Modelling Healthcare Associated Infections:

A Bayesian approach.

Theo Kypraios & Phil O'Neill

School of Mathematical Sciences, University of Nottingham

Tubingen, October 2008

Joint work with:
Ben Cooper @ Health Protection Agency

Outline

Motivation

- High-profile hospital-acquired infections such as:
  - *Methicillin-Resistant Staphylococcus Aureus* (MRSA) and
  - *Glycopeptide-Resistant Enterococcal* (GRE)
  - *Vancomycin-Resistant Enterococcal* (VRE)

  have a major impact on healthcare within the UK and elsewhere. The annual economic costs in 2002 prices were:
  - \$6.7 billion per years in the United States.
  - \$1.7 billion per years in the United Kingdom.

- Despite enormous research attention, many basic questions concerning the spread of such pathogens remain unanswered.

- Our aim is to address a range of scientific questions via analyses of detailed data sets taken from observational studies on hospital wards.

Motivation (cont.)

For instance, we are interested in answering important questions such as:

- What value do **specific control measures** have?
- Is it of **material benefit** to **increase or decrease the frequency** of swab tests?
- How is **transmission** within a ward related with **"colonisation pressure"?**
- What enables some **strains** to **spread more rapidly** than others?
- What effects do **different antibiotics** play?

Methicillin-Resistant Staphylococcus Aureus

- Staph. Aureus is a bacterium.
- Usually resides in the front part of the nose.
- About 80% of population carry it at some time.
- Transmission primarily via hands.
- Most common cause of surgical infections.

Typical Data sets

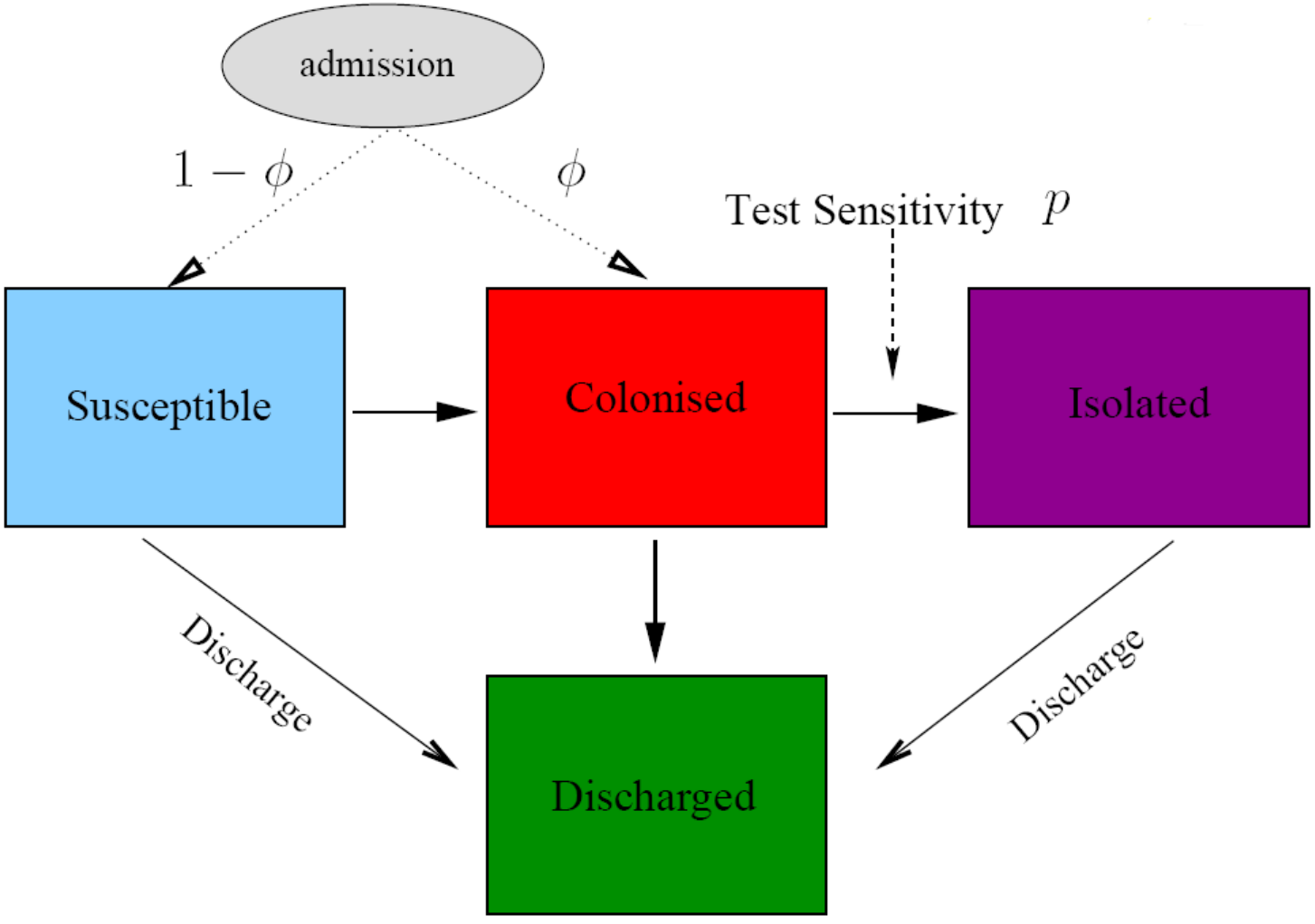
Typical data sets contain anonymised ward - level information on:

- Dates of patient admission and discharge.
- Dates when swab tests are taken (e.g. for MRSA, GRE).
- Outcomes of tests.
- Patient location (e.g. in isolation).
- Details of antibiotics administered to patients.
- Typing data.

Previous Work

- Typically, forward−simulation studies (for example, Sébille (1997), Austin et al. (1999), Lipsitch et al. (2000), Bootsma et. al. (2006)) are used to investigate the effectiveness of widely used infection control measures and rapid diagnostic testing.
- Pelupessy et al. (2002) proposed a Markov model and using maximum likelihood techniques estimated parameters by making assumption about the sensitivity of swab tests.
- Forrester et al. (2006) propose a stochastic epidemic model to infer transmission rates within a Bayesian framework.
- Bootsma et. al. (2007) propose an algorithm to estimate the importance of bacterial acquisition routes in hospital settings using maximum likelihood techniques.

A Schematic Representation of a "Standard Model"



Screening Tests

- Taken at specific times for every single patient
  - If positive then the patient becomes isolated.
- This routine swabbing procedure may be subject to imperfect sensitivity, i.e. some false negative swabs are possible.
  - Therefore, we assume that the sensitivity of this swabbing procedure is denoted by p.
  - 100% specificity is assumed, although this assumption can be relaxed.

Recall that:

- Sensitivity: $\mathbb{P}(\text{Test is positive}|\text{patient is colonised})$
- Specificity: $\mathbb{P}(\text{Test is negative}|\text{patient is uncolonised})$

Model Dynamics

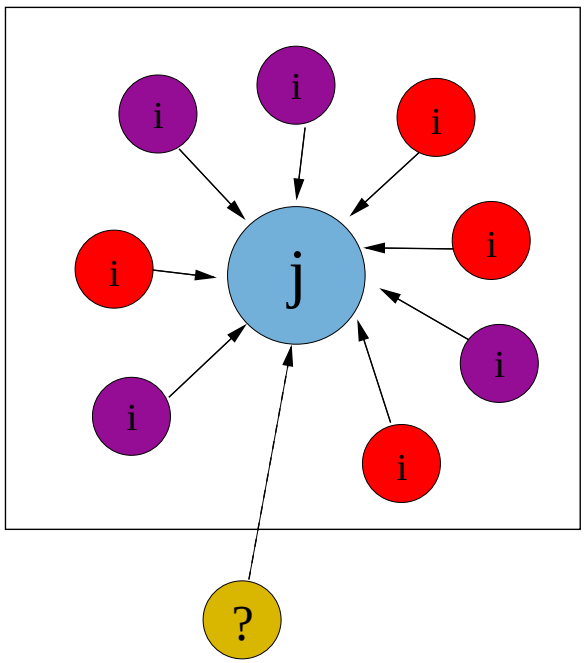
- While susceptible an individual receives indirect colonisation pressure from each colonised and non-isolated (colonised and isolated) according to a homogeneous Poisson process with intensity β_1 (β_2).
- We also allow for background transmission, i.e. an individual receives colonisation pressure from outside the ward according to homogeneous Poisson process with intensity β_0.

Model Dynamics (cont.)

In other words, the total pressure that susceptible individual j is subject to just prior to their colonisation is:

$$\lambda_j = \beta_0 + \beta_1 n_C + \beta_2 n_I$$

where n_C is number of colonised individuals on ward, n_I is number of isolated individuals on ward.



Note that this assumes **linear** colonisation pressure.

A Case Study in MRSA

The Data

Data on colonisation were collected from 9 adult intensive care units over a 17-month period.

- 10-bed ICUs in a tertiary academic medical center.
- Routine admission and weekly bilateral nares screening for MRSA (compliance 90%).
- Types of ICUs including:
  - medical,
  - cardiac,
  - general/cardiac/thoracic surgery,
  - burn trauma,
  - neurosurgery.
- Dates of MRSA-positive clinical cultures as well as positive and negative screening cultures were collected.

The Data (cont.)

- Newly-identified and previously known MRSA-positive patients were placed into contact precautions such as gown and glove use as well as use of single rooms.
- Dates of each ICU admission and discharge were obtained.
- Dates on which contact precautions were initially applied were also known.
- The first institutional date of MRSA-positive culture was also recorded even if it preceded the study period.

Summary Statistics

Ward	Number of patients	In contact precautions	Mean (SD) length of stay length of stay.
1	1293	147	3.4 (4.7)
2	706	88	5.8 (11.4)
3	1263	64	3.6 (5.2)
4	1097	66	3.8 (6.4)
5	888	67	4.8 (9.7)
6	212	50	4.8 (9.7)
7	806	144	4.3 (6.0)
8	1227	152	3.4 (5.2)
9	1030	110	4.0 (8.3)

Procedure

- We wish to make inference for the unknown parameters which govern transmission:
  - Colonisation rates: β_0, β_1, β_2.
  - Screening Test's Sensitivity: p.
  - Importation probability: ϕ.

- We fit the aforementioned "Standard Model" (assuming *linear colonisation* pressure) to the data from each ward by employing an MCMC algorithm.

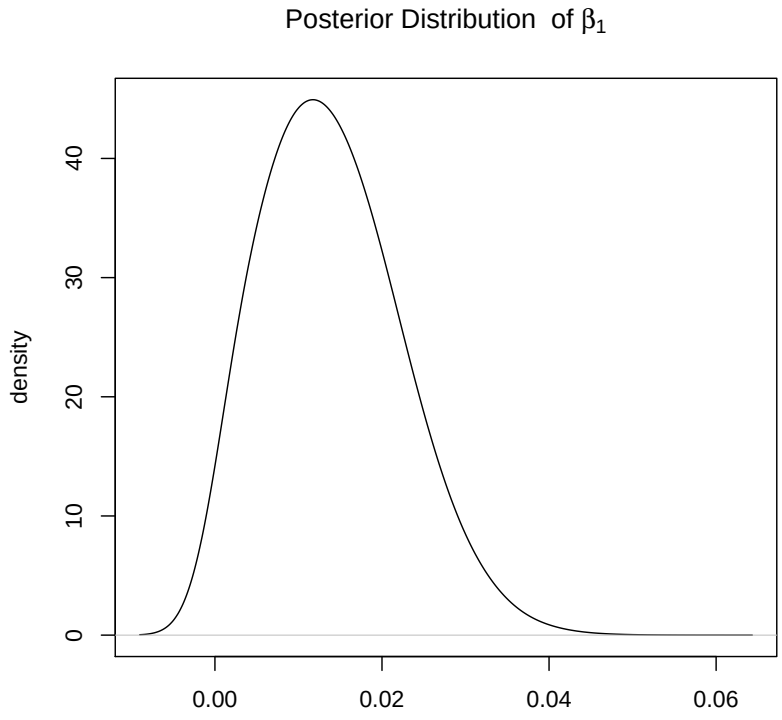
- Each MCMC algorithm runs *long enough* and then we end having samples from the posterior distribution of the parameters of interest $(\beta_0, \beta_1, \beta_2, p, \phi)$ given the observed data.

- Fairly uninformative priors were used – typically Exponential distributions with very low rate.

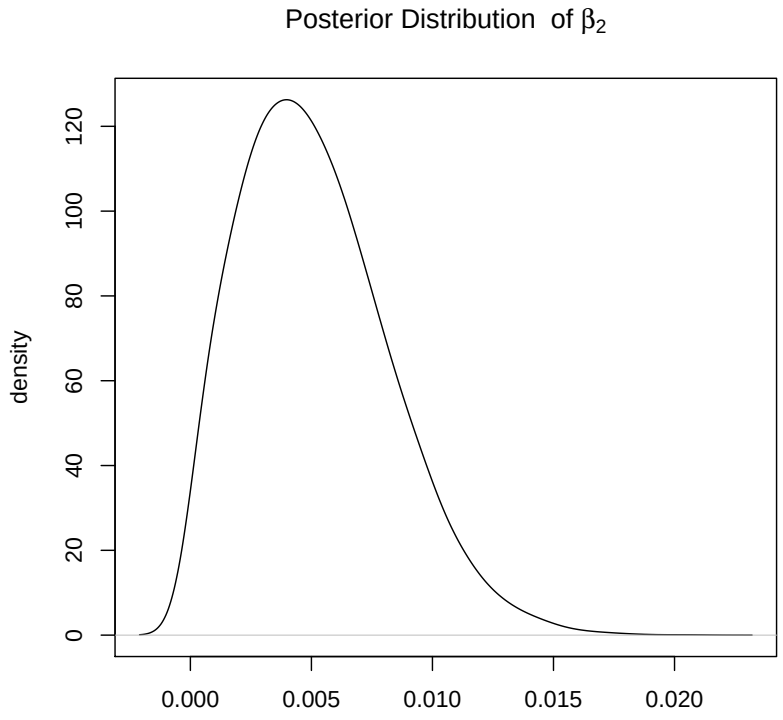
Results Within a Specific Ward

For illustration, we focus on the results obtained from the data analysis in Ward 5.

First, we concentrate on the colonisation rates β_1 and β_2:



(a) β_1

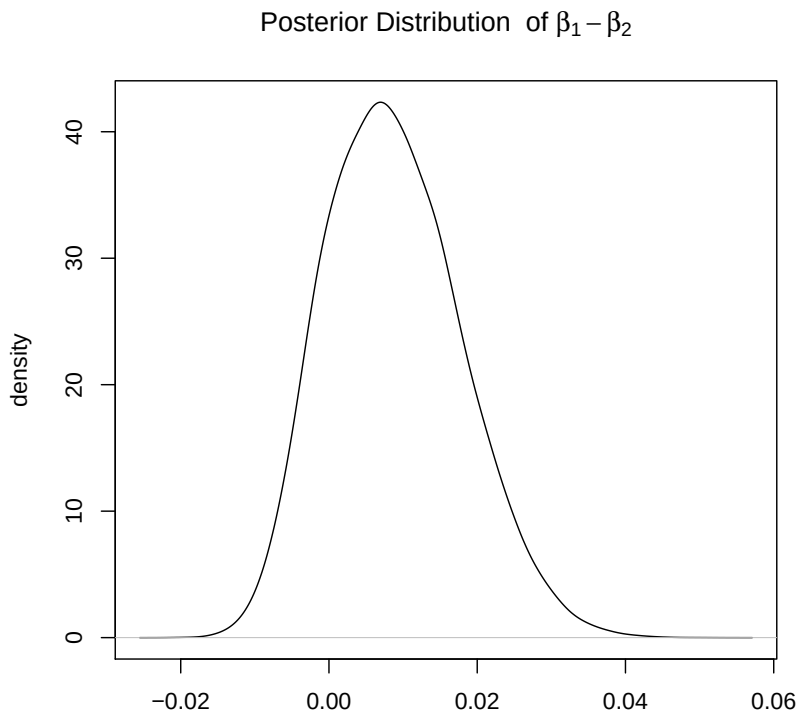


(b) β_2

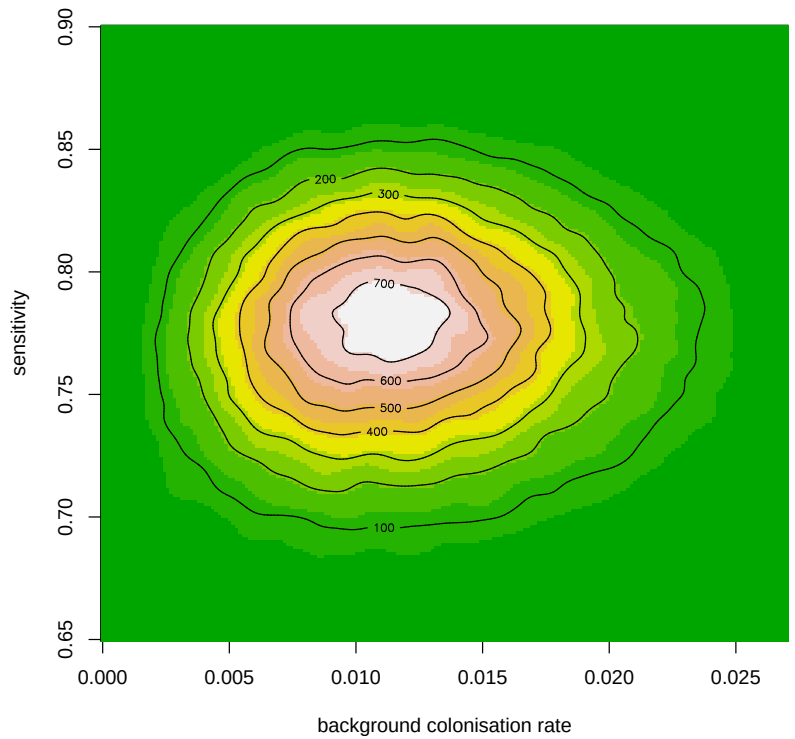
Results Within a Specific Ward

Apart from focusing on the posterior distribution of each of the model's parameters we can also look at a:

- joint distribution or a
- function of them.



(a) $\beta_1 - \beta_2$



(b) (β_0, p)

Are contact precautions actually effective?

Are Contact Precautions Effective?

Some ways to measure this include calculating

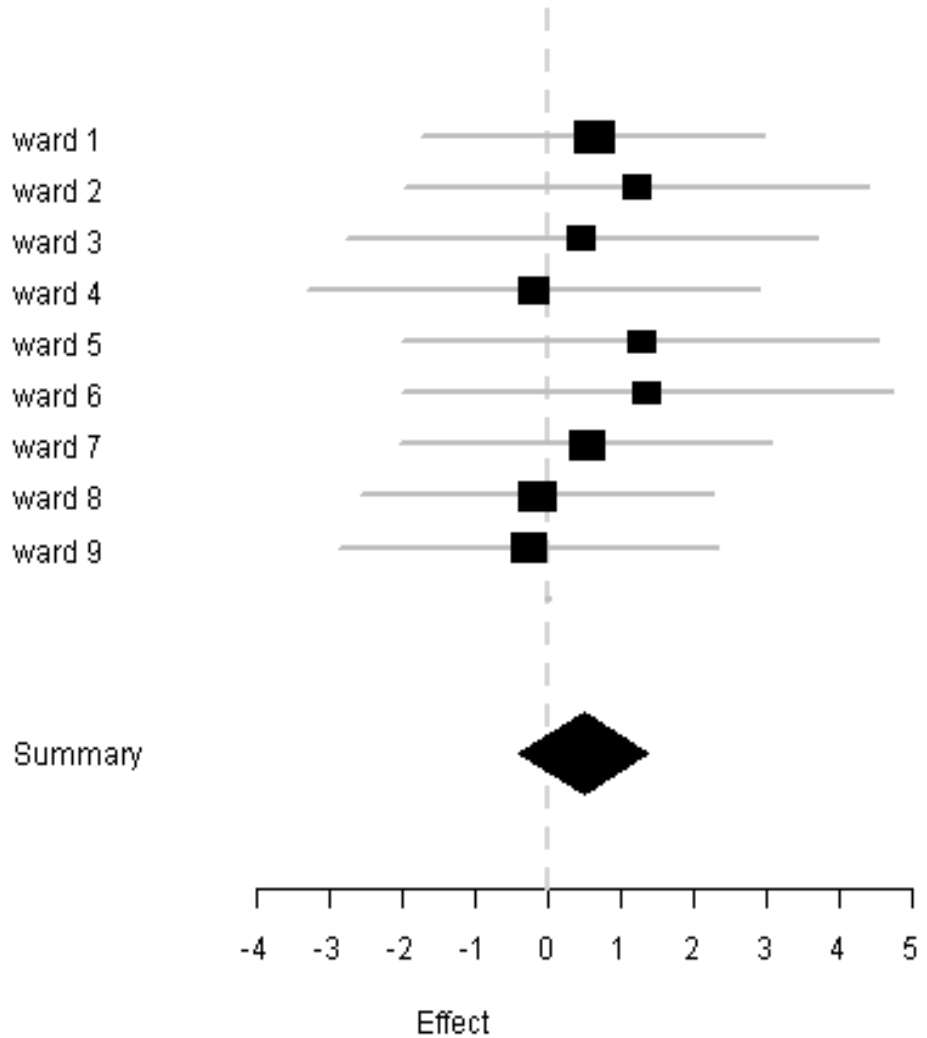
- $\mathbb{P}(\beta_1 > \beta_2|\mathbf{y})$, or
- considering the ratio β_1/β_2.

Ward	$\mathbb{P}(\beta_1 > \beta_2\|\mathbf{y})$	Median(β_1/β_2)
1	0.75	1.8
2	0.80	3.2
3	0.63	1.6
4	0.44	0.8
5	0.82	3.6
6	0.81	3.8
7	0.70	1.7
8	0.41	0.8
9	0.40	0.8

where $\mathbf{y}$ denotes the observed data.

Summarising the Results

By borrowing techniques from "Meta-Analysis" we can derive a *pooled estimate* for the $\log(\beta_1/\beta_2)$:



Are the findings model-dependent?

- Contact precautions appear to be effective - but what happens if we used a different model?
- We instead consider a simpler model in which
  - the colonisation pressure received by a susceptible individual does not increase with the number of colonised individuals.
- Specifically, the total pressure that susceptible individual j is subject to just prior to their colonisation is:

$$\lambda_j = \beta_0 + \beta_1 \mathbb{1}_{\{n_C \geq 1\}} + \beta_2 \mathbb{1}_{\{n_I \geq 1\}},$$

where n_C is number of colonised individuals on ward, n_I is number of isolated individuals on ward.

Results with simpler model

Here is the previous table with results for simpler model in brackets:

Ward	$\mathbb{P}(\beta_1 > \beta_2 \mid \mathbf{y})$	Median(β_1/β_2)
1	0.75 (0.59)	1.8 (1.3)
2	0.80 (0.60)	3.2 (1.4)
3	0.63 (0.64)	1.6 (1.7)
4	0.44 (0.65)	0.8 (1.9)
5	0.82 (0.73)	3.6 (2.3)
6	0.81 (0.67)	3.8 (1.9)
7	0.70 (0.45)	1.7 (0.9)
8	0.41 (0.39)	0.8 (0.7)
9	0.40 (0.63)	0.8 (1.5)

Results with different prior distribution

Another way of changing the model assumptions is via the choice of parameter prior distributions.

For example, consider now $\beta_0 \sim Exp(10^4)$ a priori, meaning that the background transmission rate is effectively thought to be zero.

Table: $\mathbb{P}(\beta_1 > \beta_2|\mathbf{y})$ for Full Transmission Model

Ward	Uninformative Prior	Informative Prior
1	0.75	0.59
2	0.80	0.65
3	0.63	0.80
4	0.44	0.55
5	0.82	0.75
6	0.81	0.80
7	0.70	0.74
8	0.41	0.42
9	0.40	0.40

How transmission within the ward is related to "colonisation pressure" ?

Colonisation Pressure

- We would like to investigate how does the risk of acquisition of MRSA for an individual vary depending on the number of colonised patients within the ward? We define

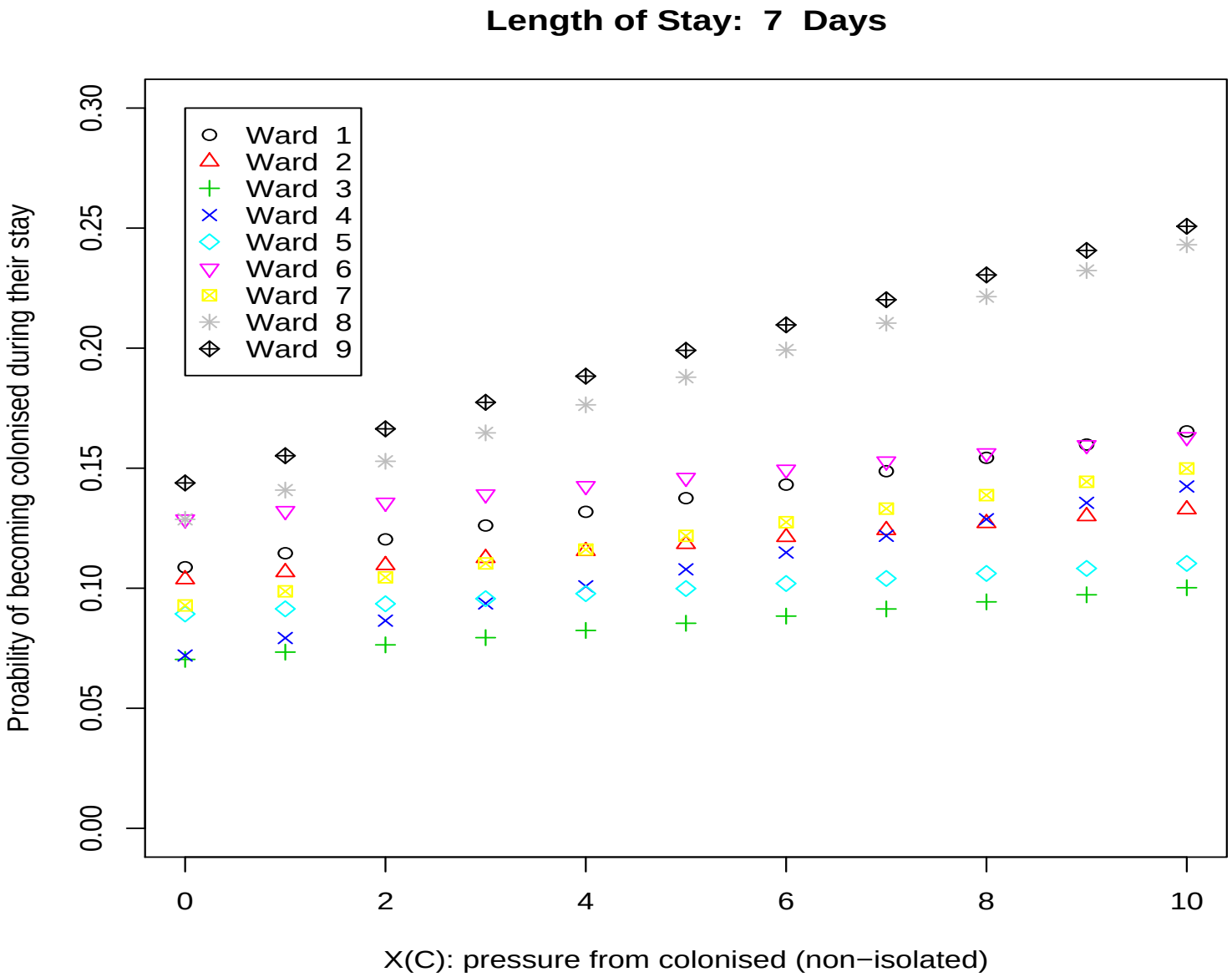
$$X_C = \int_a^d n_C(t)\, dt \quad X_I = \int_a^d n_I(t)\, dt$$

  so that X_C and X_I represent the colonisation pressure from colonised individuals who are non-isolated and isolated, respectively.

- For a patient on ward j who is exposed to X_C and X_I units of pressure, the probability that they will be colonised during their stay is

$$1 - \exp\left\{-(\beta_0^j(d-a) + \beta_1^j X_C + \beta_2^j X_I)\right\}.$$

Colonisation Pressure



Colonisation Pressure

Is there evidence to support the assumption of linear colonisation pressure?

- Bayesian Model Choice.
- Posterior Model Probabilities - Bayes Factors
- Within-Model prior distributions and Lindley's paradox : Prior's Matching & Prior Senstivity
- Trans-dimensional MCMC algorithm

Matched Prior Distributions

Both models have three parameters:

$$\beta_0, \beta_1 \text{ and } \beta_2.$$

We match the prior distributions for the two models by trying to make the pressure experienced by a susceptible individual similar in both models.

For instance, the pressure from the colonised (but non-isolated) individuals under the standard and the simpler model are:

$$\begin{aligned} \text{Standard Model:} \quad & \lambda_F = \beta_1^F \cdot n_C \\ \text{Simpler Model:} \quad & \lambda_S = \beta_1^S \cdot \mathbb{1}(n_C > 0) \end{aligned}$$

Matched Prior Distributions (cont.)

The idea is to assign a prior distribution to β_1^F say, and then derive the prior distribution for β_1^S by matching the moments of the prior distributions:

$$E[\lambda^F] = E[\lambda^S]$$
$$V[\lambda^F] = V[\lambda^S]$$

For example, if we assign $\beta_1^F \sim Exp(\mu)$ then it is easy to derive $E[\lambda^F]$ and $V[\lambda^F]$ for some fixed value of n_C

Assuming a Gamma prior for β_1^S, i.e. $\beta_1^S \sim Ga(c, d)$ we can derive moment estimators for c and d as follows:

$$c = \frac{\left(E[\lambda^F]\right)^2}{V[\lambda^F]} \text{ and } d = \frac{E[\lambda^F]}{V[\lambda^F]}$$

RJMCMC results

Ward	$\mathbb{P}$(Simpler Model|data)
1	0.37
2	0.72
3	0.75
4	0.71
5	0.89
6	0.85
7	0.90
8	0.23
9	0.21

- Results do not suggest much support for the full model.
- However, closer scrutiny reveals that, typically, n_C and n_I are *on average* 0, 1, or 2.
- Thus, for these data, it is hard to distinguish between the two models.

A Semi-Parametric Model

- We propose that the total pressure that susceptible individual j is subject to just prior to their colonisation is given as follows:

$$\lambda_j = \begin{cases} \beta_0, & \text{if} \quad n_{C+Q} \in [a, b] \\ \beta_1, & \text{if} \quad n_{C+Q} \in [b+1, c] \\ \beta_2, & \text{if} \quad n_{C+Q} \in [c+1, \infty] \end{cases}$$

  where $b > a$ and $c > b + 1$ are fixed and known.

- Note that we don't make any assumption regarding the relationship of β_0, β_1 and β_2. For example, we don't imply *a-priori* the constraint that $\beta_2 > \beta_1$.

A Semi-Parametric Model (cont).

- In order to fit such a model to our data, we should first choose values for the different levels of colonisation pressure: a, b and c.
- We try the following levels:
  - $[0, 1]$, $[2, 4]$ and $[5, \infty]$
  - $[0, 2]$, $[3, 4]$ and $[5, \infty]$
- An MCMC algorithm is employed in order to draw samples from the posterior distribution of the parameters β_0, β_1 and β_2.
- Note that for this particular model, we do not make the distinction between colonised and isolated or colonised but non-isolated.

A Semi-Parametric Model (cont.)

Assume that the levels of colonisation pressure are

$$[0, 1], [2, 4] \text{ and } [5, \infty]$$

Ward	$P(\beta_1 > \beta_0)$	$P(\beta_2 > \beta_1)$	$P(\beta_2 > \beta_0)$
1	0.86	0.92	0.96
2	0.22	0.95	0.91
3	0.29	0.98	0.97
4	0.94	0.94	0.99
5	0.27	0.51	0.40
8	0.74	1.00	1.00
9	0.13	0.83	0.15
10	0.74	0.68	0.83

A Semi-Parametric Model (cont.)

Assume that the levels of colonisation pressure are

$$[0, 2],\ [3, 4] \text{ and } [5, \infty]$$

Ward	$P(\beta_1 > \beta_0)$	$P(\beta_2 > \beta_1)$	$P(\beta_2 > \beta_0)$
1	0.86	0.92	0.96
2	0.23	0.96	0.92
3	0.30	0.97	0.97
4	0.95	0.95	0.99
5	0.26	0.54	0.42
8	0.73	1.00	1.00
9	0.12	0.81	0.14
10	0.76	0.72	0.85

A Semi-Parametric Model (cont.)

These results ...

- are pretty similar for the different selection of levels of colonisation pressure.
- suggest that for most of the wards there is some evidence to assume that the probability of becoming colonised increases with the number of colonised individuals in the ward.

But ...

- The choice of the levels is arbitrary
- The mixing of MCMC algorithm can be problematic in the case where there is not enough data and the results are driven from the prior.

The impact of undetected colonised patients on transmission.

Undetected cases and test delays

- Our methodology enable us to assess:
  - how much transmission is due to patients who are colonised but not yet detected and
  - how much transmission is due to patients who are colonised and have been tested, but who are awaiting results.
- Define 1 CPD to be one Colonised-Patient-Day, i.e. each colonised patient contributes one unit of CPD for each day they remain colonised.
- We are interested in the mean percentage of total CPD that arose
  - from patients who were colonised but not detected (p_{hidden})
  - from patients who were colonised and tested but awaiting test results (p_{wait}).

Undetected cases and test delays (cont.)

Table: p_{hidden} and p_{wait} fitting the "standard" model

Ward	p_{hidden}	p_{wait}
1	16.5 (15.1, 17.9) %	12.4 (11.6, 13.1) %
2	8.7 (7.3, 10.1)%	4.8 (4.4, 5.0)%
3	19.8 (16.8, 23.1) %	10.9 (9.6, 12.1) %
4	15.5 (13.0, 18.0)%	8.4 (7.6, 9.1)%
5	10.2 (8.6, 12.0)%	5.9 (5.5, 6.2)%
6	13.5 (9.8, 18.3)%	7.1 (5.6, 8.1)%
7	10.8 (9.8, 11.7)%	9.6 (8.9, 10.2)%
8	13.2 (11.8, 14.7)%	8.7 (7.9, 9.3)%
9	17.1 (8.8, 13.2)%	7.8 (7.0, 8.4)%

So, roughly speaking:

- about 10% - 15% of patient-colonised days are undetected
- about 10% of patient-colonised days occur due to delays in obtaining test results.

Future Research Directions

Bayesian Model Choice

- The implementation of the Trans-dimensional MCMC algorithm involves the computation of the likelihood, i.e. calculation such as:
  - $$\int_{T_S}^{T_E} \beta_0 S_t \, \mathrm{d}t, \int_{T_S}^{T_E} \beta_1 C_t S_t \, \mathrm{d}t, \int_{T_S}^{T_E} \beta_1 Q_t S_t \, \mathrm{d}t$$

    (for the "standard model").
  - $$\int_{T_S}^{T_E} \beta_0 S_t \, \mathrm{d}t, \int_{T_S}^{T_E} \beta_1 \mathbb{I}(C_t > 0) S_t \, \mathrm{d}t, \int_{T_S}^{T_E} \beta_1 \mathbb{I}(Q_t > 0) S_t \, \mathrm{d}t$$

    (for the "simpler model")
- Although these integrals can be translated into sums, nevertheless such calculations can be very computationally expensive and this is especially the case for "simpler model".

Application-Driven Directions

- Fit the same or similar models to other pathogens, e.g. VRE, GRE etc?
- How typing data can be incorporated into such or different models?
- Models taking into account on which antibiotics usage will enable us to assess the effectiveness of some and reveal the role they are playing in transmission.

Conclusions

Conclusions

- A Bayesian framework for modelling healthcare associated infections has been presented.
- State-of-the-art Markov Chain Monte Carlo methods have been developed to efficiently draw inference for the model parameters.
- Trans-dimensional MCMC algorithms have been employed to explore different models.
- Identified future research directions.

Acknowledgments

- Wellcome Trust (funding).
- Dr Susan Huang @ Harvard Medical School.