Using Bayesian Methods to Predict Self-Discharge in the Sydney Children's Hospital Network

Presented by
Dr. Gordon McDonald
Sydney Informatics Hub
The University of Sydney

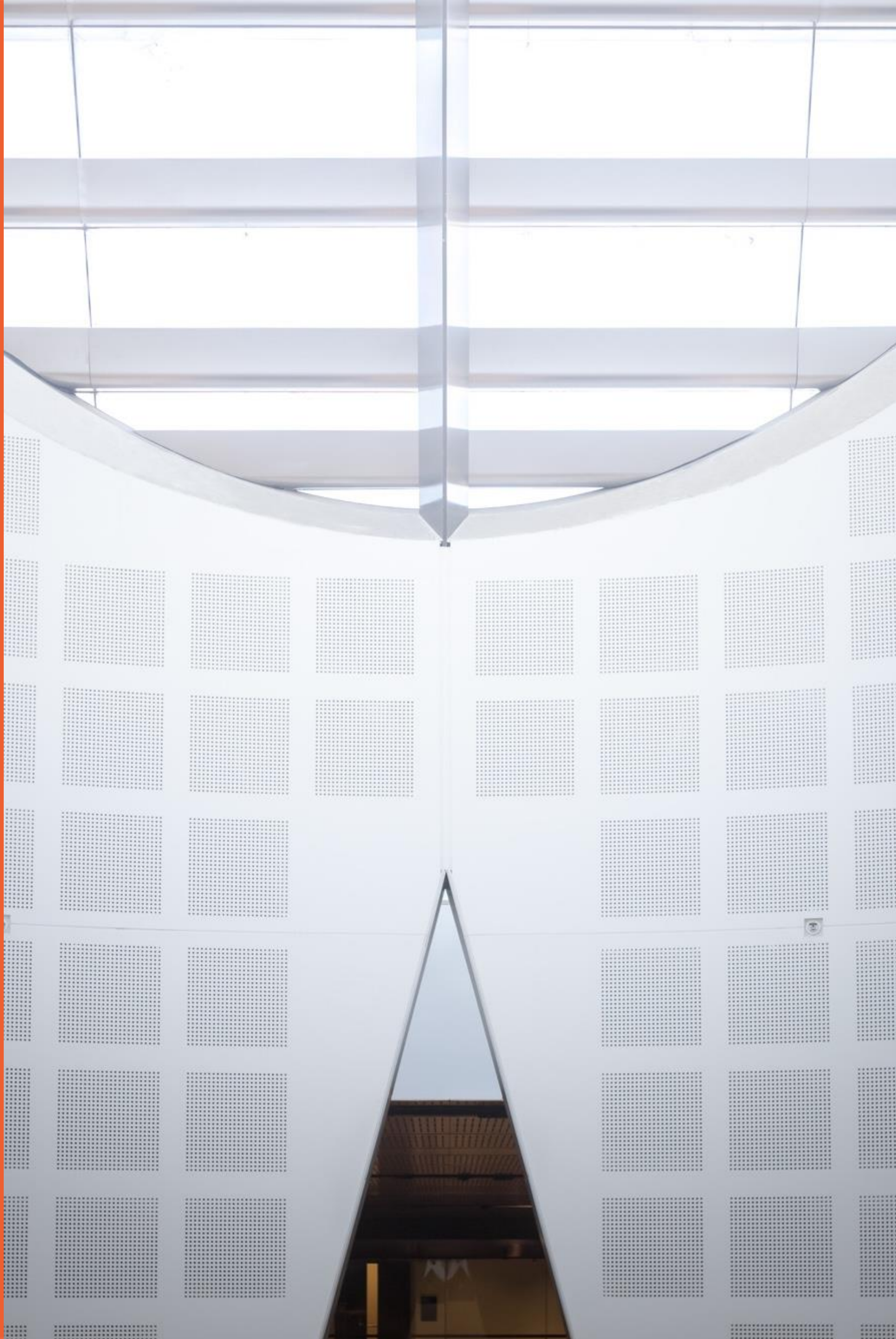
UNSW SYDNEY

The Sydney children's Hospitals Network
care, advocacy, research, education

On behalf of

Dr. Louise Sealy
Prof. Karen Zwi
Lisa Crawford
Prof. Hasantha Gunasekera
Dr. Gordon McDonald
Dr. Aldo Saavedra

THE UNIVERSITY OF SYDNEY



Literature

Subgroup	Place	Outcome	Rate / Risk / OR	Uncert.	Study
Aboriginal Children	NSW	Avoidable Hospitalizations	Rate = 2x	0.02x	1
Aboriginal Heart patients	WA	DAMA	OR = 2.3	1.5 - 3.5	2
Emergency admission Heart patients	WA	DAMA	OR = 6	3 - 12	2
Heart patients with previous alcohol related admission	WA	DAMA	OR = 3	2 - 4	2
Aboriginal Patients	Australia	DAMA	5x to 11x depending on age	-	3
Age group	Australia	DAMA	DAMA rate peaks around 30, lower for children and elderly	-	3
Rurality	Australia	DAMA	Higher DAMA rate for more rural areas	-	3
Children's Hospital	Sydney	DAMA	?????		This one

Literature

[1] Judith M Katzenellenbogen, Frank M Sanfilippo, Michael ST Hobbs, Matthew W Knuiman, Dawn Bessarab, Angela Durey and Sandra C Thompson. Voting with their feet - predictors of discharge against medical advice in Aboriginal and non-Aboriginal ischaemic heart disease inpatients in Western Australia: an analytic study using data linkage. **BMC Health Services Research,** 2013 Vol. 13, Issue 1, page 330. DOI: http://dx.doi.org/10.1186/1472-6963-13-330

[2] Kathleen Falster, Emily Banks, Sanja Lujic, Michael Falster, John Lynch, Karen Zwi, Sandra Eades, Alastair H. Leyland and Louisa Jorm. Inequalities in pediatric avoidable hospitalizations between Aboriginal and non-Aboriginal children in Australia: a population data linkage study. **BMC Pediatrics,** 2016 Vol. 16, Issue 1, page 169. DOI: http://dx.doi.org/10.1186/s12887-016-0706-7

[3] Department of Prime Minister and Cabinet, Aboriginal and Torres Strait Islander Health Performance Framework 2014 Report, Section 3.09 - Discharge against medical advice.
Url: https://www.dpmc.gov.au/sites/default/files/publications/indigenous/Health-Performance-Framework-2014/tier-3-health-system-performance/309-discharge-against-medical-advice.html

The Sydney children's Hospitals Network
care, advocacy, research, education

Question: Which covariates predict **discharge against medical advice**?

5 years of Admissions : 2011-2015

250k admissions from 125k patients

Link by postcode to:

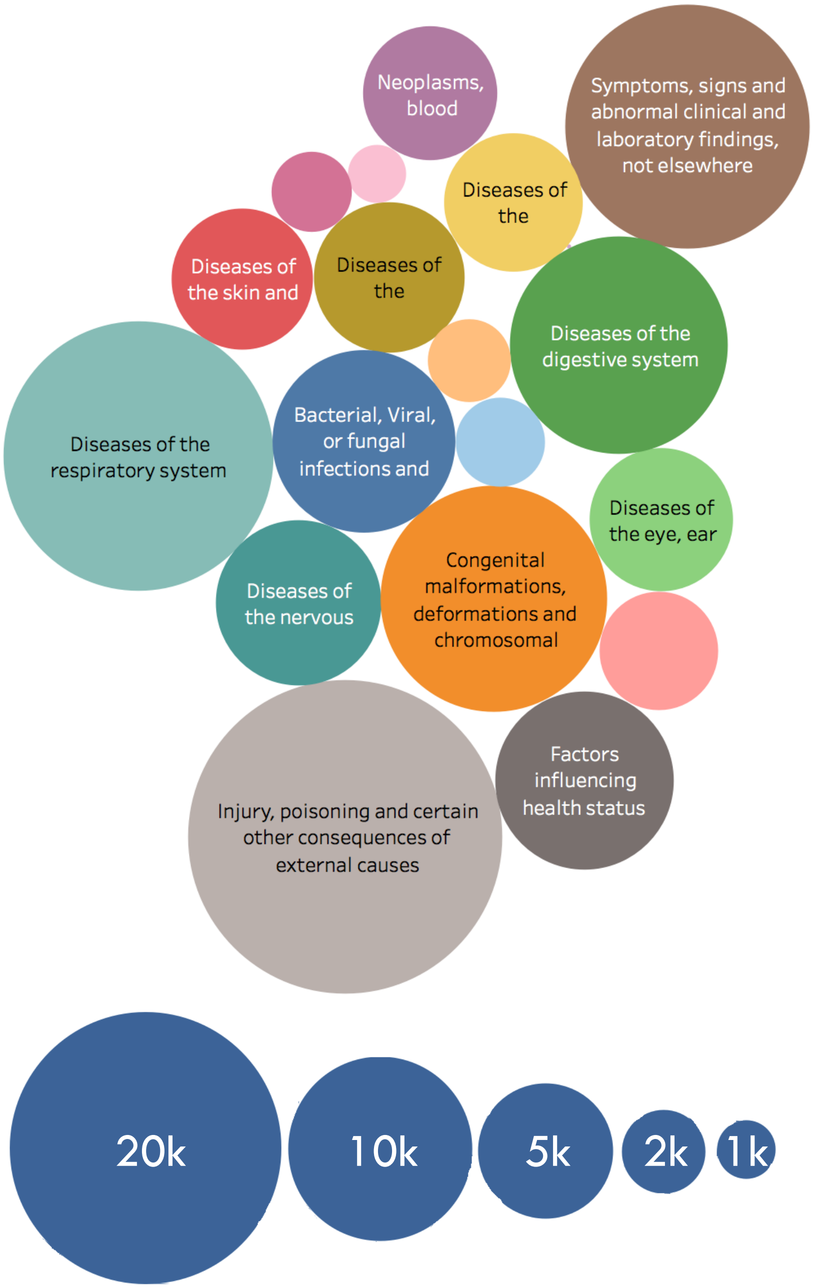
Australian Bureau of Statistics

ARIA+ index of remoteness

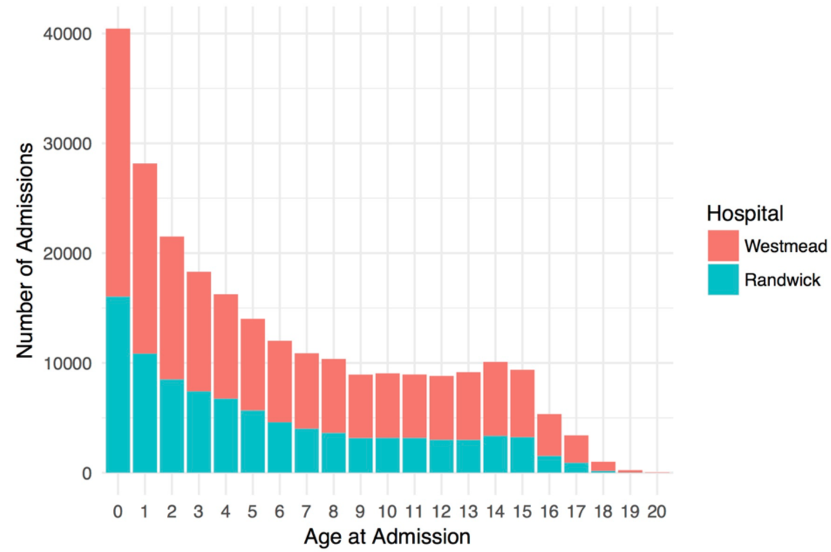
and SEIFA indicies

Factor / Covariate in data	Hypothesis (effect on DAMA rate)	Result?
Gender	No effect	
Index of Relative Socioeconomic Disadvantage	Higher DAMA rate in more disadvantaged areas	
Hospital	No effect	
Aboriginality	Increased DAMA rate for aboriginal children	
Emergency admission	Higher for Emergency admission	
Preventable illnesses	?	
Previous DAMA	Higher for previous DAMA	
Age group	Increasing rate of DAMA with age	
Rurality	Increasing rate of DAMA with remoteness	
ICD-10 diagnosis code	Larger effect for some diagnosis subgroups	

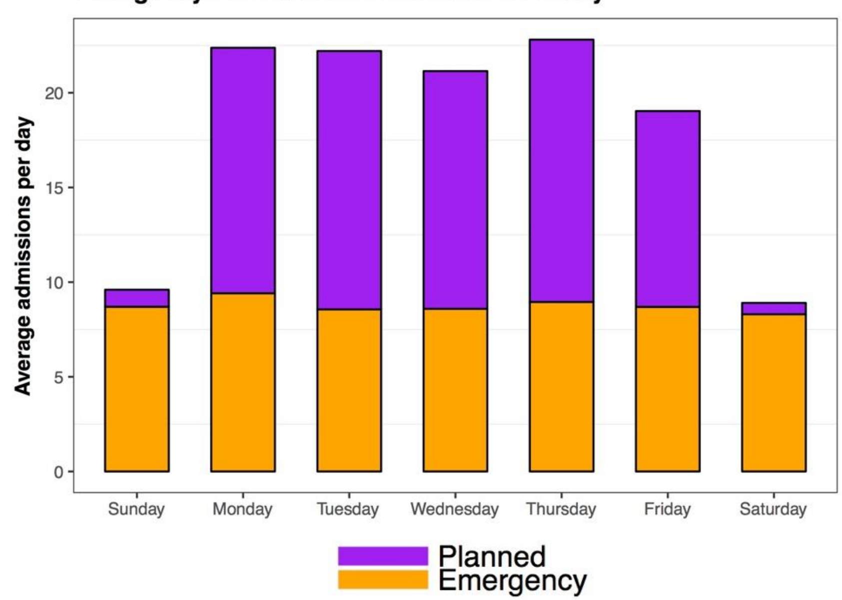
ICD10 Category of Admission



Age distribution of Admissions

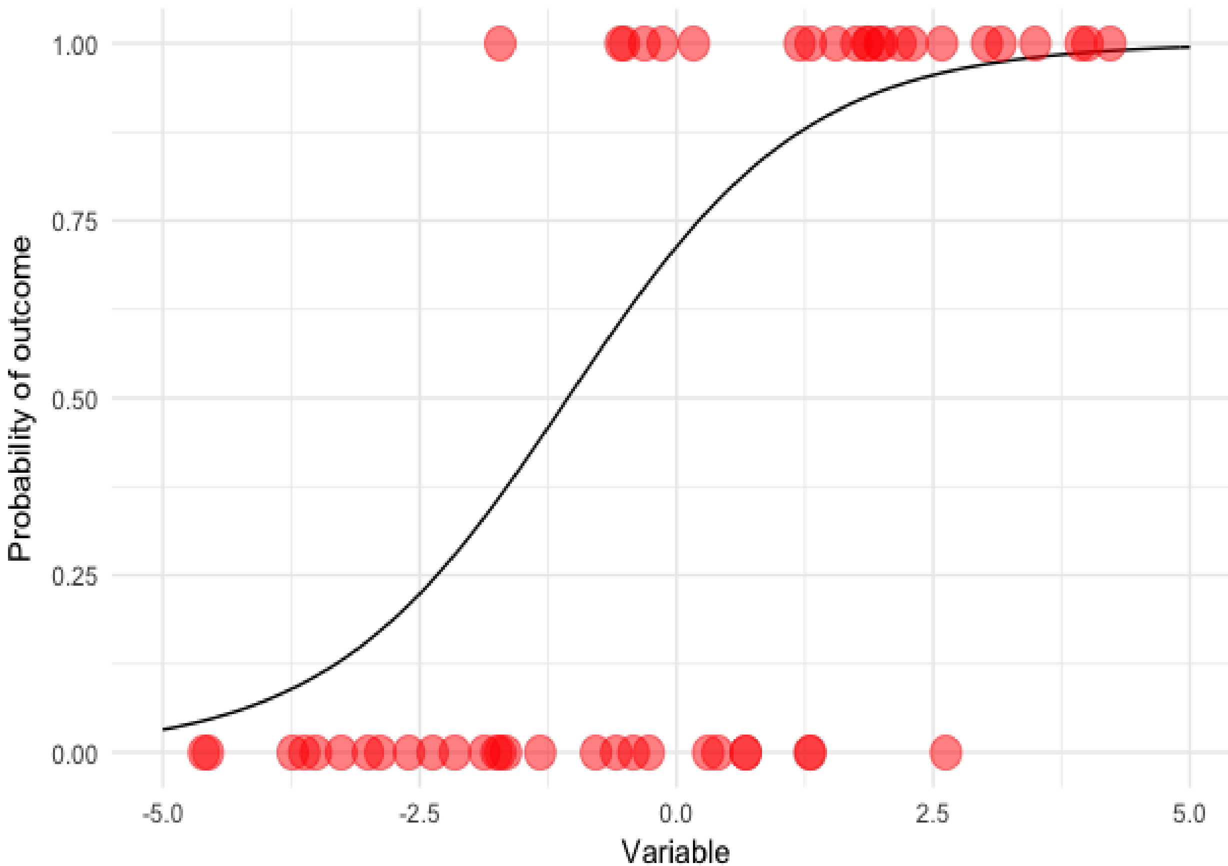


Emergency/Planned Admissions vs. Weekday



What is our model for DAMA?

Logistic regression



p

y

Patients with given outcome

Patients without given outcome

Income, Age, IRSD

Why use Bayesian methods?

1. Proper quantification of uncertainty from multiple sources

- Domain Knowledge
- Previous studies
- Low Sample size

2. Answers the question you really want answered:

What is the probability my model is accurate, given my experimental data?

rather than

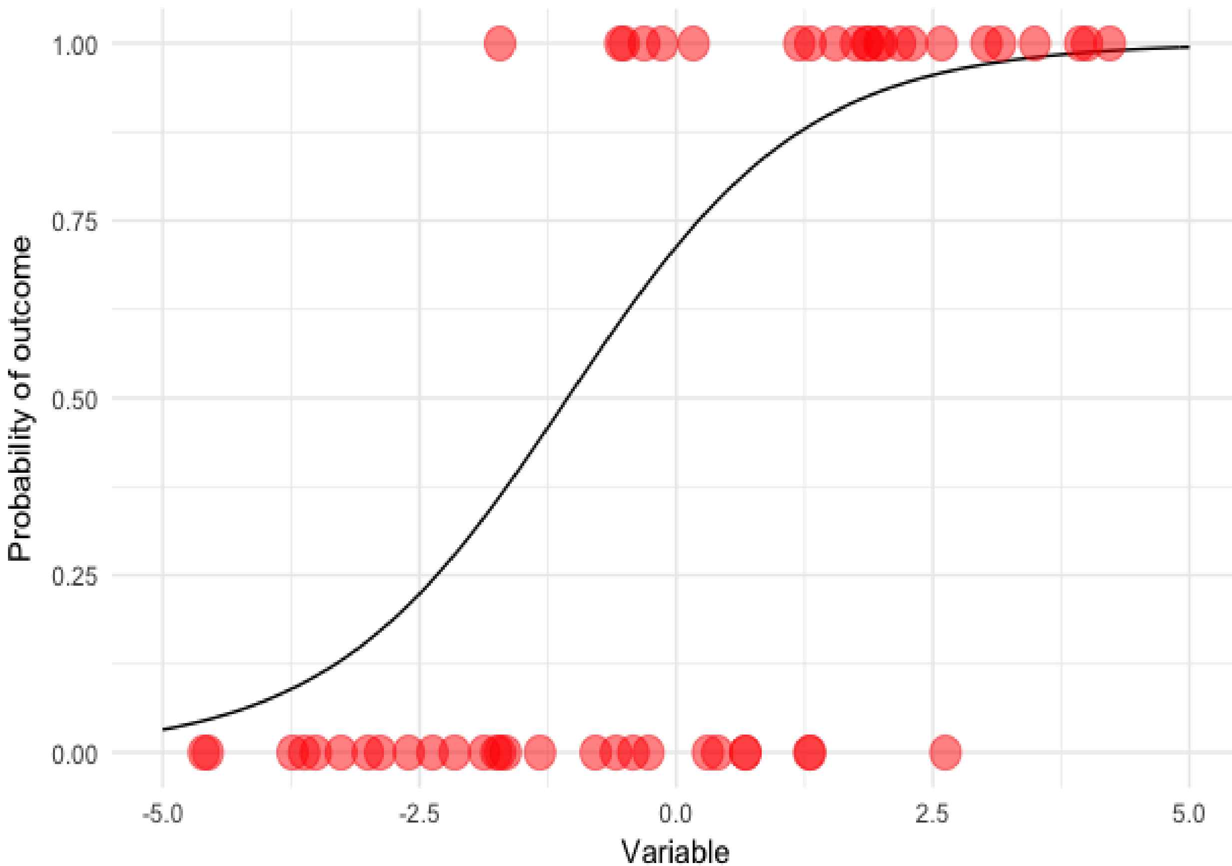
How likely is my data to have been generated by my (possibly incorrect) model?

3. Confidence intervals and *p*-values are much easier to interpret and understand in a Bayesian framework

4. More varied types of models can be explored and compared easily

Logistic regression

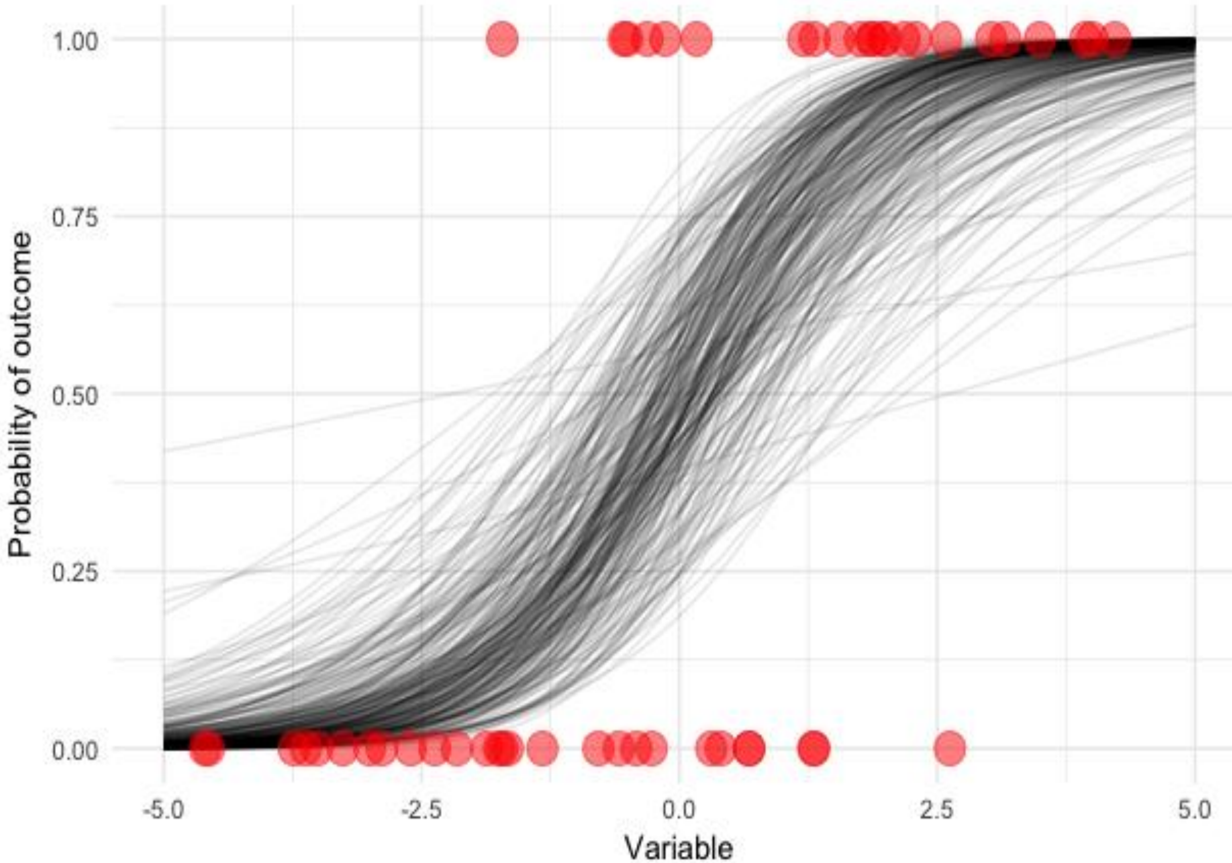
p



y

$$\frac{1}{1+e^{-y}}$$

Bayesian Logistic regression



p

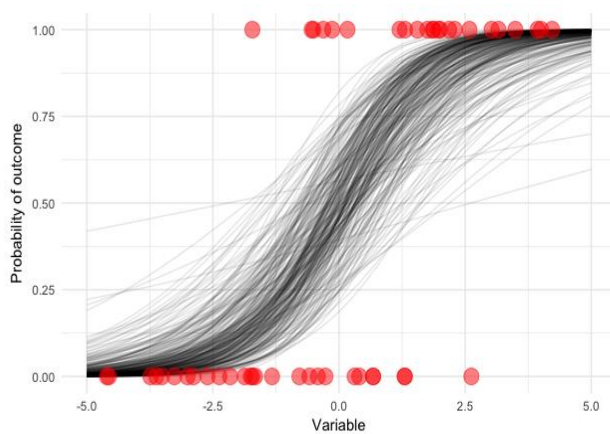
y

$$\frac{1}{1+e^{-y}}$$

Bayesian Logistic regression

$$y = \text{base rate } + \beta_1 \times \text{age}$$
$$+\beta_2 \times \text{IRSD } +\beta_3 \times \text{gender}$$
$$+\beta_4 \times \text{which hospital}$$
$$+\beta_5 \times \ldots.$$

p



y

$$p = \frac{1}{1 + e^{-y}}$$

Bayesian Logistic regression

$$y = \ln\left(\frac{p}{1-p}\right)$$

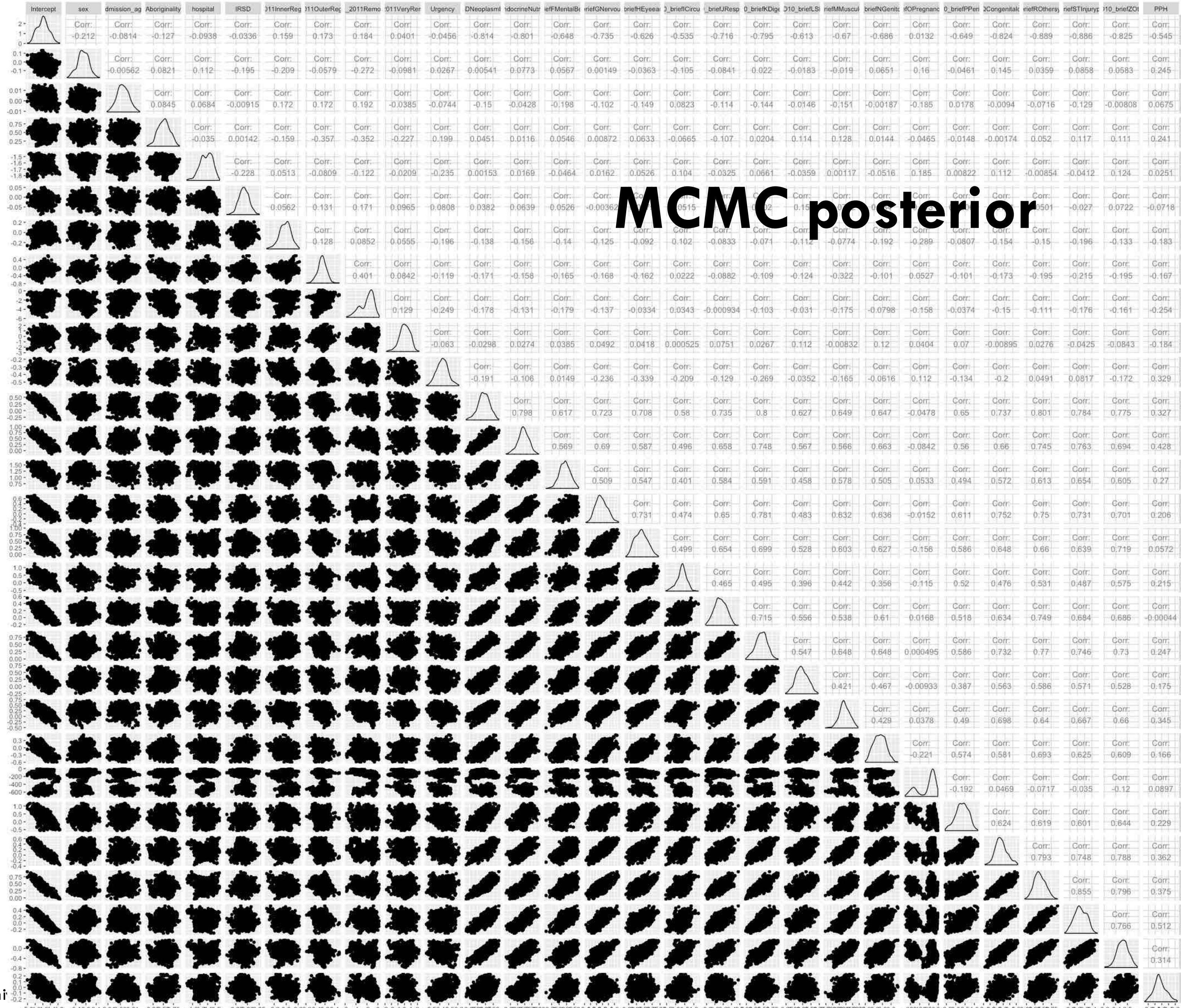
$$= \ln(\text{Odds}) \qquad \text{Odds} = \frac{p}{1-p}$$

$$\text{Odds Ratio}_{(\text{this case})} = \frac{\text{Odds}_{(\text{this case})}}{\text{Odds}_{(\text{reference})}}$$

Q: What factors influence DAMA?

1. Bayesian MCMC using Polya-Gamma latent variable technique, to find the posterior probability of each covariate in the logistic regression, using the R package 'BayesLogit'

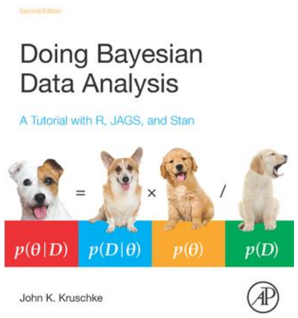
MCMC posterior



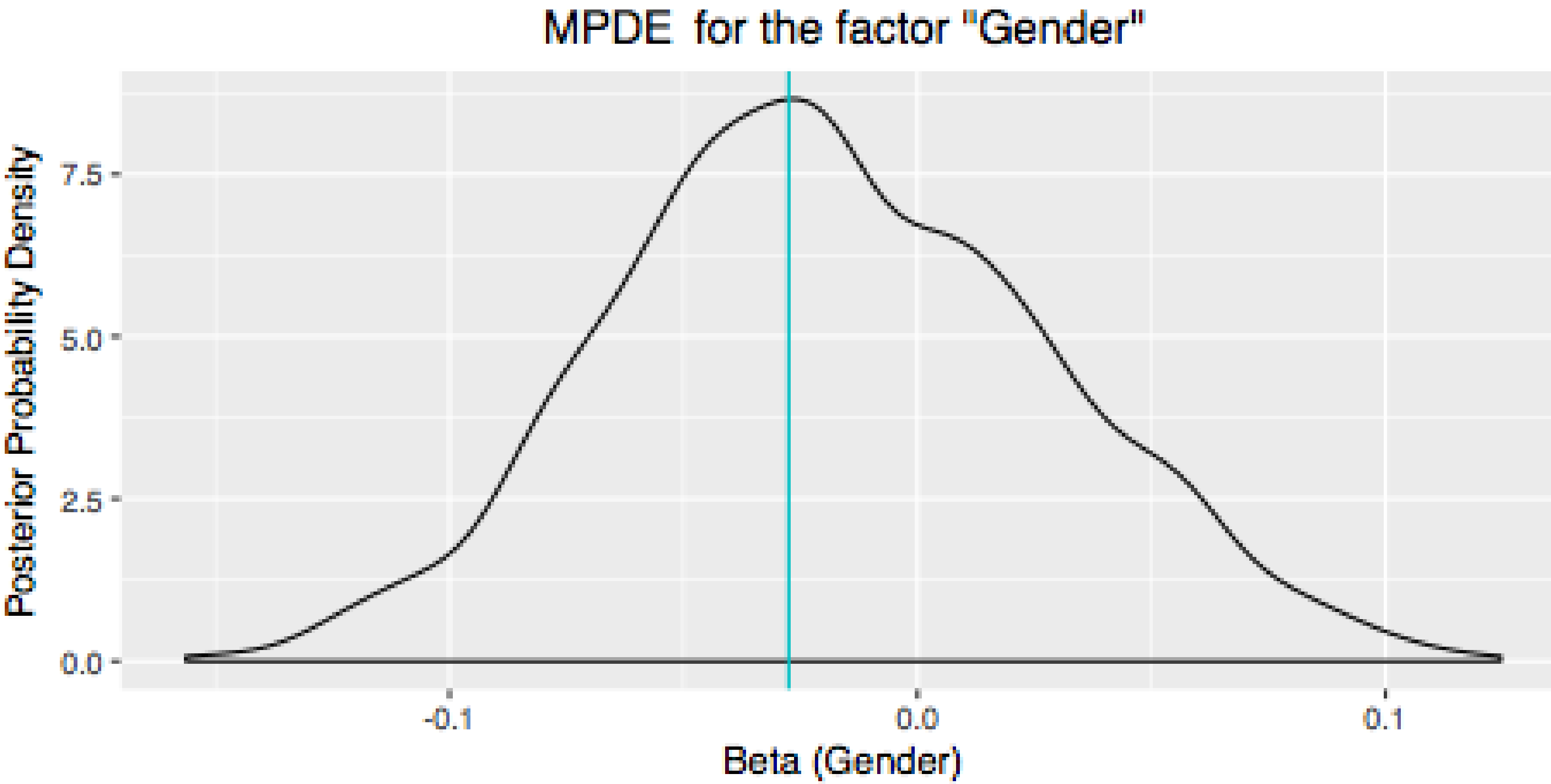
Q: What factors influence DAMA?

1. Bayesian MCMC using Polya-Gamma latent variable technique, to find the posterior probability of each covariate in the logistic regression, using the R package 'BayesLogit'

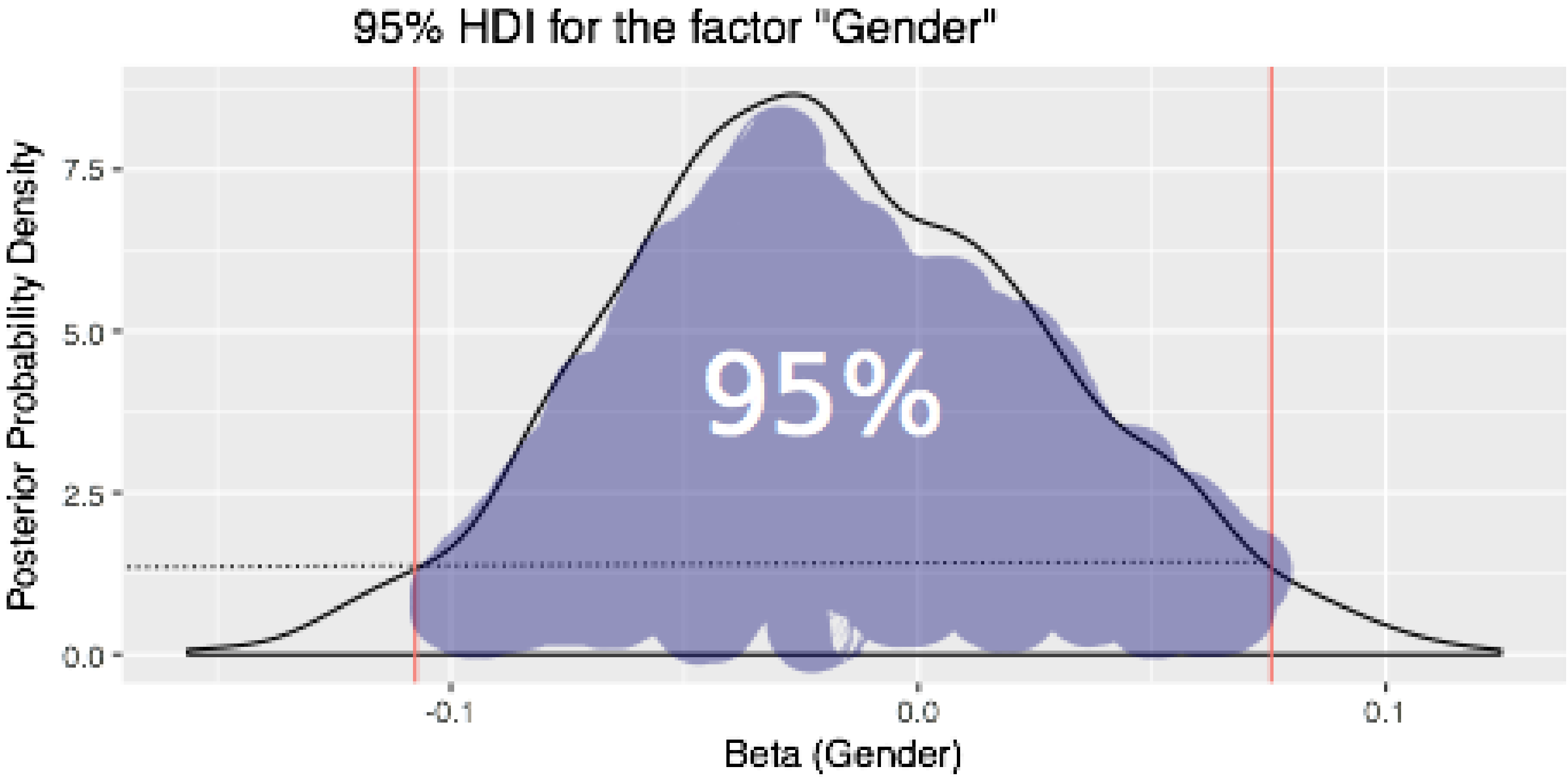
2. Estimate of each coefficient and Determination of significance using MPDE, HDI and ROPE



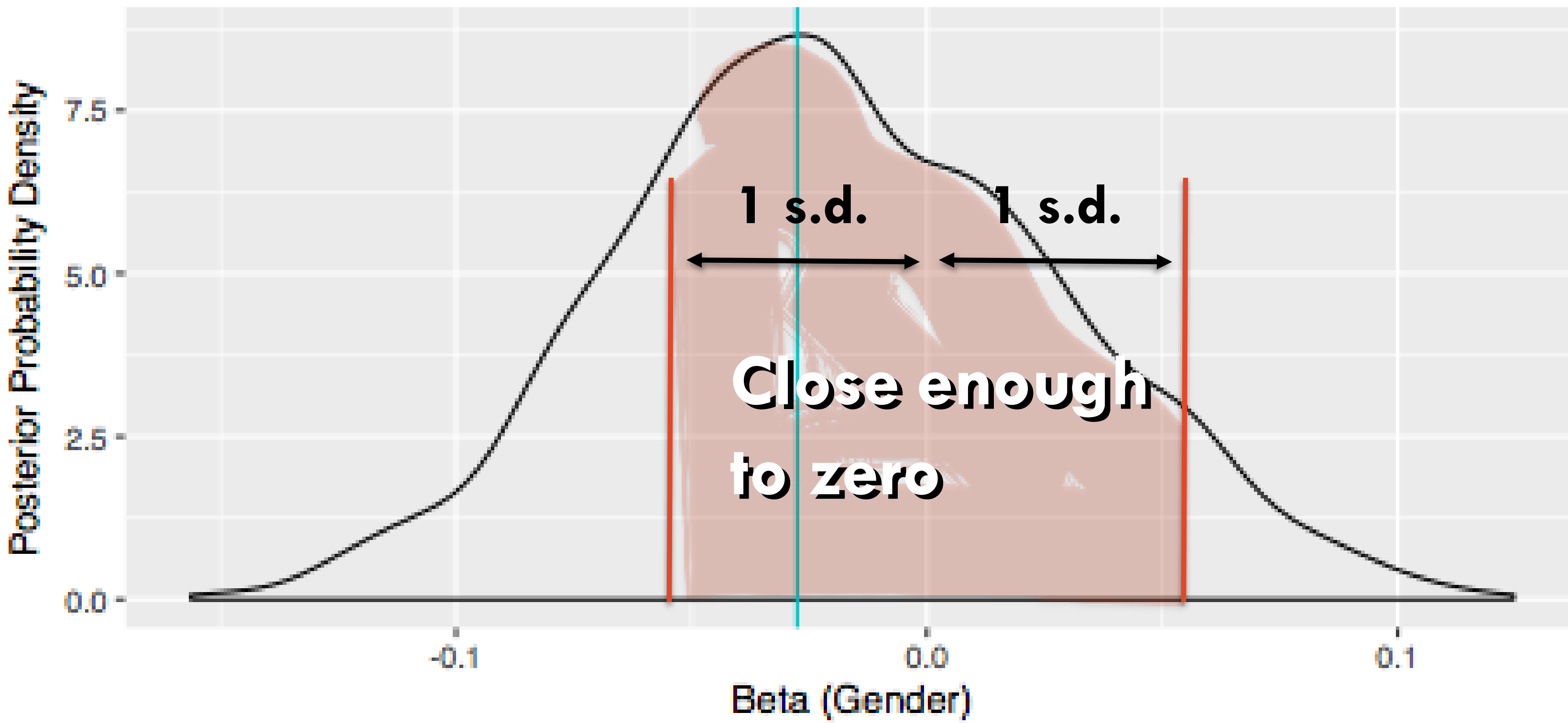
Maximum Posterior Density Estimate
"Best Guess"



95% Highest Density Interval



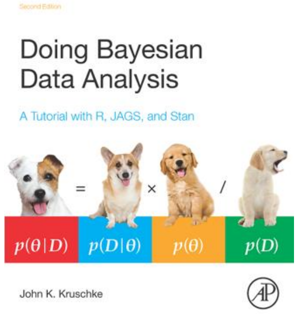
Define a region of practical equivalence (ROPE) to zero effect



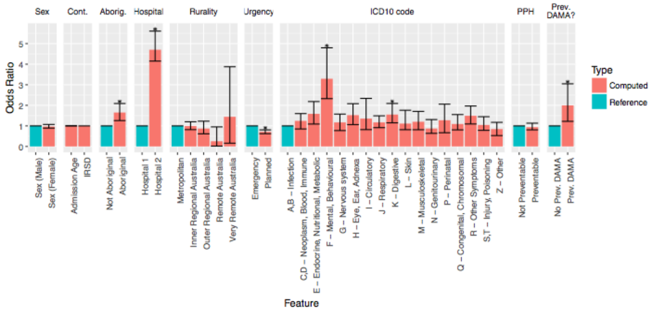
Q: What factors influence DAMA?

1. Bayesian MCMC using Polya-Gamma latent variable technique, to find the posterior probability of each covariate in the logistic regression, using the R package ‘BayesLogit’

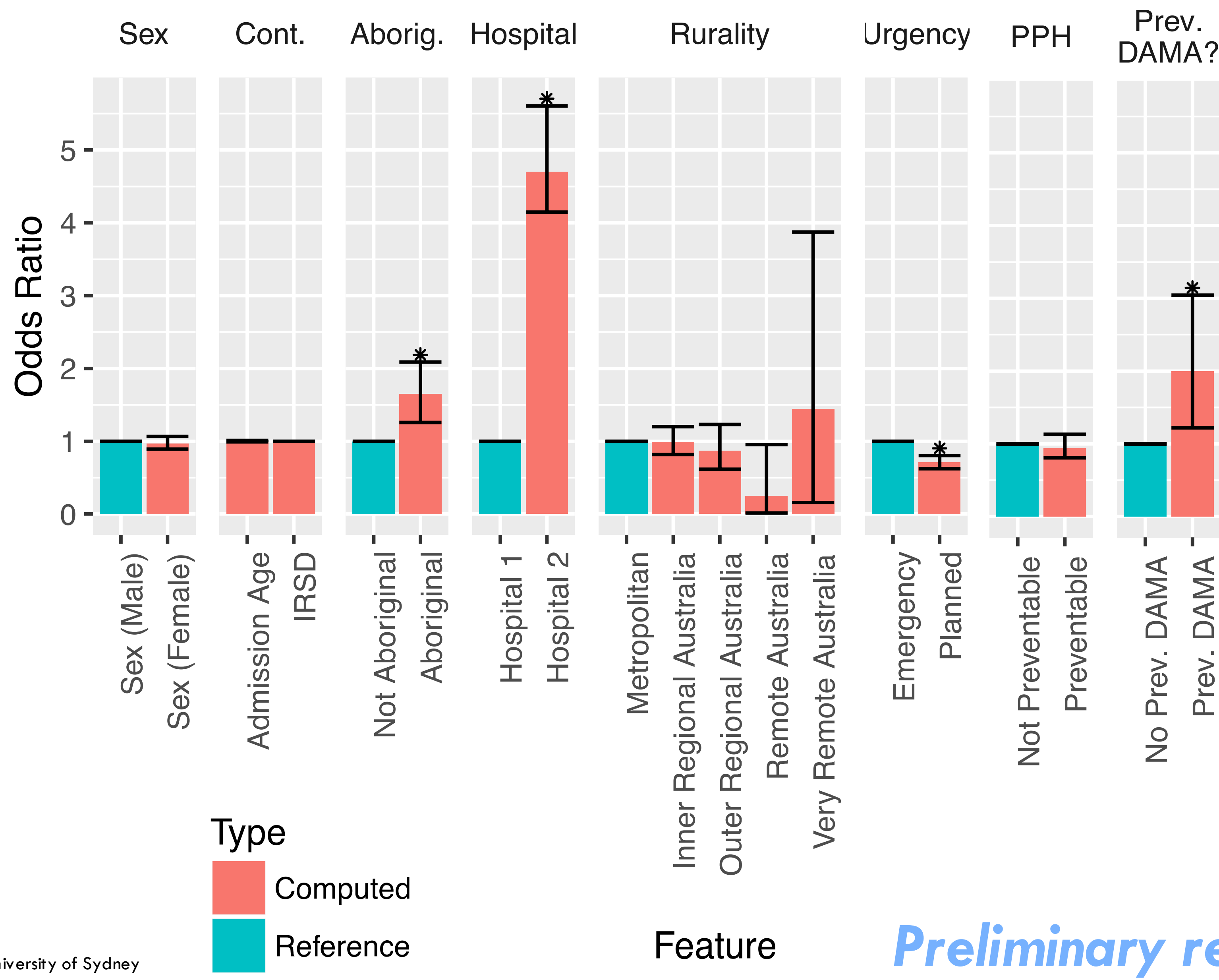
2. Estimate of each coefficient and Determination of significance using MPDE, HDI and ROPE (following Kruschke)



3. Plots of the result

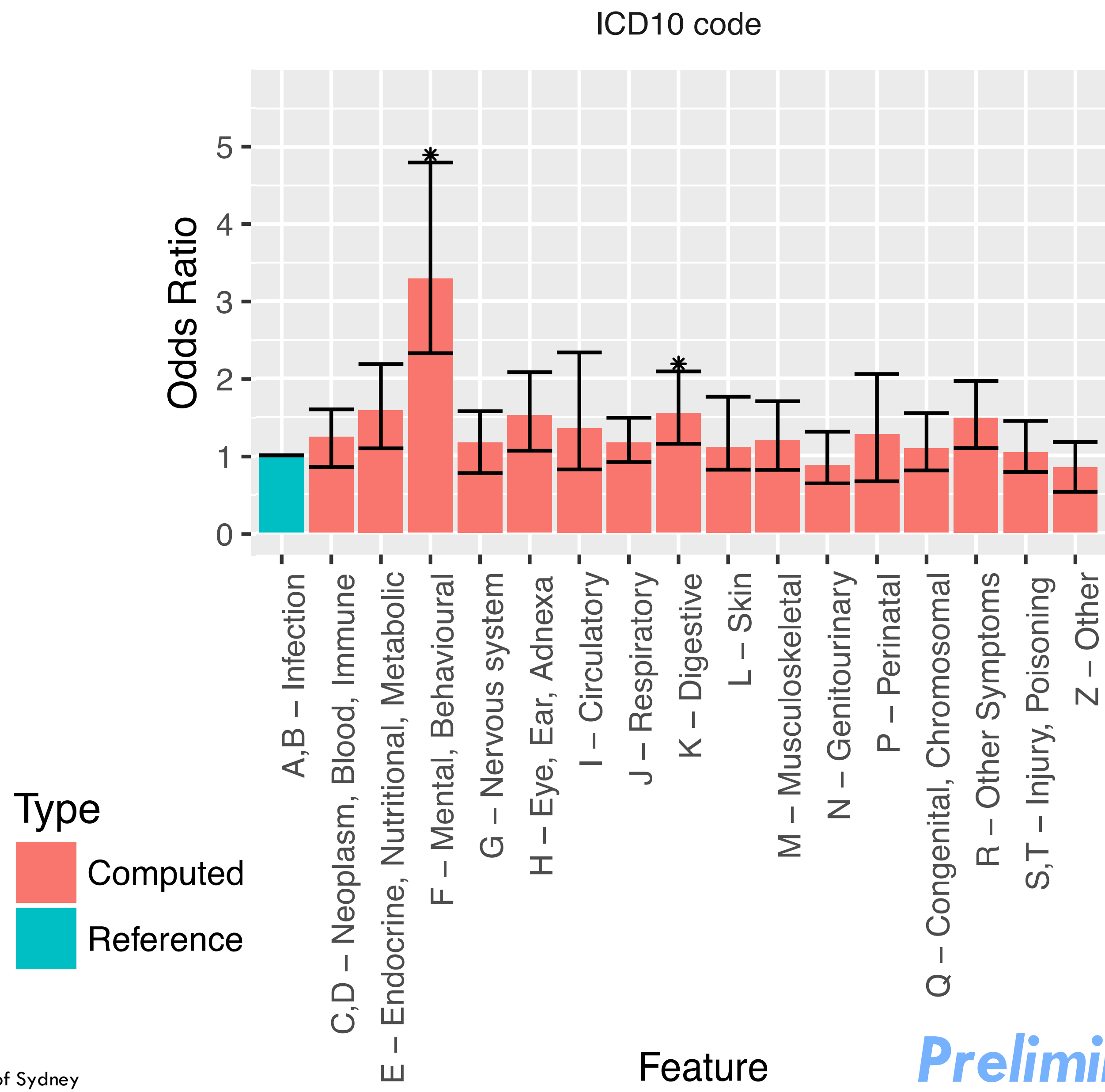


Odds Ratio of Discharge Against Medical Advice



Preliminary results

Odds Ratio of Discharge Against Medical Advice



Preliminary results

Factor / Covariate in data	Hypothesis (effect on DAMA rate)	Result?
Gender	No effect	No effect
Index of Relative Socioeconomic Disadvantage	Higher DAMA rate in more disadvantaged areas	No effect
Hospital	No effect	OR = 4.7 (4.2-5.6)
Aboriginality	Increased DAMA rate for aboriginal children	OR = 1.7 (1.3-2.1)
Emergency admission	Higher for Emergency admission	OR = 0.7 (0.6-0.8)
Preventable illnesses	?	No effect
Previous DAMA	Higher for previous DAMA	OR = 2 (1.2 – 3.0)
Age group	Increasing rate of DAMA with age	No effect
Rurality	Increasing rate of DAMA with remoteness	No effect
ICD-10 diagnosis code	Larger effect for some diagnosis subgroups	Large effect for Mental health: OR = 3.3 (2.3 – 4.8)

Preliminary results

Take away messages

1. Doing the analysis in a Bayesian framework allows proper quantification of uncertainty and makes interpretation of results easier.

2. Further study needs to be done to analyze the link between mental health and DAMA, and the link between CALD and DAMA, in order to understand and prevent discharge against medical advice.

Using Bayesian Methods to Predict Self-Discharge in the Sydney Children's Hospital Network

Presented by
Dr. Gordon McDonald
Sydney Informatics Hub
The University of Sydney

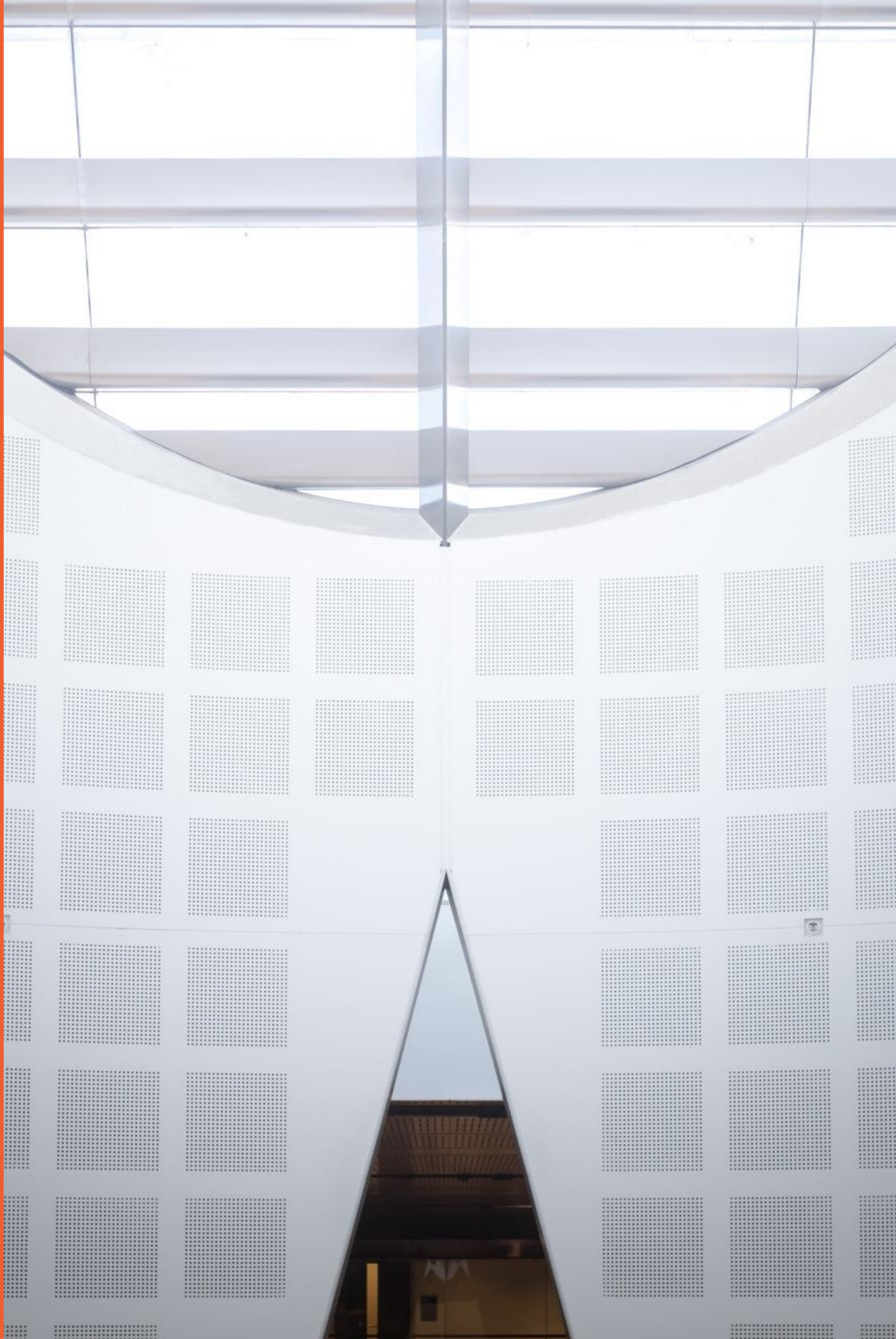
UNSW SYDNEY

On behalf of

Dr. Louise Sealy
Prof. Karen Zwi
Lisa Crawford
Prof. Hasantha Gunasekera
Dr. Gordon McDonald
Dr. Aldo Saavedra

The Sydney children's Hospitals Network
care, advocacy, research, education

THE UNIVERSITY OF SYDNEY



Appendix - why not p-values?

What is a *p*-value?

$$\Pr(D|H_1)$$

The probability of getting data as extreme (or more extreme) than you got,

given the assumption that your hypothesis (or your model) H_1 is true

What do we really want?

$$\Pr(H_1|D)$$

The probability that your hypothesis (or your model) H_1 is true

Given that you got the data that you got

$$\Pr(H_1|D) = \frac{\Pr(D|H_1)\Pr(H_1)}{\Pr(D)}$$

$$\Pr(H_1|D) = \frac{\Pr(D|H_1)\Pr(H_1)}{\Pr(D)}$$

Likelihood of the data given the hypothesis

$$\Pr(H_1|D) = \frac{\Pr(D|H_1)\Pr(H_1)}{\Pr(D)}$$

Prior belief about the hypothesis

$$\Pr(H_1|D) = \frac{\Pr(D|H_1)\Pr(H_1)}{\Pr(D)}$$

Posterior probability of the hypothesis given the data

$$\Pr(H_1|D) = \frac{\Pr(D|H_1)\Pr(H_1)}{\Pr(D)}$$

$$\Pr(H_1|D) = \frac{\Pr(D|H_1)\Pr(H_1)}{\Pr(D)}$$

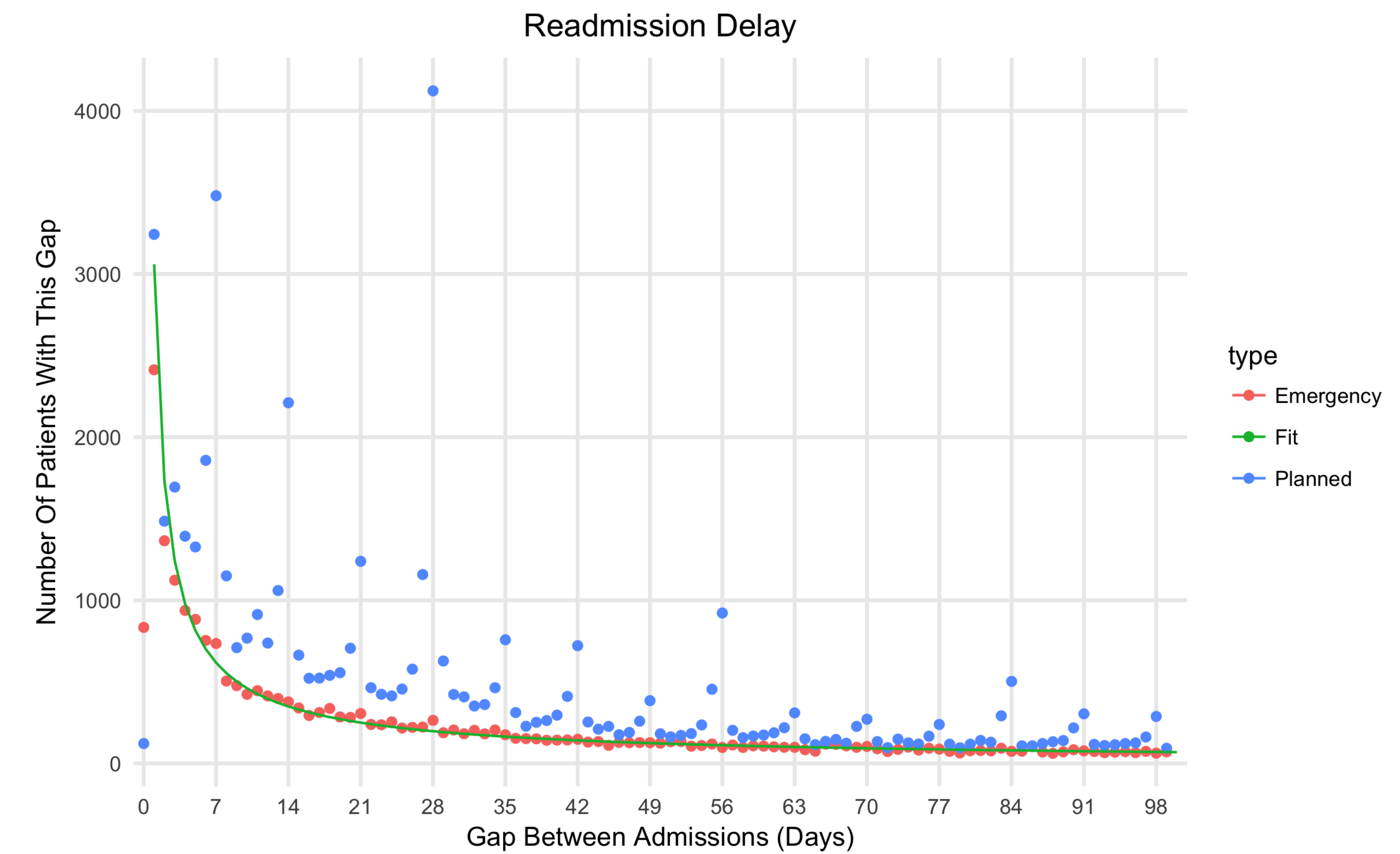
Chance of this evidence occurring

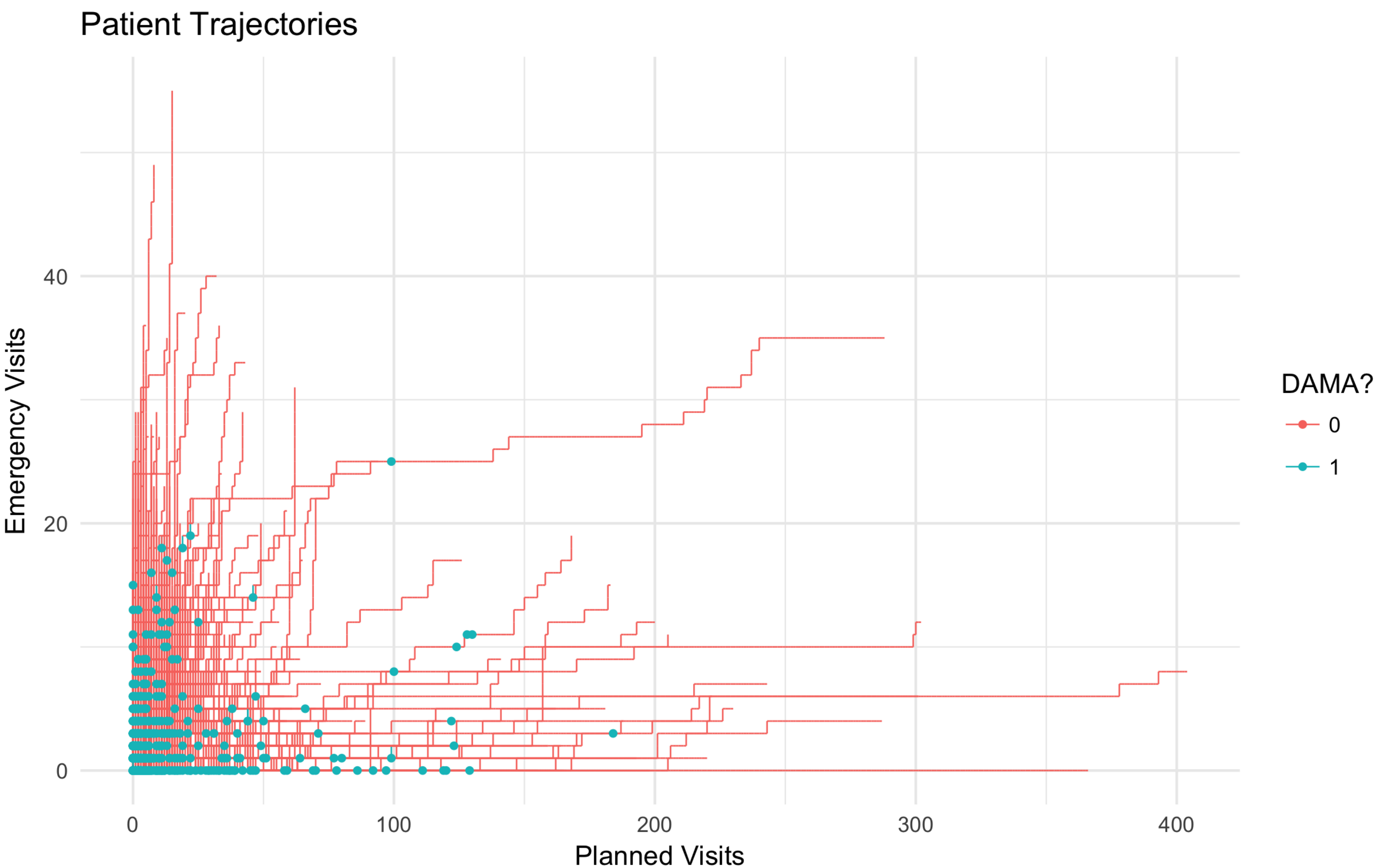
$$\Pr(H_1|D) = \frac{\Pr(D|H_1)\Pr(H_1)}{\Pr(D)}$$

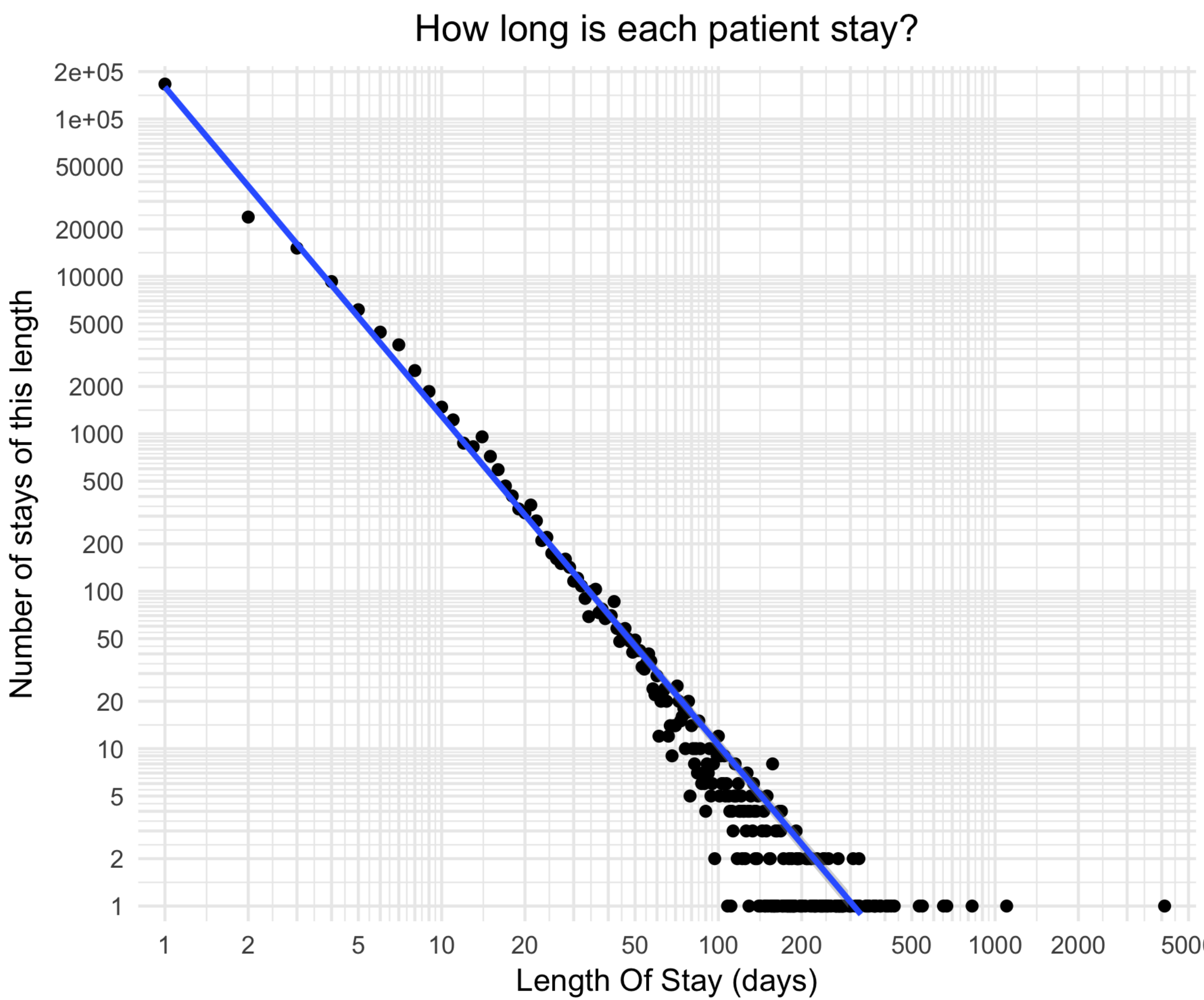
$$\begin{aligned}\Pr(D) = \quad & \Pr(D|H_1)\Pr(H_1) \\ + \; & \Pr(D|H_2)\Pr(H_2) \\ + \; & \Pr(D|H_3)\Pr(H_3) \\ + \; & \ldots\end{aligned}$$

But this is difficult to compute directly

So in the end we approximate by calculating thousands of different options and computing to see how they compare relative to one another.


Readmission Delay
Number Of Patients With This Gap
4000
3000
2000
1000
0
0
7
14
21
28
35
42
49
56
63
70
77
84
91
98
Gap Between Admissions (Days)
type
Emergency
Fit
Planned


Patient Trajectories
Emergency Visits
40
20
0
0
100
200
300
400
Planned Visits
DAMA?
0
1


How long is each patient stay?
Number of stays of this length
2e+05
1e+05
50000
20000
10000
5000
2000
1000
500
200
100
50
20
10
5
2
1
1
2
5
10
20
50
100
200
500
1000
2000
5000
Length Of Stay (days)