Hierarchical modelling of performance indicators, with application to MRSA & teenage conception rates

Hayley E Jones
School of Social and Community Medicine,
University of Bristol, UK

Thanks to David Spiegelhalter,
Statistical Laboratory / MRC Biostatistics Unit, Cambridge

Young Statisticians working in Medical Research

Avon RSS Local Group meeting
Bristol, 10th February 2011

Introduction

- Large quantities of routine ‘performance’ data now collected on many healthcare providers at regular intervals.
- Often strict government targets for performance improvements.
- Population level performance & ‘unusual’ providers both of interest.
- Standard to first adjust for risk factors beyond the influence of providers (‘case mix’).

Introduction

- Large quantities of routine ‘performance’ data now collected on many healthcare providers at regular intervals.
- Often strict government targets for performance improvements.
- Population level performance & ‘unusual’ providers both of interest.
- Standard to first adjust for risk factors beyond the influence of providers (‘case mix’).
- **We will compare models for risk-adjusted performance measures based on their predictive ability.**

e.g.1: MRSA bacteraemia rates in NHS Trusts

- Cases of Methicillin resistant *Staphylococcus Aureus* bacteraemia infections in hospital are high profile example.
- Mandatory surveillance in all NHS Trusts since 2001.

e.g.1: MRSA bacteraemia rates in NHS Trusts

- Cases of Methicillin resistant *Staphylococcus Aureus* bacteraemia infections in hospital are high profile example.
- Mandatory surveillance in all NHS Trusts since 2001.

- $i = 1, ..., 171$ NHS Trusts
- $t = 1, ..., 11$ six-month time periods

- O_{it} = observed no. of infections
- E_{it} = 'expected' infections, based on Trust 'size' & average rate for Trust type

e.g.1: MRSA bacteraemia rates in NHS Trusts

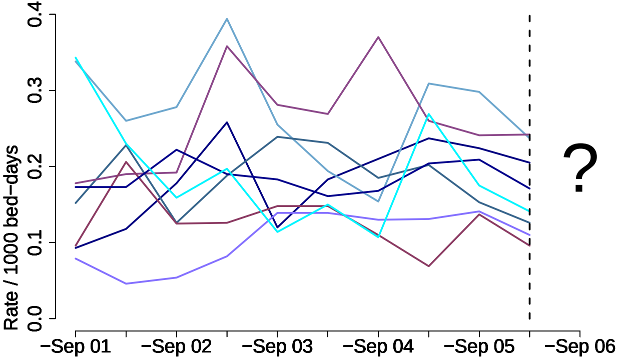
- Cases of Methicillin resistant *Staphylococcus Aureus* bacteraemia infections in hospital are high profile example.
- Mandatory surveillance in all NHS Trusts since 2001.

- $i = 1, \ldots, 171$ NHS Trusts
- $t = 1, \ldots, 11$ six-month time periods

- O_{it} = observed no. of infections
- E_{it} = 'expected' infections, based on Trust 'size' & average rate for Trust type

- Assume $\boxed{O_{it} | r_{it} \sim Poisson(r_{it} E_{it})}$

e.g.1: MRSA bacteraemia rates in NHS Trusts



Data on 8 of 171 NHS Trusts

e.g.2: Teenage conceptions in English Local Authorities

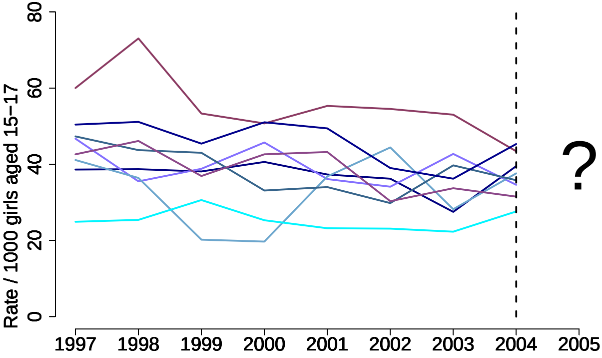
- Britain's teenage pregnancy rate is the highest in Western Europe.
- Government targets for rate reductions.

- $i = 1, \ldots, 352$ Local Authorities
- $t = 1, \ldots, 9$ years

- O_{it} = observed no. of under-18 conceptions
- E_{it} = 'expected' infections, based on regression model (adjust for population size, deprivation, education, rurality)

- Assume $\boxed{O_{it}|r_{it} \sim Poisson(r_{it}E_{it})}$

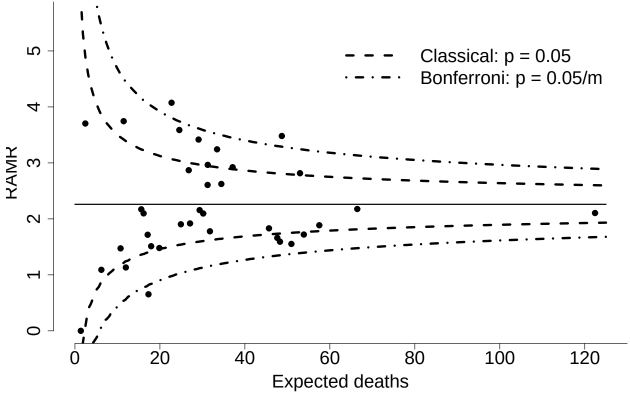
e.g.2: Teenage conceptions in English Local Authorities



Data on 8 of 352 Local Authorities.

Modelling cross-sectional data

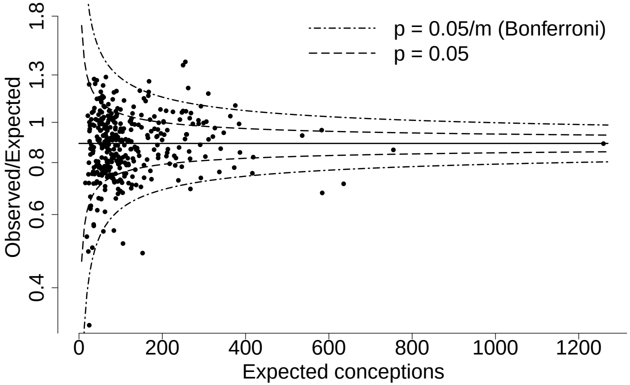
Funnel plots: Plot each risk-adjusted rate against a measure of its precision.



Test if each provider's rate is equal to the average.

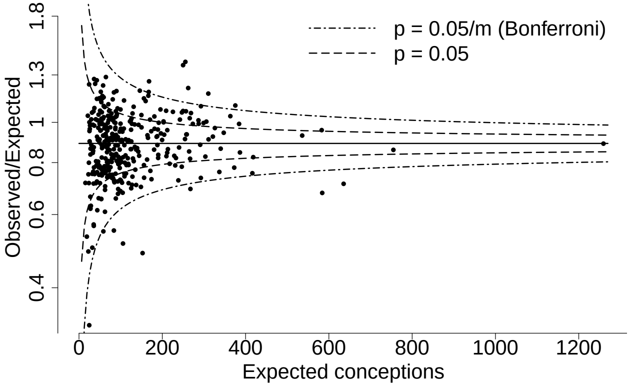
Modelling cross-sectional data

For teenage conceptions data (& MRSA), considerable heterogeneity in rates in any one year.



Modelling cross-sectional data

For teenage conceptions data (& MRSA), considerable heterogeneity in rates in any one year.



Motivates considering a hierarchical model in this context.

Motivation for a hierarchical model

- Some dispersion of true rates around the mean is to be expected, due to **imperfect risk adjustment**.
  $\rightarrow$ Allow some leeway in the null, to prevent too many providers being identified as ‘unusual’.

Motivation for a hierarchical model

- Some dispersion of true rates around the mean is to be expected, due to **imperfect risk adjustment**.
  $\rightarrow$ Allow some leeway in the null, to prevent too many providers being identified as 'unusual'.

- Resulting **shrinkage estimates** of underlying rates have attractive properties:
  Improve estimation performance by drawing on info from other units.
  - Deal with small counts effectively, increasing precision.
  - Improve predictions / Handle 'regression-to-the-mean'.

Hierarchical model

e.g. Poisson-gamma model:
$r_i|\mu, \tau \sim IID\ Gamma[\mu, \tau^2]$.

Assuming μ and τ are known, we obtain

Shrinkage estimate:

$$\hat{r}_i = w_i \frac{O_i}{E_i} + (1 - w_i)\mu$$

where

$$w_i = \frac{\tau^2}{\tau^2 + \mu/E_i}$$

- Empirical Bayes: use plug-in estimates of μ and τ.
- Implied predictive distribution for next period is negative binomial.

Modelling longitudinal performance data

- Well-known advantages of hierarchical models / shrinkage estimators (*i.e.* **smoothing between providers**).
- With longitudinal data we also have the option to **smooth within providers** over time.

Modelling longitudinal performance data

- Well-known advantages of hierarchical models / shrinkage estimators (*i.e.* **smoothing between providers**).
- With longitudinal data we also have the option to **smooth within providers** over time.
- We could smooth independently within each unit (e.g. Poisson regressions / EWMA), or combine both types of smoothing in a hiearchical longitudinal model
  = **'Bidirectional smoothing'**.

Bidirectional smoothing

Hierarchical AR(1) model (Lin *et al.*, 2009)

$$O_{it}|r_{it} \sim Poisson(r_{it}E_{it})$$

Simple hierarchical model assumed to hold marginally in each time period:

$$log(r_{it}) \sim Normal(\mu_t, \tau_t^2) \ .$$

Bidirectional smoothing

Hierarchical AR(1) model (Lin *et al.*, 2009)

$$O_{it}|r_{it} \sim Poisson(r_{it}E_{it})$$

Simple hierarchical model assumed to hold marginally in each time period:

$$log(r_{it}) \sim Normal(\mu_t, \tau_t^2) \ .$$

Time series structure on each standardised process:

$$\frac{log(r_{it})-\mu_t}{\tau_t} = \phi\left(\frac{log(r_{i,t-1})-\mu_{t-1}}{\tau_{t-1}}\right) + \eta_{it} \ , \qquad t = 2, ..., T$$

Fit in WinBUGS.

Extension to Lin et al model

To automatically make **predictions**, we extended the model to incorporate a **random walk for the population mean:**

Random walk on μ_t:

$$\mu_t = \mu_{t-1} + \delta_t$$
$$\delta_t \sim Normal(0, \sigma_\mu^2)$$

Extension to Lin et al model

To automatically make **predictions**, we extended the model to incorporate a **random walk for the population mean:**

Random walk on μ_t:

$$\mu_t = \mu_{t-1} + \delta_t$$
$$\delta_t \sim Normal(0, \sigma_\mu^2)$$

Similarly for the log of the **population standard deviation:**

Random walk on $log(\tau_t)$:

$$\log(\tau_t) = log(\tau_{t-1}) + \epsilon_t$$
$$\epsilon_t \sim Normal(0, \sigma_\tau^2)$$

Bidirectional smoothing

Such models for performance data, resulting in 'bidirectional' smoothing, are increasingly being suggested in the literature (e.g. West & Aguilar, 1998; Van Houwelingen *et al*).

But:

- Involve considerable extra complexity.
- No systematic evaluation has been made, comparing these models with simpler 'one-way' smoothing alternatives.

Evaluation criteria

We use 3 approaches, drawing heavily on recommendations of Gneiting *et al.* (2005) in the field of weather forecasting:

1. **Accuracy of point predictions:**
   $MSE = \frac{1}{m}\sum_{i=1}^{m}(O_{iT} - \hat{O}_{iT})^2$
   $MAE = \frac{1}{m}\sum_{i=1}^{m}|O_{iT} - \hat{O}_{iT}|$

Evaluation criteria

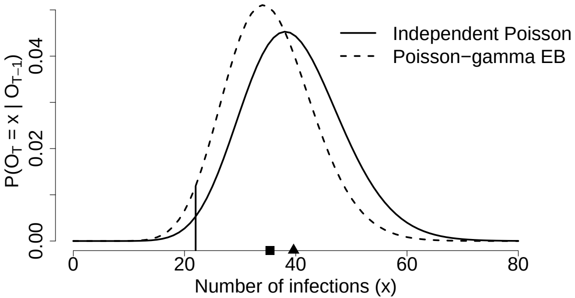
We use 3 approaches, drawing heavily on recommendations of Gneiting *et al.* (2005) in the field of weather forecasting:

1. **Accuracy of point predictions:**
   $MSE = \frac{1}{m}\sum_{i=1}^{m}(O_{iT} - \hat{O}_{iT})^2$
   $MAE = \frac{1}{m}\sum_{i=1}^{m}|O_{iT} - \hat{O}_{iT}|$

   But also evaluate the full forecasting distributions:
2. Uniformity of predictive p-values.
3. Proper scoring rules: log-score and CRPS.

Example of 2 predictive densities

Number of MRSA infections in a particular NHS Trust.



▲ Point prediction under independent Poisson model.

■ Point prediction under Poisson-gamma EB model.

Uniformity of predictive p-values

- A correctly calibrated forecasting distribution
  $\equiv$ Events declared to have probability p occur a proportion p of the time on average.

- If this is the case, then the set of predictive p-values should have an approximately *Uniform(0,1)* distribution.

- Can assess this:
  1. **Visually**, using histograms and plots of ordered p-values.
  2. **Using test statistics** *e.g.* Kolmogorov-Smirnov D or Cramér-von-Mises W^2.

Proper scoring rules

- Uniformity of the predictive p-values is necessary but not sufficient condition for forecasting system to be 'ideal' (Gneiting *et al.*, 2007).
- Therefore also consider 2 'proper scoring rules'.

Logarithmic Score

$$LS_i = -log(f(O_i))$$

Continuous Ranked Probability Score

$$CRPS(F_i, O_i) = E|O_i^{pred} - O_i| - \frac{1}{2}E|O_i^{pred} - O_i^{pred'}|$$

- Examine mean of these over providers: lower preferred.

Results

Model comparison for the **teenage pregnancies** data:

		MSE	*MAE*	*D*	W^2	CRPS	LS
0	Pois indep	249	11.2	0.04	0.08	8.00	3.98
1	PG EB	197	10.0	0.07	0.41	7.17	3.87
1	P-LN Bayes	197	10.0	0.06	0.28	7.16	3.87

Results

Model comparison for the **teenage pregnancies** data:

		MSE	*MAE*	*D*	W^2	CRPS	LS
0	Pois indep	249	11.2	0.04	0.08	8.00	3.98
1	PG EB	197	10.0	0.07	0.41	7.17	3.87
1	P-LN Bayes	197	10.0	0.06	0.28	7.16	3.87
2	HR AR(1)	189	9.7	0.05	0.33	6.88	3.83

0) No smoothing, 1) Smooth between, 2) Bidirectional smoothing.

Results

Model comparison for the **MRSA** data:

		MSE	*MAE*	*D*	W^2	CRPS	LS
0	Pois indep	52	5.5	0.09	0.24	3.81	3.19
1	PG EB	36	4.5	0.06	0.07	3.15	3.01
1	P-LN Bayes	37	4.6	0.06	0.09	3.17	3.01
2	HR AR(1)	33	4.3	0.06	0.14	2.96	2.96

0) No smoothing, 1) Smooth between, 2) Bidirectional smoothing.

Conclusions

- Smoothing observed performance measures has clear benefits in terms of predictive accuracy.
- This is now well recognised & hierarchical models are increasingly used for modelling cross-sectional data.

Conclusions

- Smoothing observed performance measures has clear benefits in terms of predictive accuracy.
- This is now well recognised & hierarchical models are increasingly used for modelling cross-sectional data.
- Smoothing *within* in addition to *between* providers using the 'bidirectional' smoothing model of Lin *et al.* was found to have additional benefits for 2 example datasets.
- This model is highly interpretable, automatically adapts to characteristics of dataset & is straightforward to program in WinBUGS.

Conclusions

- Seems reasonable to suggest this model should be used as a default.
- However, it requires use of specialist software (WinBUGs) & takes a very long time to fit!

Conclusions

- Seems reasonable to suggest this model should be used as a default.
- However, it requires use of specialist software (WinBUGs) & takes a very long time to fit!
- **Future research should focus on faster / simpler methods for fitting bidirectional models**.

References

[1] H E Jones and D J Spiegelhalter. Improved probabilistic prediction of healthcare performance indicators using bidirectional smoothing models. *Under revision for Journal of the Royal Statistical Society Series A*.

[2] T Gneiting, A E Raftery, A H Westveld III, and T Goldman. Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Monthly Weather Review*, 133:1098–1118, 2005.

[3] R Lin, T A Louis, S M Paddock, and G Ridgeway. Ranking USRDS provider-specific SMRs from 1998-2001. *Health Services and Outcomes Research Methodology*, 9:22–38, 2009.