)



# Very high contamination ($\phi = 2.0$)



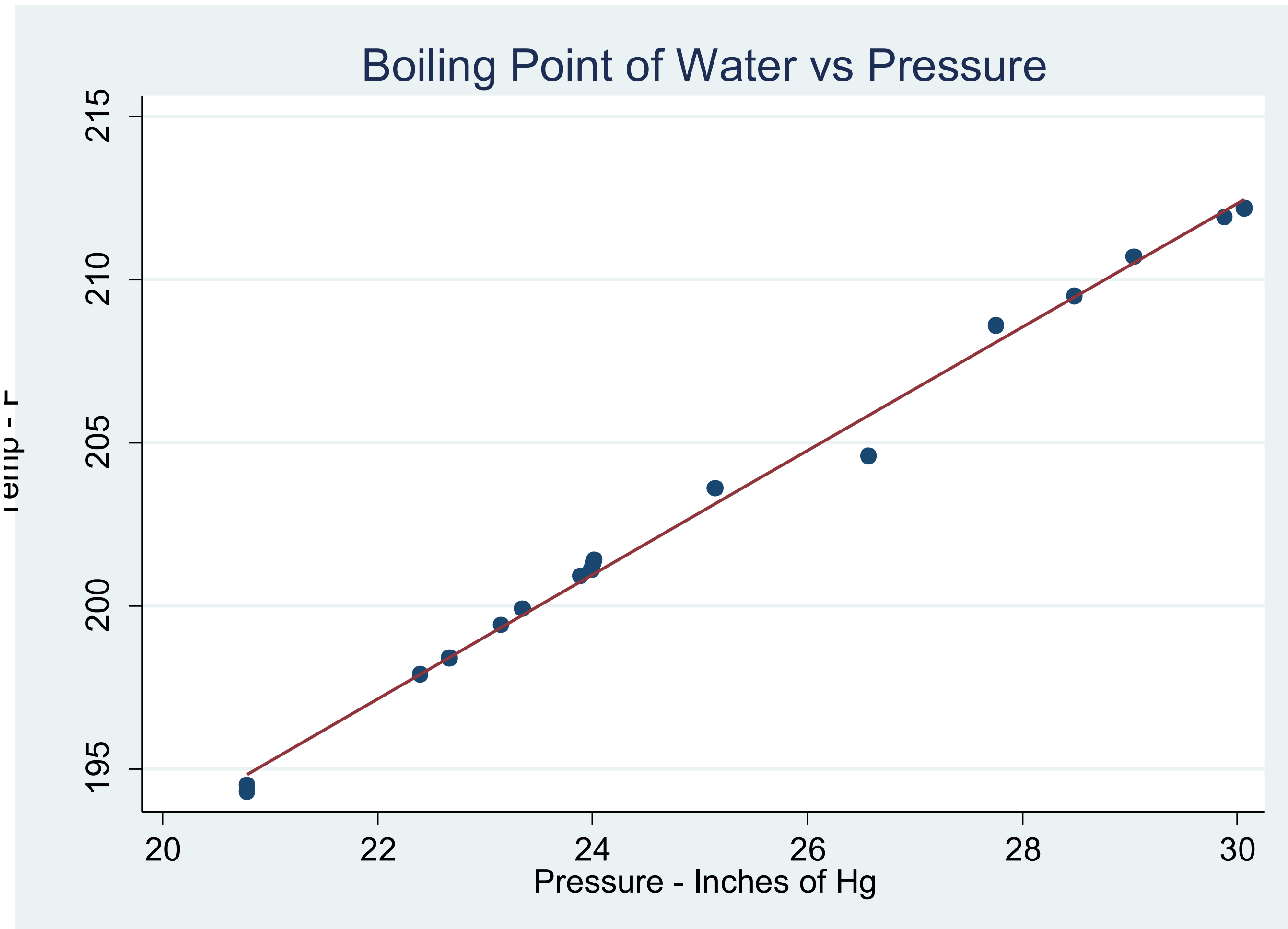
# Desirable Characteristics

- Accurate (i.e. low bias)
- Descriptively realistic
- Precise (i.e. low variability)
- General
- Robust
- Simple / Parsimonious

# Simple / Parsimonious

- A model is *parsimonious* if it can “accomplish a lot without much”
  - E.g. a model that selects a relatively small number of the most useful parameters
  - Simple isn’t always better
  - The research question should drive the complexity of the model

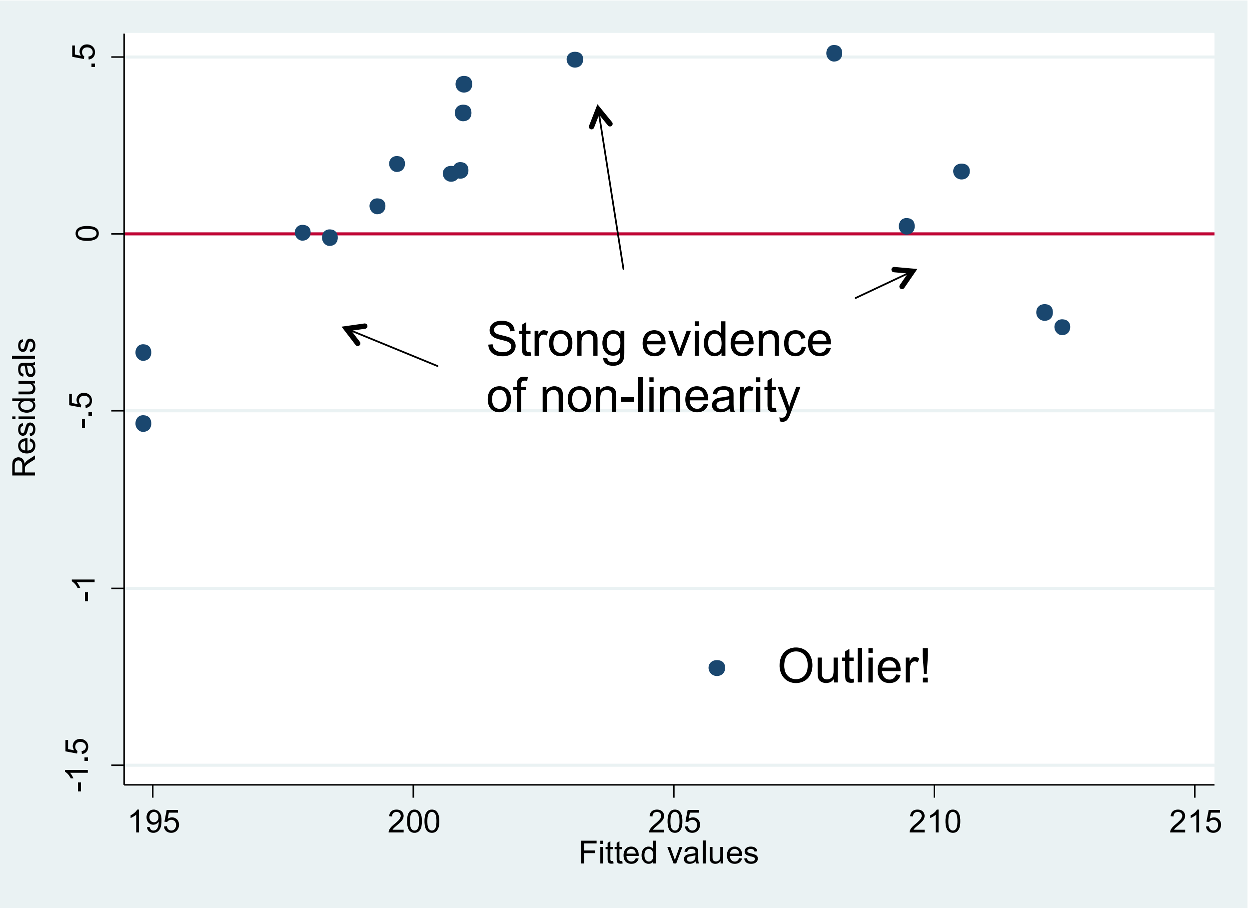
# Boiling Point of $H_2O$ and Pressure




Source: Wikipedia

**r = 0.9972** **Est. Temp = 155.3 + 1.90 * Pressure**

# Is the Relationship Linear?



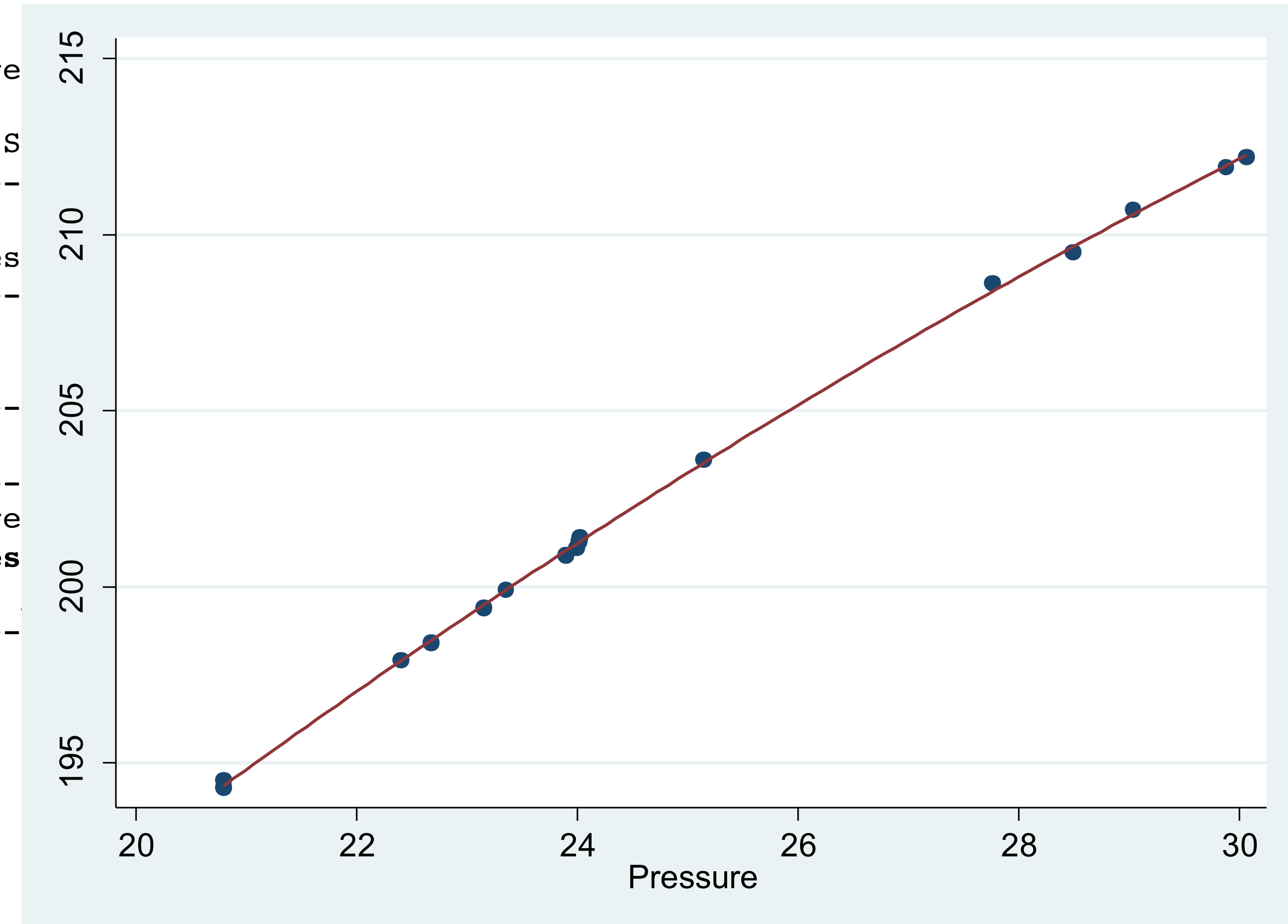

**Hmmm…**

**What does this mean?**

# Significant non-linearity

```
. regre
      S                                  16
-------                                 .88
                                        000
    Res                                 997
-------                                 996
                                        802

-------                                 ---
                                        al]
-------                                 ---
    pre                                 629
   pres                                 322
                                        357
-------                                 ---
```




**Est. Temp = 131.8 + 3.75 Pressure – 0.036 Pressure$^2$**

# Model Comparison

Epidemics 3 (2011) 119–124

Contents lists available at ScienceDirect

Epidemics

journal homepage: www.elsevier.com/locate/epidemics

ELSEVIER


## The epidemiological dynamics of infectious trachoma may facilitate elimination

Thomas M. Lietman [a,b,c,d], Teshome Gebre [e], Berhan Ayele [e], Kathryn J. Ray [a], M. Cyrus Maher [a], Craig W. See [a], Paul M. Emerson [e,f], Travis C. Porco [a,b,c,*]
and The TANA Study Group [a]

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I \quad \text{vs.} \quad \frac{dI}{dt} = \beta\left(\nu_1 \frac{\bar{I}}{N} + \frac{I}{N} + \nu_2 \left(\frac{I}{N}\right)^{\phi+2}\right)(N-I) - \gamma I$$

# AICc: Akaike Information Criterion (correction)

- Used as a model selection tool
  - Penalizes models with excessive parameter spaces
  - $AIC = 2k - 2ln(L)$
  - $AICc = AIC + 2k(k+1) / (n - k - 1)$
- *AICc* is often used to avoid over-fitting when the sample size is small or the parameter space is large
- Lower *AICc* → more parsimonious model

# Model Comparison

| Model | Transmission | | | | Recovery |
|---|---|---|---|---|---|
| | Outside | Within community | | | |
| | $\nu_1$ (95% CI) | $\beta$ (95% CI) | $\nu_2$ (95% CI) | $\phi$ (95% CI) | $\gamma$ (95% CI) |
| 1 | 0.1047 (0.0111, 0.2951) | 0.0139 (0.0072, 0.0214) | 2.706 (1.402, 5.261) | * | 0.0168 (0.0100, 0.0243) |
| 2 | 0.0747 (0.0178, 0.2412) | 0.0192 (0.0156, 0.0522) | 2.673 (1.238, 6.127) | 0.8064 (−0.5674, 1.591) | 0.0173 (0.0133, 0.0414) |
| 3 | * | 0.0143 (0.0072, 0.0287) | 1.779 (0.694, 3.241) | * | 0.0136 (0.0087, 0.0231) |
| 4 | * | 0.0191 (0.0149, 0.0368) | 2.321 (1.565, 6.638) | 1.3137 (−0.5517, 3.131) | 0.0140 (0.0114, 0.0240) |
| 5 | * | 0.0190 (0.0140, 0.0382) | * | * | 0.0109 (0.0079, 0.0226) |
| 6 | 0.0112 (−0.0020, 0.0310) | 0.0196 (0.0129, 0.0302) | * | * | 0.0113 (0.0073, 0.0179) |

$$\frac{dI}{dt} = \frac{\beta SI}{N} - \gamma I$$

$$\frac{dI}{dt} = \beta\left(v1\frac{\bar{I}}{N} + \frac{I}{N} + v2\frac{I}{N}^{2}\right)$$

| Loglikelihood | $AIC_c$ |
|---|---|
| −142.40 | 291.38 |
| −141.84 | 292.66 |
| −144.64 | 293.56 |
| −143.67 | 293.92 |
| −152.81 | 307.72 |
| −151.99 | 308.25 |

“Simple” Model (#5)

“Best” model (#1)

# Desirable Characteristics

- Accurate (i.e. low bias)
- Descriptively realistic
- Precise (i.e. low variability)
- General
- Robust
- Simple / Parsimonious
- Useful
- 

# Useful

- A model is *useful* if:
  - its conclusions are useful
  - it points the way to other good models
  - E.g. Modeling HIV exercise
    - The early models weren't necessarily accurate but they were *useful*

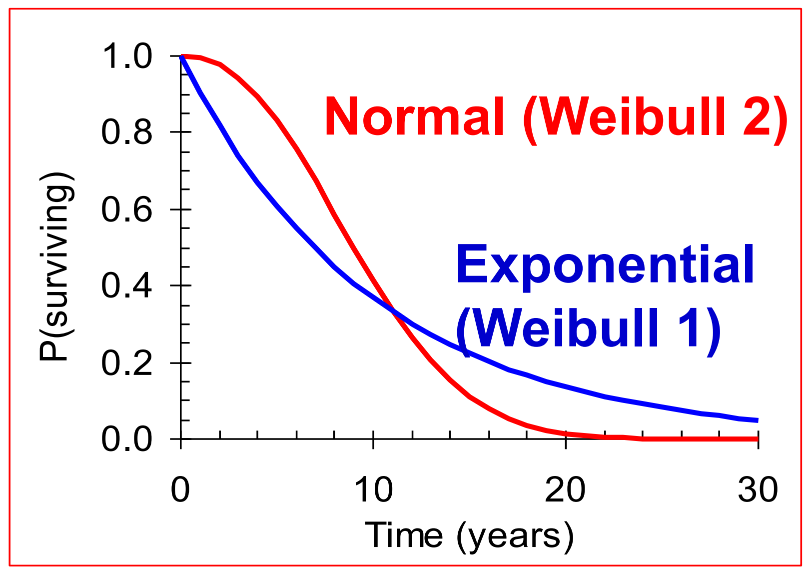
# This is a useful model


$\beta$ = birth rate

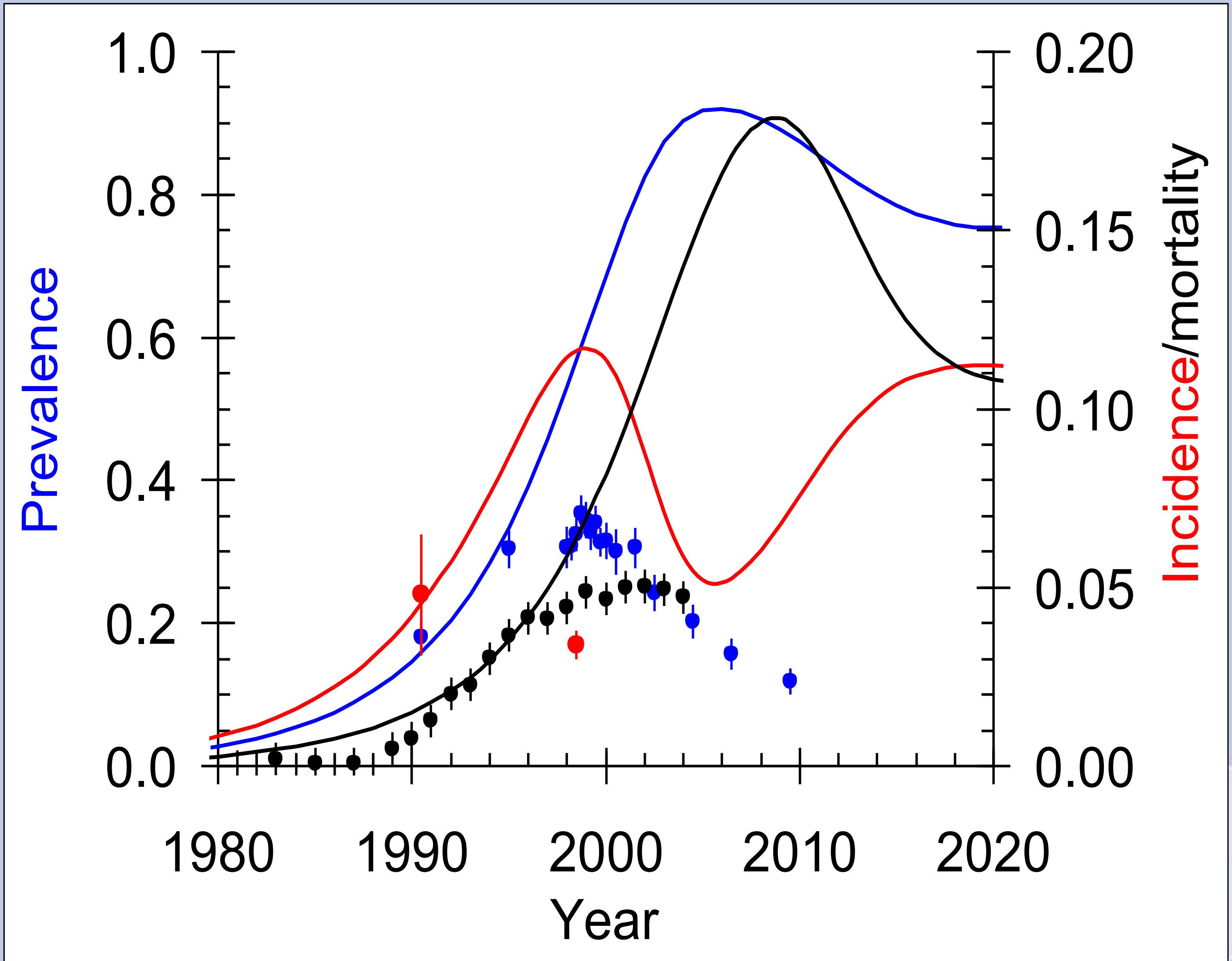
$N$ = $S + I$

$\lambda$ = infection rate

$\overline{\delta I}$ = Weibull mortality

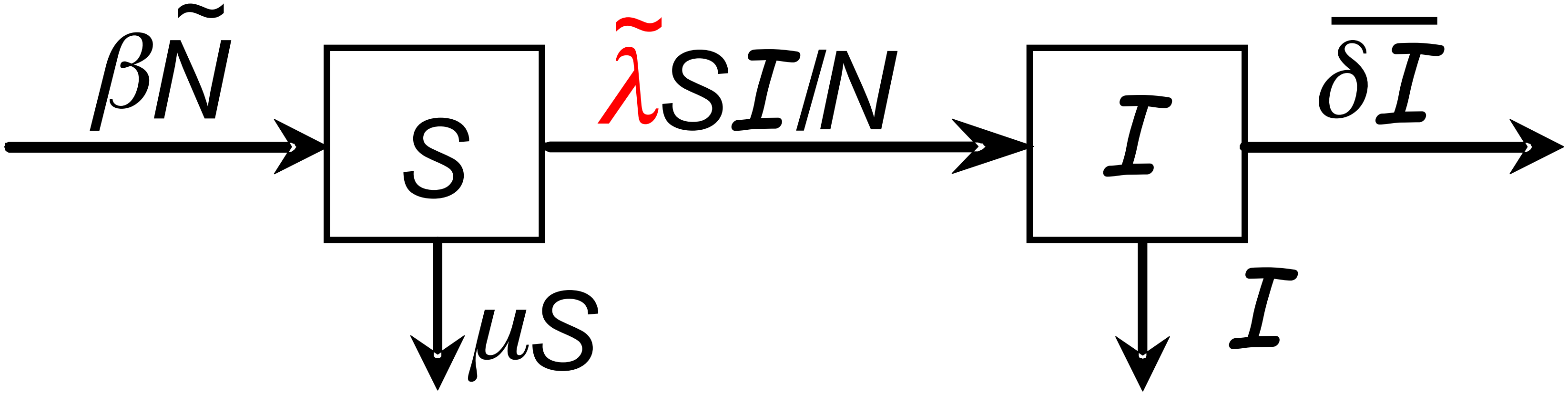



Slide credit: J. Hargrove/B. Williams




SACEMA
DST/NRF Centre of Excellence in Epidemiological Modelling and Analysis
Funded by DST, Administered by NRF, Hosted by Stellenbosch University

Slide credit: J. Hargrove/B. Williams

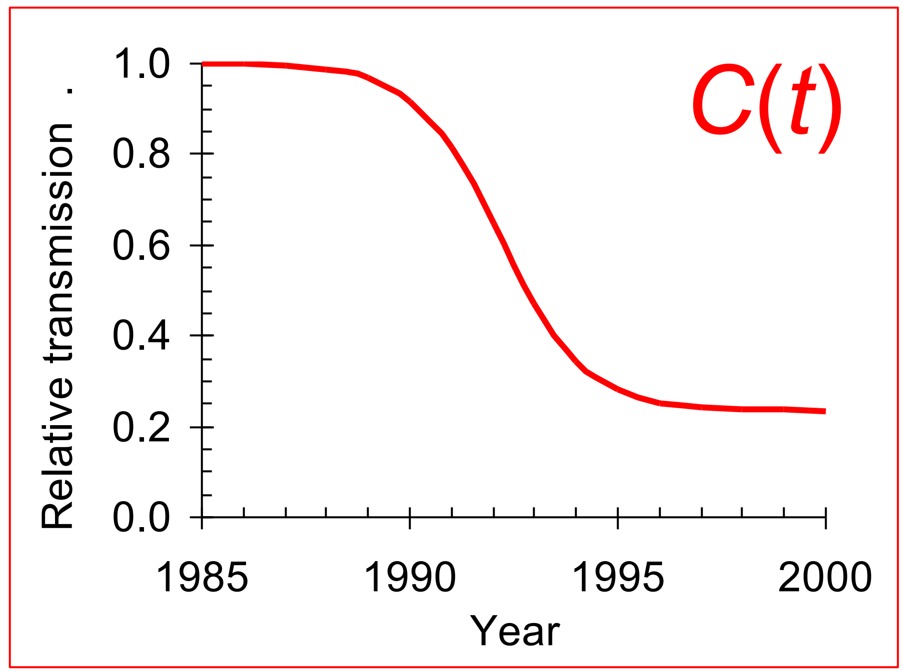



$\beta$ = birth rate

$\tilde{N}$ = population

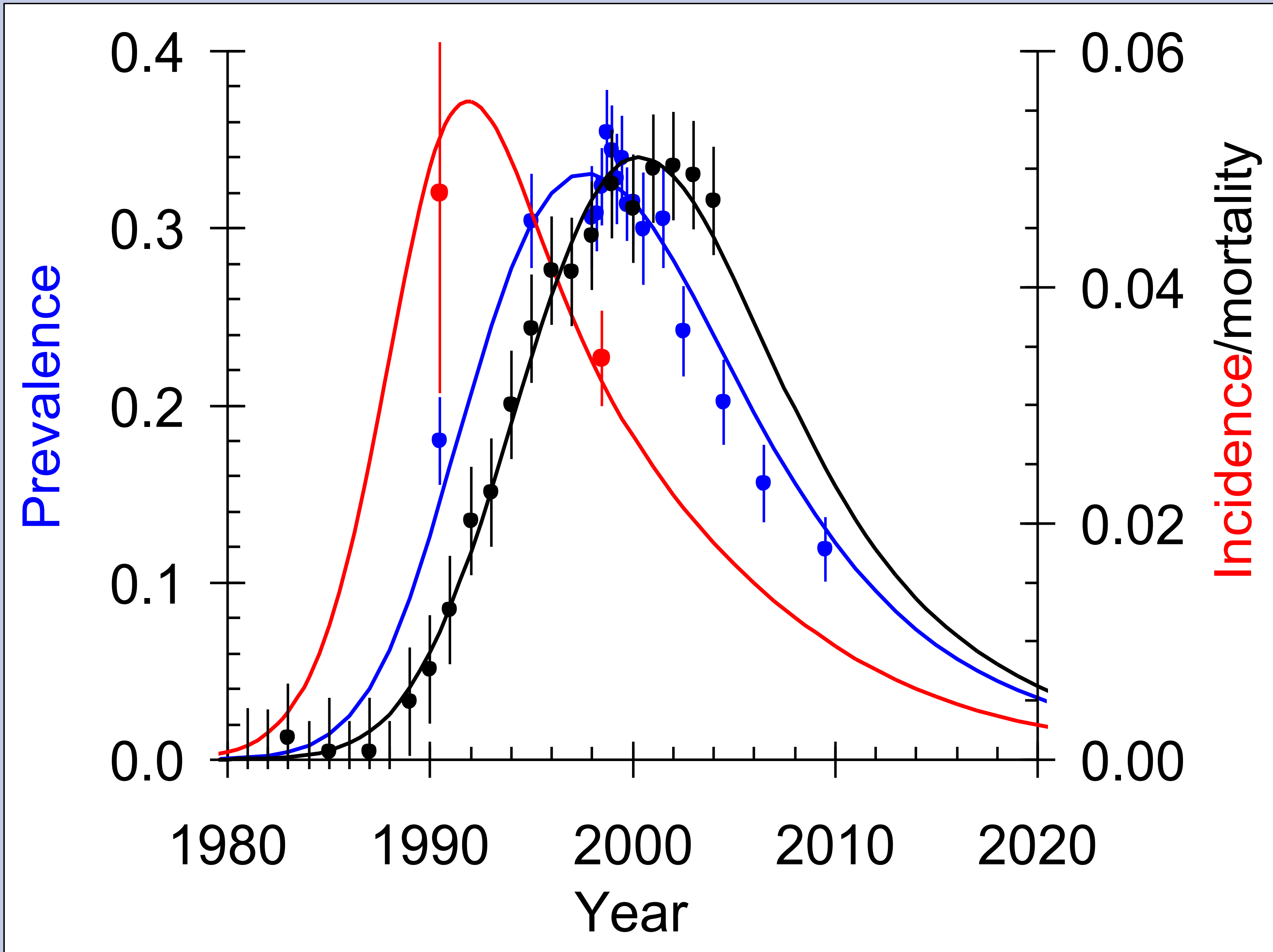
$\tilde{\lambda} = \hat{\lambda}C(t)$

$\overline{\delta I}$ = mortality




Slide credit: J. Hargrove/B. Williams

# Including control




SACEMA

DST/NRF Centre of Excellence in Epidemiological Modelling and Analysis

Funded by DST, Administered by NRF, Hosted by Stellenbosch University

Slide credit: J. Hargrove/B. Williams

# Desirable Characteristics

- Accurate (i.e. low bias)
- Descriptively realistic
- Precise (i.e. low variability)
- General
- Robust
- Simple / Parsimonious
- Useful
- Inexpensive
- Others???

# What makes a model "good"?

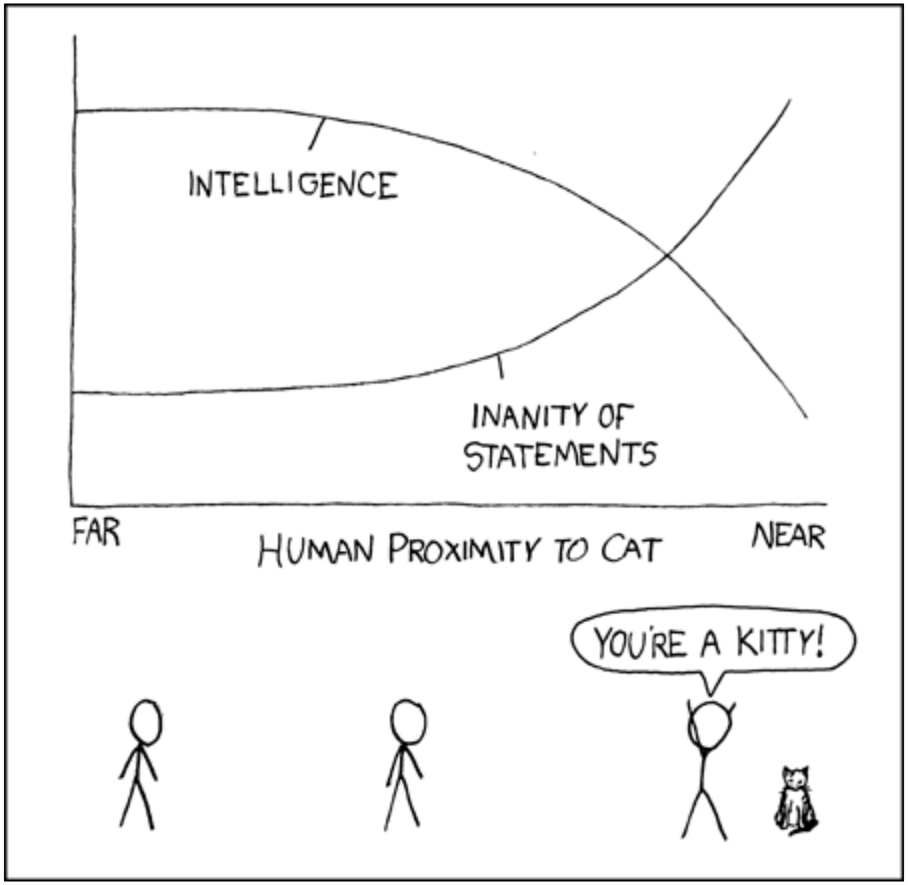
- Accurate (i.e. low bias)
- Descriptively realistic
- Precise (i.e. low variability)
- General
- Robust
- Simple / Parsimonious
- Useful
- Inexpensive
- Others???

# What makes a model "good"?

- It really depends on what your original research question was
  - Was the goal to accurately predict something?
  - Was the goal to determine a relationship between two or more parameters?
  - Was the goal to understand a system in general terms?
  - Was the goal to test a hypothesis?  Or to generate one?

# Exercise

- Consider each of the models presented today
  - What are the good things about each model?
  - What are the shortcomings of each model?
- Final word:




Source: XKCD

# Sources


- Concepts of Mathematical Modeling, Walter Meyer, McGraw-Hill, 1984
- Probability and Statistics, Charles Stone, Duxbury, 1996
- Modeling Infectious Diseases in Humans and Animals, Keeling and Rohoni, Princeton, 2008
- Mathematical Models for Communicable Diseases, Brauer and Castillo-Chavez, SIAM, 2013
- The Analysis of Biological Data, Whitlock and Schluter, Roberts and Company, 2008
- Wikipedia
- XKCD