Bayesian Model Averaging Applied to Tuberculosis and HIV Research Studies

Brock Stewart, PhD
jnn6@cdc.gov

DTBE Brown Bag
October 6, 2016

Bayesian Model Averaging in Tuberculosis Research – couple of motivating examples

OPEN ACCESS Freely available online

PLoS one

Baseline Predictors of Sputum Culture Conversion in Pulmonary Tuberculosis: Importance of Cavities, Smoking, Time to Detection and W-Beijing Genotype

Marianne E. Visser[1]*, Michael C. Stead[2,3], Gerhard Walzl[2], Rob Warren[2], Michael Schomaker[4], Harleen M. S. Grewal[5], Elizabeth C. Swart[6], Gary Maartens[7]

1 School of Public Health, University of the Western Cape, Cape Town, South Africa, **2** MRC Centre for Molecular and Cellular Biology, University of Stellenbosch, Cape Town, South Africa, **3** Division of Medical Microbiology, Institute of Infectious Diseases and Molecular Medicine, University of Cape Town, Cape Town, South Africa, **4** School of Public Health and Family Medicine, Centre for Infectious Disease Epidemiology and Research, University of Cape Town, Cape Town, South Africa, **5** Section of Microbiology and Immunology, The Gade Institute, University of Bergen, Haukeland University Hospital, Bergen, Norway, **6** Division of Dietetics, University of the Western Cape, Cape Town, South Africa, **7** Division of Clinical Pharmacology, Department of Medicine, University of Cape Town, Cape Town, South Africa

Abstract

Background: Time to detection (TTD) on automated liquid mycobacterial cultures is an emerging biomarker of tuberculosis outcomes. The *M. tuberculosis* W-Beijing genotype is spreading globally, indicating a selective advantage. There is a paucity of data on the association between baseline TTD and W-Beijing genotype and tuberculosis outcomes.

Aim: To assess baseline predictors of failure of sputum culture conversion, within the first 2 months of antitubercular therapy, in participants with pulmonary tuberculosis.

Design: Between May 2005 and August 2008 we conducted a prospective cohort study of time to sputum culture conversion in ambulatory participants with first episodes of smear and culture positive pulmonary tuberculosis attending two primary care clinics in Cape Town, South Africa. Rifampicin resistance (diagnosed on phenotypic susceptibility testing) was an exclusion criterion. Sputum was collected weekly for 8 weeks for mycobacterial culture on liquid media (BACTEC MGIT 960). Due to missing data, multiple imputation was performed. Time to sputum culture conversion was analysed using a Cox-proportional hazards model. Bayesian model averaging determined the posterior effect probability for each variable.

Results: 113 participants were enrolled (30.1% female, 10.5% HIV-infected, 44.2% W-Beijing genotype, and 89% cavities). On Kaplan Meier analysis 50.4% of participants underwent sputum culture conversion by 8 weeks. The following baseline factors were associated with slower sputum culture conversion: TTD (adjusted hazard ratio (aHR) = 1.11, 95% CI 1.02; 1.2), lung cavities (aHR = 0.13, 95% CI 0.02; 0.95), ever smoking (aHR = 0.32, 95% CI 0.1; 1.02) and the W-Beijing genotype (aHR = 0.51, 95% CI 0.25; 1.07). On Bayesian model averaging, posterior probability effects were strong for TTD, lung cavitation and smoking and moderate for W-Beijing genotype.

Conclusion: We found that baseline TTD, smoking, cavities and W-Beijing genotype were associated with delayed 2 month sputum culture. Larger studies are needed to confirm the relationship between the W-Beijing genotype and sputum culture conversion.

Journal of Biomedical Informatics 53 (2015) 237–242

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Towards probabilistic decision support in public health practice: Predicting recent transmission of tuberculosis from patient attributes

Hiroshi Mamiya [a,e,*], Kevin Schwartzman [d,e,f], Aman Verma [a,e], Christian Jauvin [a], Marcel Behr [b,e,f], David Buckeridge [a,c,e]

[a] McGill Clinical and Health Informatics, McGill University, 1140 Avenue Pine, Montréal, Québec H3A 1A3, Canada
[b] McGill University Health Centre, 1650 Cedar Avenue, Room A5.156, Montreal, H3G 1A4, Canada
[c] Agence Sociosanitaire de Montréal, Direction de la santé publique, 1301 Rue Sherbrooke Est, Montreal, Quebec H2L 1M3, Canada
[d] Respiratory Epidemiology and Clinical Research Unit, Montreal Chest Institute, McGill University Health Centre, 3650 Rue Saint-Urbain, Montreal, Quebec H2X 2P4, Canada
[e] Department of Epidemiology, Biostatistics, and Occupational Health, McGill University, 1020 Pine Avenue West, Montreal, Quebec H3A 1A2, Canada
[f] McGill International TB Centre, McGill University Health Centre, 1650 Cedar Avenue, Room A5.156, Montreal, Quebec, H3G 1A4, Canada

ARTICLE INFO

Article history:
Received 29 September 2014
Accepted 11 November 2014
Available online 20 November 2014

Keywords:
Decision support model
Statistical prediction
Public health
Tuberculosis
Transmission

ABSTRACT

Objective: Investigating the contacts of a newly diagnosed tuberculosis (TB) case to prevent TB transmission is a core public health activity. In the context of limited resources, it is often necessary to prioritize investigation when multiple cases are reported. Public health personnel currently prioritize contact investigation intuitively based on past experience. Decision-support software using patient attributes to predict the probability of a TB case being involved in recent transmission could aid in this prioritization, but a prediction model is needed to drive such software.

Methods: We developed a logistic regression model using the clinical and demographic information of TB cases reported to Montreal Public Health between 1997 and 2007. The reference standard for transmission was DNA fingerprint analysis. We measured the predictive performance, in terms of sensitivity, specificity, negative predictive value, positive predictive value, the Receiver Operating Characteristic (ROC) curve and the Area Under the ROC (AUC).

Results: Among 1552 TB cases enrolled in the study, 314 (20.2%) were involved in recent transmission. The AUC of the model was 0.65 (95% confidence interval: 0.61–0.68), which is significantly better than random prediction. The maximized values of sensitivity and specificity on the ROC were 0.53 and 0.67, respectively.

Conclusions: The characteristics of a TB patient reported to public health can be used to predict whether the newly diagnosed case is associated with recent transmission as opposed to reactivation of latent infection.

© 2014 Elsevier Inc. All rights reserved.

Limitations of this presentation

- This is not a "how-to"
- For a tutorial, see
-  Caution: some minimal use of notation

Statistical Science
1999, Vol. 14, No. 4, 382–417

Bayesian Model Averaging: A Tutorial

Jennifer A. Hoeting, David Madigan, Adrian E. Raftery and Chris T. Volinsky

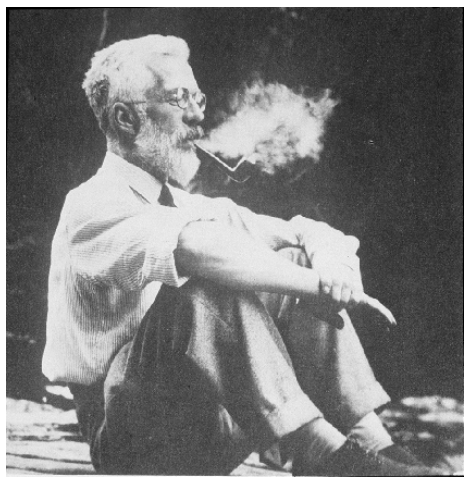
Abstract. Standard statistical practice ignores model uncertainty. Data analysts typically select a model from some class of models and then proceed as if the selected model had generated the data. This approach ignores the uncertainty in model selection, leading to over-confident inferences and decisions that are more risky than one thinks they are. Bayesian model averaging (BMA) provides a coherent mechanism for accounting for this model uncertainty. Several methods for implementing BMA have recently emerged. We discuss these methods and present a number of examples. In these examples, BMA provides improved out-of-sample predictive performance. We also provide a catalogue of currently available BMA software.

Key words and phrases: Bayesian model averaging, Bayesian graphical models, learning; model uncertainty, Markov chain Monte Carlo.

CONTENTS

Goals of this presentation

- Describe Bayesian Model Averaging (BMA)
- Give a very, *very* brief intro to the "Bayesian" part of BMA
- Describe what one may do with BMA
- Provide brief real-life examples of BMA in research studies



An absurdly crude history of "modern" Statistics [wikipedia]

- **mid 1700's:** The term "statistics" for demographic, economic data by states, mostly for tax, war purposes – Latin *statisticum collegium*, "council of state"
- **Late 1700's:** Bayes, Laplace – Establish "Bayes' theorem"
- **1800's – early 1900's:** Gauss, Pearson, Fisher, Neyman, Wald – Large-sample theory, least squares, use of probability theory, experimental design, hypothesis testing
- **Most of 1900's:** Bayesian-style ideas and proponents largely dismissed
- **1950** The term "Bayesian" used by those dissatisfied with limitations of frequentist statistics
- **1990:** Bayesian approach sees revival, less acrimony between frequentist and Bayesian, advancements in computing and algorithms

Which hoe is the best?

No need to choose only one and stick with it forever, acquire and learn to use both

When you need a hoe, use whichever you believe is best for the situation

How are "Bayesian Statistics" different from "Statistics"

- Bayesian approach inherently promotes thought, description, and reasoning rather than cookbook or black-box solutions
  - Note, this implies more work for the analyst, but in a good way
- All stat's you've learned may be called "frequentist" statistics

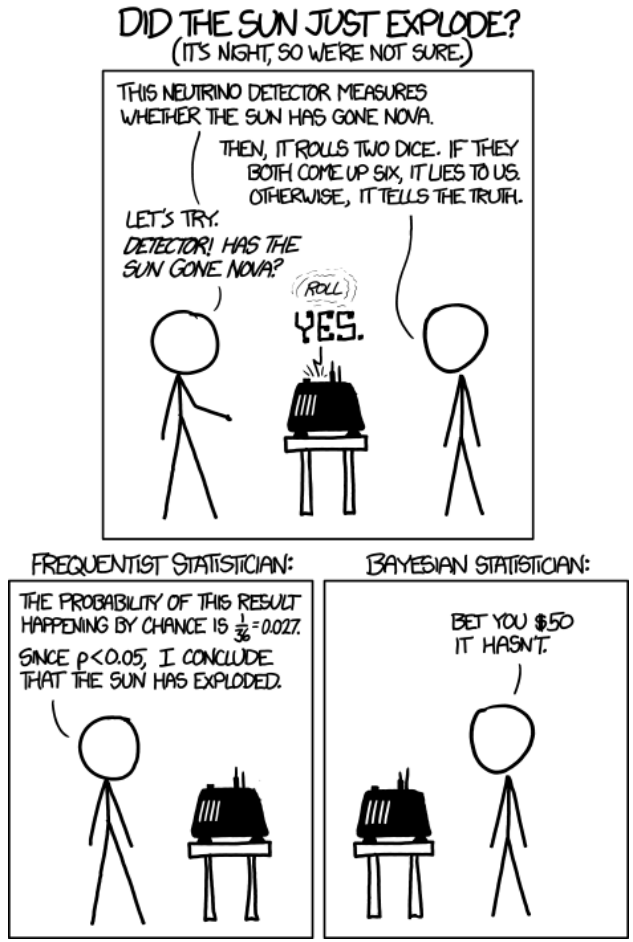
Frequentist

- $Pr(Y|H_0)$
- Parameters are fixed, data are random
- Uncertainty: frequency in hypothetical repeated samples
- Conclusions based on point estimates, CI, p-values

Bayesian

- $Pr(H_0|Y)$
- Data are fixed, parameters are random
- Uncertainty: random parameters
- Conclusions based on the complete *posterior* probability

Frequentist vs. Bayesian – XKCD



Not all experiments are repeatable

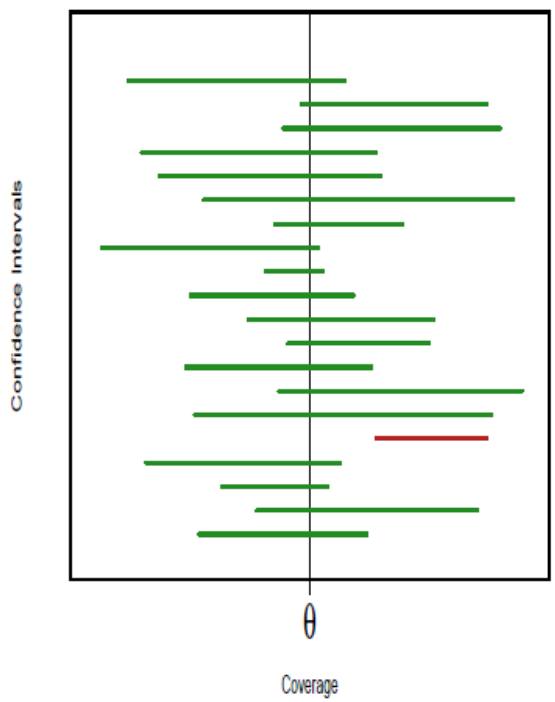
Frequentist vs. Bayesian 90% Intervals

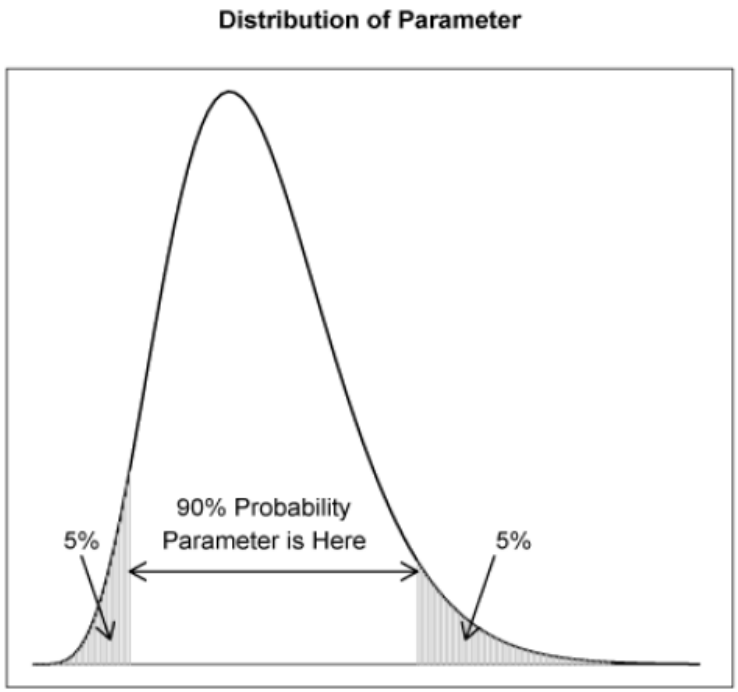
Frequentist

- If we repeat the experiment many times, 90% of the CI will encompass the true value *and* 10% will *not*

Bayesian

- Given this data, this is the interval that has 90% chance of containing the true value





Baye's Rule

To make probability statements about θ given y, we start with a model. The model is a joint probability of θ and y. Specify a *prior* distribution on the paraemter(s):

$$p(\theta)$$

Called *prior* to reflect our belief before considering data. The data are generated from a *sampling distribution*:

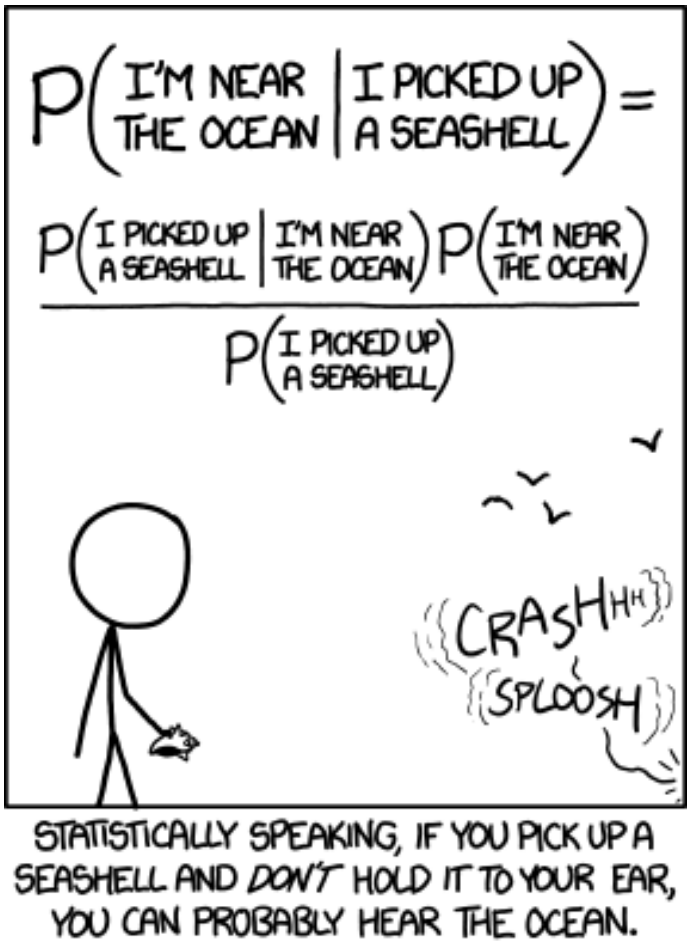
$$p(y|\theta)$$

From the probability axioms (laws), we can get

$$p(\theta, y) = p(\theta)p(y|\theta) = p(y)p(\theta|y)$$

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{p(y)}$$

Baye's Rule – XKCD



$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

$\theta =$ "I'm Near The Ocean"
$y =$ "I Picked Up a Seashell"

In this case, for me,

$$p(y|\theta) \approx p(y)$$

General layout of a Bayesian model

Data is generated from (e.g. Normal, Binomial, Poisson):

$$p(y|\theta)$$

But we don't know the value of θ, and we reflect our uncertainty by

$$p(\theta)$$

What we really want is to update our knowledge of θ using the data

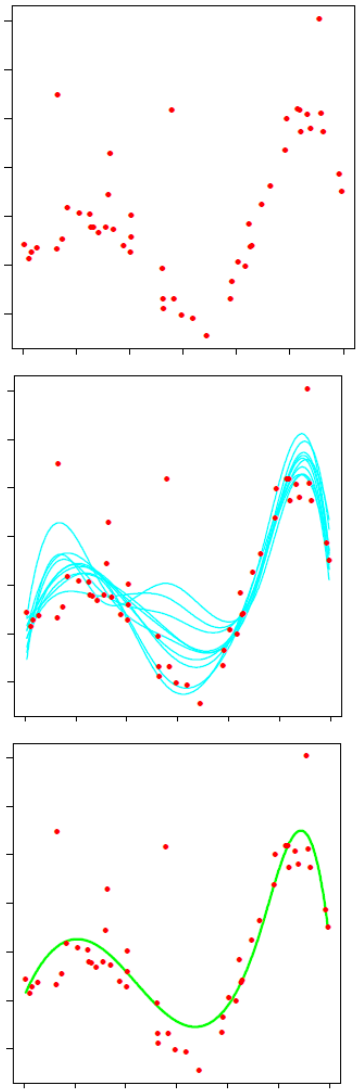
$$p(\theta|y)$$

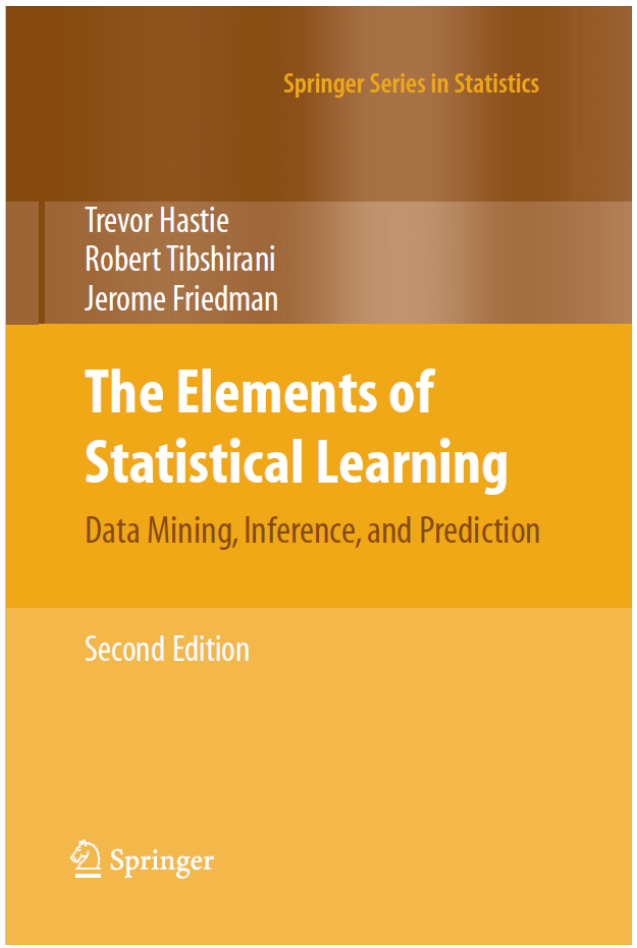
Which we can obtain from Bayes' Rule

$$p(\theta|y) = \frac{p(\theta)p(y|\theta)}{p(y)}$$

Uses of BMA

- Combine models of differing forms for prediction
- Choose between model forms
- Choose which variables are important





Model Uncertainty

(Hoeting, et al. (1999), abstract)

- When we choose a model form, e.g. logistic, we make the tacit assumption that the observed data were generated from that model
- When we choose which independent variables go in a model we make a similar assumption
- This approach ignores the uncertainty in model selection, leading to over-confident inferences and decisions

Bayesian Model Averaging provides a method for accounting for this model uncertainty

BMA Inferential Products

- **A superior prediction model**
  - Predictions = weighted average of predictions from all models.
  - RMSE is never worse than any of the component models.

- **Model (variable) selection**
  - Which independent variables should be included in the model?
  - Stepwise et al. do not address uncertainty in final choice.
  - Choose model with largest posterior probability.

- **Variable importance**
  - $p(\beta \neq 0|y)$ is an output.
  - Sum the probabilities of models containing β.
  - Serves as a variable importance measure.

Variable Selection & Importance

- This is my personal motivation for pursuing BMA
- A common data analytic scenario in our work:
  - Clearly identified dependent variable
  - Solid choice for model form (e.g. logistic, Cox PH, etc.)
  - Large number of available independent variables
  - Little idea which independent variables are actually important
- BMA gives us a measure – the posterior model probabilities – that serves as an objective ranking system for different subsets of independent variables
- BMA gives us a measure – $p(\beta \neq 0|y)$ – that objectively quantifies just how important is a potential independent variable

3HP Discontinuation – Background

(from the unpublished manuscript in clearance)

- **Importance:**
  Recent clinical trials showed that a new treatment for latent *Mycobacterium tuberculosis* infection (LTBI), 12 weekly doses of directly observed isoniazid and rifapentine (3HP) was as efficacious as 9 months of isoniazid (9H) with greater completion rates.

- **Research Question:**
  Will these favorable completion rates also be present in non-research settings?

- **Objective:**
  Evaluate 3HP treatment completion rates in non-research setting.

3HP Discontinuation Study– BMA Results (abridged)

Factor	$p(\beta \neq 0 \mid Data)$	EV	M_1	M_2	M_3	M_4	M_5
Intercept	100	-2.24	-2.23	-2.24	-2.17	-2.17	-2.30
Age	100.0	0.01	0.01	0.01	0.01	0.01	0.01
Gender	0.0						
raceEth	0.0						
RxReasonContact	90.3	-0.47	-0.53	-0.54	-0.55	-0.54	-0.47
newhomeless	49.1	0.32	0.66			0.64	
newcorrections	20.9	0.09					0.42
RxreasonStudent	20.8	-0.19			-0.95	-0.91	
RxreasonHealthcareworker	0.0	0.00					
StateCode	100.0						
nVar			4	3	4	5	4
post prob			0.29	0.27	0.10	0.08	0.07

9 independent variables considered; $2^9 = 512$ possible models.
11 models fit in "Occam's Window", top 5 shown above.
Bottom 2 rows: no. variables, $p(M_k|y)$
EV = expected value, i.e. coefficient averaged over all models.
$M_1 \ldots M_5$ = coefficients in models 1 ... 5.
Coefficients of factors having multiple levels are omitted.

3HP Discontinuation Study– BMA Results (abridged)

Factor	$p(\beta \neq 0 \mid Data)$	EV	M_1	M_2	M_3	M_4	M_5
Intercept	100	-2.24	-2.23	-2.24	-2.17	-2.17	-2.30
Age	100.0	0.01	0.01	0.01	0.01	0.01	0.01
Gender	0.0						
raceEth	0.0						
RxReasonContact	90.3	-0.47	-0.53	-0.54	-0.55	-0.54	-0.47
newhomeless	49.1	0.32	0.66			0.64	
newcorrections	20.9	0.09					0.42
RxreasonStudent	20.8	-0.19			-0.95	-0.91	
RxreasonHealthcareworker	0.0	0.00					
StateCode	100.0						
nVar			4	3	4	5	4
post prob			0.29	0.27	0.10	0.08	0.07

Top 5 models accounted for 82% of the total posterior probability.
Top 2 models accounted for 56% of the total posterior probability.
This indicates moderate model uncertainty.
Age, RxReasonContact, StateCode are highly important.
newhomeless is of moderate importance.
newcorrections, RxreasonStudent are of low importance.
Gender, raceEth, RxreasonHealthcareworker are of very low importance.

Chuck Rose & Yi Pan (DHAP): Assessing new HIV cases

factor	$P(p \neq 0)$	EV	M_1	M_2
Intercept	100.00	3.7	x	x
Factor A	100.00	-0.06	x	x
Factor B	0.0	1.3		
Factor C	9.7	1.7e-5		x
Factor D	99.8	4.5e-2		
nVar			2	3
$P(M_k\|D)$			0.828	0.09

Also 9 independent variables, leading to $2^9 = 512$ possible models.
All 512 models assessed; 4 variables and 2 models shown above.
Best model accounted for 83% of the total posterior probability.
2nd best model only accounted for 9%.
The remaining 510 models only accounted for 8% combined.
This indicates a *very* low level of model uncertainty.
2 variables, "Factor A" and "Factor G" (not shown) were dominant in predicting new HIV cases.

Var (ordered by $P(\beta \neq 0 \mid D)$)	β [95% CI]	P-value	μβ [95% HDI]	$P(\beta \neq 0 \mid D)$
Chronic wound (pos. dis.)	4.70 [2.27 - 9.73]	<0.0001	4.99 [2.44 - 10.51]	0.9999
Male sex	2.22 [1.33 - 3.71]	0.0024	2.61 [1.70 - 4.01]	0.9998
MSA colonization	7.71 [3.60 -16.51]	<0.0001	4.11 [2.14 - 7.86]	0.9996
Discharge to a nursing home	2.65 [1.41 - 4.99]	0.0013	2.20 [1.18 - 4.22]	0.8621
CVC (insert post. dis.)			3.77 [0.95 - 14.34]	0.7714
Charlson Score	1.35[1.17 - 1.55]	<0.0001	1.20 [1.01 - 1.45]	0.7527
antibiotic exposure (tot.)			1.85 [0.84 - 3.91]	0.6892
Dialysis			1.58 [0.16 - 2.15]	0.6109
CVC (present at discharge)	2.16 [1.13 - 4.11]	0.0253	1.77 [0.85 - 3.75]	0.5408
Atitibiotic exposure (out.)			1.41 [0.77 - 2.52]	0.4603
Decolonization therapy			0.66 [0.20 - 2.41]	0.4447
Dialysis (out.)			2.73 [0.56 - 12.60]	0.4332
Hospital admission			1.34 [0.67 - 2.61]	0.4321
Admission diagnosis	1.84 [1.05 - 3.22]	0.0405	1.68 [0.93 - 3.15]	0.4152
Discharge with non-CVC invasive device	3.03 [1.24 - 7.39]	0.0096	2.23 [0.87 - 6.00]	0.3917

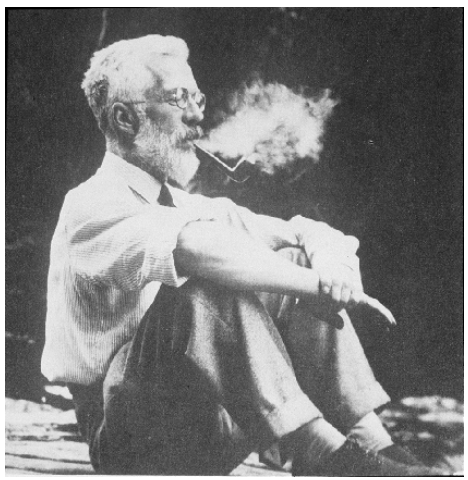
Study of patients developing Methicillin-Resistant Staphylococcus aureus (MRSA) infection within 3 months of hospital discharge, a prospective matched (1:2) case-control study.

Var (ordered by $P(\beta \neq 0 \mid D)$)	β [95% CI]	P-value	μβ [95% HDI)	$P(\beta \neq 0 \mid D)$
Chronic wound (pos. dis.)	4.70 [2.27 - 9.73]	<0.0001	4.99 [2.44 - 10.51]	0.9999
Male sex	2.22 [1.33 - 3.71]	0.0024	2.61 [1.70 - 4.01]	0.9998
MSA colonization	7.71 [3.60 -16.51]	<0.0001	4.11 [2.14 - 7.86]	0.9996
Discharge to a nursing home	2.65 [1.41 - 4.99]	0.0013	2.20 [1.18 - 4.22]	0.8621
CVC (insert post. dis.)			3.77 [0.95 - 14.34]	0.7714
Charlson Score	1.35[1.17 - 1.55]	<0.0001	1.20 [1.01 - 1.45]	0.7527
antibiotic exposure (tot.)			1.85 [0.84 - 3.91]	0.6892
Dialysis			1.58 [0.16 - 2.15]	0.6109
CVC (present at discharge)	2.16 [1.13 - 4.11]	0.0253	1.77 [0.85 - 3.75]	0.5408
Atitibiotic exposure (out.)			1.41 [0.77 - 2.52]	0.4603
Decolonization therapy			0.66 [0.20 - 2.41]	0.4447
Dialysis (out.)			2.73 [0.56 - 12.60]	0.4332
Hospital admission			1.34 [0.67 - 2.61]	0.4321
Admission diagnosis	1.84 [1.05 - 3.22]	0.0405	1.68 [0.93 - 3.15]	0.4152
Discharge with non-CVC invasive device	3.03 [1.24 - 7.39]	0.0096	2.23 [0.87 - 6.00]	0.3917

BMA gives a more clear, broad picture.
Overstated effects? MSA colonization 7.71 vs. 4.11.
CVC eliminated, but has 77% probability of inclusion.
Admission diagnosis has $p<0.05$, but has moderate inclusion prob.

Special **Thank You** to
Bunie Nwana, Amy Sandul, Christine Ho (DTBE)
Chuck Rose, Yi Pan (DHAP)
Yi Mu (NCEZIED/DHQP)

Questions?



Now let's all sit in quiet contemplation of this quagmire.
(Anonymous)