Bayesian methods in Pharmacoepidemiology

- Stephen Evans
- *London School of Hygiene & Tropical Medicine*
- stephen.evans@Lshtm.ac.uk

Declaration of Interests

- Formerly employed at UK Medicines Control Agency (MHRA)
  - & at Quintiles, UK
- Co-opted (Independent) Member of Pharmacovigilance Working Party at European Medicines Agency (EMA)
- Member - WHO Global Advisory Committee on Vaccine Safety
- LSHTM (but not SE) funded by several pharmaceutical companies & via DSRU, Southampton.
- OMOP fee for review
- Formerly a Statistical advisor & "hanger" at BMJ

Thanks & Acknowledgements

- Thanks to
  - ISPE for inviting me
  - David Prieto-Merino – joint work on Bayesian Hierarchical Models & some slides
  - Laurence McCandless – joint work & some ideas for slides
  - David Spiegelhalter and Bill DuMouchel for many ideas and discussions

What this session will cover

- Some History & philosophy – *but skipped for sake of time?*
- What is controversial & what is not
  - *Bayes Theorem & diagnostic tests*
- Basic Bayesian analysis
- Analysis of data from spontaneous report databases (could also be applied in OS & RCTs)
- Some applications in pharmacoepidemiology

History

- Probability was of interest to gamblers and also to philosophers from the 16th and 17th centuries
- There were various views about probability and one of the early people interested in the theory was the Reverend Thomas Bayes (1702—1761)
- He was a Non-conformist minister from Tunbridge Wells, UK
- His main work was edited and published by his friend, Richard Price, who presented the work in 1763, "An Essay towards solving a Problem in the Doctrine of Chances"
- Laplace replicated and extended the work in 1774, apparently unaware of Bayes' work.

Thomas Bayes

- Very little is known about him, but he was a member of the Royal Society
- "Dissenters" could not be buried in Church of England Graveyards, and London has a Cemetery in which noted dissenters were buried at Bunhill Fields:
  - Bayes, Bunyan, Defoe, Wesley

Probability

- What is the probability that a coin will show heads after being tossed?
- The *Frequentist* view is that it is the proportion of times a head will appear in a very large number of tosses (possibly an infinite number)
- If it is a “fair” coin, fairly tossed it is 0.5
- This is sometimes referred to as a *Classical* view

“Classical” inference

- Mainly developed in 20^{th} century, but earlier work related to gambling
- RA Fisher, Neyman, Pearson
- Significance testing; hypothesis tests
- Confidence intervals – strange interpretation
- Its proponents emphasise it is objective
- The observed data give all the information relevant to that experiment

Probability based on belief

- An alternative view says probability is based on belief
  - Lindley, de Finetti, Savage in 20^{th} century, but possibly Bayes and others in pre-20^{th} century times regarded probability as based on belief
- Bayes’ theorem can be applied to probabilities of both types, but is usually applied to the belief-based form of probability

Changing Attitudes

Interview with Don Berry *Stat Sci 2012*

- *"If you read Bayesian polemic from the 1970s and 1980s—including my own—it's usually arrogant and even insulting. Some of the terms were excessively pointed. For example, Bayesians identified which frequentist methods were "incoherent," or more accurately, lamented that none seemed to be coherent. On the other hand, Bayesians were accused of being "biased."*

Don Berry, continued

- *The rhetoric was not all that different from that of the Fisher/Pearson duels. But we Bayesians have stopped saying derogatory things, partly because we have changed and partly because frequentists have been listening.*

- www.imstat.org/sts/future_papers.html

Classical confidence intervals

- The "true" proportion of people getting an adverse reaction to a drug is unknown, but an estimate may be obtained
- A 95% confidence interval can be calculated, and it can be shown that the true value will, in the long run, be contained within 95% of those 95% confidence intervals

- E.g. Greenland (*IJE 2006;**35:765-75)***

Methods of calculating CIs differ
E.g. Greenland *(IJE 2006;35:765-75)*

	Exposed (>3mG)	Unexposed
Cases	3	33
Controls	5	193

OR =3.51		
	"Exact" 95% CI	0.52 18.9
	Cornfield	0.89 14.1
	Test-based	0.87 14.2
	Woolf and Logistic regression	0.80 15.4

Bayes' Theorem

- Conditional probability
- 1% of women at age 40 have breast cancer
- 80% with BrCa have +ve mammograms
- 9.6% without BrCA have +ve mammograms

- What is probability that a woman with a +ve mammogram will have BrCa?
- (+ve predictive value)

Screen 10,000 women

mammogram	Without BrCa	With BrCa
+ve	950	80
-ve	8950	20

Chance of having breast cancer =80/(80+950) = 7.8%

This is heavily dependent on not just the properties of the screening test but the prevalence of breast cancer

Bayes Theorem as formula

$$P(Ca|m+) = \frac{P(m+|Ca) * P(Ca)}{P(m+|Ca) * P(Ca) + P(m+|\sim Ca) * P(\sim Ca)}$$

This changes the prior probability of cancer as a result of the test

The prior was 1%, the posterior was 7.8%

The probability of the data given a hypothesis is called the ***likelihood***

A P value is P(data or a more extreme result|null hypothesis)

Bayesian inference

- This uses Bayes' theorem to combine the prior probability of a hypothesis, with the likelihood from the observed data, to obtain a *posterior* probability for that hypothesis
- In practice, we have a range of possible prior values, and obtain the probability for a range of possible values when we have taken both prior belief and the observed data into account

Likelihood ratio

- This is simply the ratio of likelihoods for different hypotheses for the same set of data
- Bayes' Theorem can also be set out as

posterior odds= prior odds * likelihood ratio

log(posterior odds)= log(prior odds)+ log(LR)

Prior probabilities

- These can be very uncertain- a *vague* prior
- They can be rather more certain
- The range of prior values can be very wide or very narrow
- This is the same as effectively having very little prior data or having quite a bit of prior data

Greenland's example

- Prior is that OR ~1, 95% prob 0.25-4 (prior variance of log(OR)=0.5
- Can use weighted means (inverse variance) of priors and data, assuming a Normal distribution
- Used for mean OR and its variance
- Original OR 3.51(0.80-15.4)
- "Bayesian" OR=1.8 (0.65-4.94)
- This has "shrunk" the estimate of the OR and its CI towards an OR of 1 – obviously as it is an average of the prior which was 1, and the observed OR of 3.51

An alternative view by Greeenland

- Can regard the prior as being represented by a set of data which are added as a separate stratum and a stratified frequentist analysis done
- Add a set of data corresponding to

4	4
10000	10000

- RR=1 ~95% CI 0.25-4
- (similar values for 100 instead of 10000)
- using LR gives OR = 1.76 (95%CI 0.59 - 5.23)

Empirical Bayes

- This uses (old) data to obtain the prior probabilities, rather than just beliefs, but then uses the same logic to obtain posterior probabilities combining it with some particular (new) data
- It can be that the prior distribution is itself a mixture of two distributions
  - E.g. one might have a relative risk of 1, while another has a relative risk centred on 0.5

Signals from spontaneous reports

- New evidence about an association between a drug and an adverse event (AE)
- More spontaneous reports (SRs) than we expect
- Process of detection → evaluation (impact) → new data required → regulatory action

Brief review of PRR

- Proportional reporting Ratios (PRR) are the simplest method of scanning a database for signals of disproportionate reporting
- They use the information solely from within a database of spontaneous reports
- Usage data can be used to look at overall reporting rate for a drug in conjunction with PRRs but is subject to even more bias than the PRR

Focusing on one of the drug/event combinations

Sum all other Events, & all other suspect drugs

	E1	E2	E3	.	.	.	.	.	.	.	Em
D1	o_{11}	o_{12}	o_{13}	.	.	.	.	.	.	.	.
D2	o_{21}	o_{22}	o_{23}	.	.	.	.	.	.	.	.
D3	o_{31}	o_{32}	o_{33}	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.	.	.
Dn	.	.	.	.	.	.	.	.	.	.	.

Drugs by events condensed to 2 x 2 table

	E1	E2-m
D1	a	b
D2-n	c	d

PRR Calculation

	Specific event	All other events
Specific drug	a	b
All other drugs	c	d

(a/(a+c)) / (b/(b+d)) is PRR

Equivalent to observed/expected count

we can use odds ratio (ad/bc) – ROR

this has some mathematical advantages but difference is trivial – Commentary {Waller et al PDS 2004}

Extension to "data mining"

- The calculation for a single drug/event combination shown for one cell, is then repeated for the next drug/event combination
- Every drug/cell combination with >0 in the cell has the same calculation done
- This produces a PRR (or other method's summary) for every cell –**n.b. multiplicity**
- The frequentist method has a cut-off to detect a potential signal (PRR/ROR + uncertainty)

Two major Bayesian Methods

- Gamma Poisson Shrinker (GPS)
- Multi-item GPS (MGPS)

  All "shrink" estimates of O/E towards null value- by a variable amount, larger if O or E count is small
- Bayesian Confidence Propagation neural network (BCPNN)
  - Result expressed as Information Component –IC

Bayesian approaches

- DuMouchel uses empirical Bayes
- He fits a mixture of 2 distributions to the totality of the data- all calculated PRRs
- First distribution is assumed to be "noise" – essentially centred on a PRR of 1 with a wide scatter
- Second is "true" ADRs centred on a PRR >1 again with a wide scatter
- The mixture can vary in % in each distribution

Empirical Bayes Geometric Mean

- Then, the prior distribution for each cell is based on the overall data so most cells are assumed to be “noise” effectively
- The data from that cell is then used to get a likelihood and hence a posterior distribution
- The EBGM value is the mean of this posterior and it has a “credible interval” and the lower 5% value is often used for to “signal” - EB05
- This results in “shrinkage” of the cell PRR towards 1- depending on the amount of data

WHO BCPNN

- Bayesian Confidence Propagation Neural Network
- Similar principles to DuMouchel
- Computer science rather than statistical approach
- Different prior distribution – fixed for WHO data

Bayesian methods

- Can be seen as related to

{(O+k)/(E+k)}

where when k=½ result is close to IC (BCPNN) method

This minimises effect of a small E, especially when O is small

O = 5, E = 0.01

PRR = 5/0.01 = 500

IC = 5.5/0.51 = 10.8 {IC actually uses $\log_2(10.8)$}

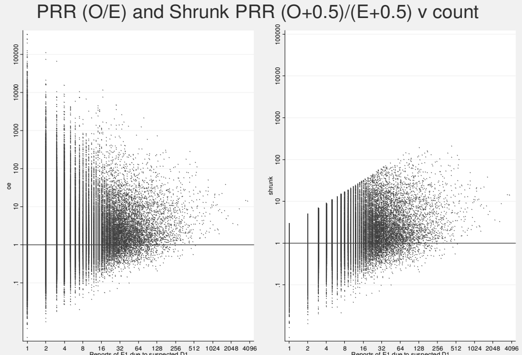
The effect of Bayesian methods

- Their effects are very important with low expected counts & quite important with low observed counts –aid **multiplicity problem**
- They "shrink" the value of O/E towards 1.
- For BCPNN it is close to-
  (O+½)/(E+½)
- E.g. with first report for a drug that is rash
- (1+½)/(.036+½) = 2.8 (a lot less than 27.8)

Shrinkage

- If expected value is even smaller, then even more extreme results can be obtained
- E.g. a drug levocarnitine had a single report for Alopecia effluvium, in a database of 1,000,116 combinations. Alopecia effluvium was only reported 3 times, so the expected rate was 3 in 1 million and the PRR is 300,000
- The "shrunk" value of PRR is never >3 with count =1

PRR (O/E) and Shrunk PRR (O+0.5)/(E+0.5) v count



Problem of using Preferred Term

- There are 24 different PTs that have "rash"
- These in total form nearly 7% of the PTs
- Our current methods usually treat each PT individually
- Some work has been done at looking at higher levels but this has problems
- Ignoring other PTs that involve rash may lower sensitivity of any automatic method

Suggestion

- Berry & Berry approach to AEs
  - Bayesian analysis using information from similar ADRs to address problems of multiplicity
- *Biometrics (*June) 2004 60:418-26.
  - (there may be statistical problems with this paper)
- if several reports of different e.g. cardio-vascular events occur, then each is more likely to be causal than if single reports come from medically unrelated areas (e.g. skin, neurological, thrombosis, cancer)

"Borrowing" information

- This is an important idea that reflects clinical judgement
- Berry & Berry used SOC to group terms
- It will be better to use more narrow, medically meaningful groupings- experience from spontaneous reports will be helpful

Suggestion

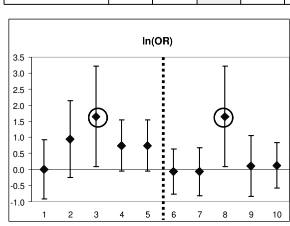
- The use of standardised MedDRA queries allows for this in a different way
- Grouping of medically related terms to describe a single medical condition
- Signal detection methods should use these wherever possible
- MedDRA MSSO with CIOMS has developed some e.g. Rhabdomyolysis, Torsades de Pointes

Bayesian Hierarchical Models (BHM) allow us:

1. To adjust for multiple testing but "penalising" each combination with a more rational criteria.
2. To estimate mean effects of groups of drugs and adverse events and put the effect of each drug into "context".
3. To incorporate previous pharmacological or clinical knowledge.

A brief explanation of BHM:

AEs:	AE_1	AE_2	AE_3	AE_4	AE_5	AE_6	AE_7	AE_8	AE_9	AE_{10}
SD_1 N= 200	10	10	10	20	20	17	15	10	10	18
SD_0 N= 200	10	4	2	10	10	18	16	2	9	16
OR	1	2.6	**5.2**	2.1	2.1	0.9	0.9	**5.2**	1.1	1.1



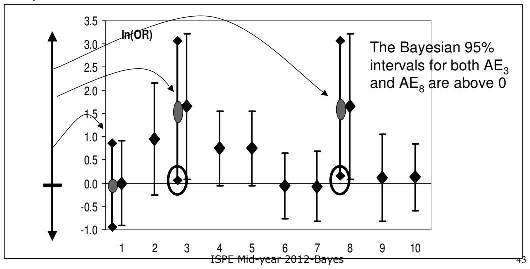
AE_3 and AE_8 show SAME NUMBERS, a high OR and are suspicious but... they could be false positives...!!!

Should we penalise them equally?

Assume it makes sense to group the AEs as shown, and expect similar behaviour in all members of each group. Should we have the same belief now on AE_3 and AE_8 being false positive?

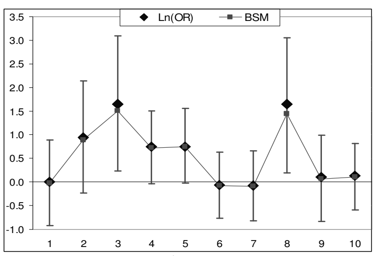
Bayesian Simple Model (Non hierarchical, the groups are not used)

- Each AE is analysed separately with a non-informative prior with mean "0" and large variance.
- The posterior Bayesian interval for the AE are very similar to the data and the frequentist confidence intervals.



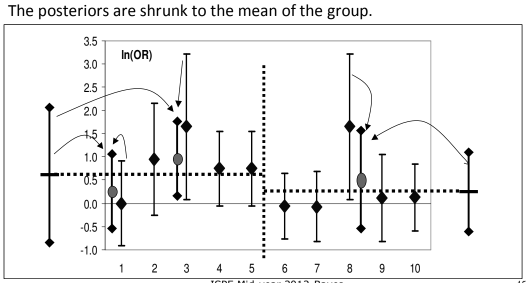
Bayesian Simple Model (Non hierarchical, the groups are not used)

- Each AE is analysed separately with a non-informative prior with mean "0" and large variance.
- The posterior Bayesian interval for the AE are very similar to the data and the frequentist confidence intervals.



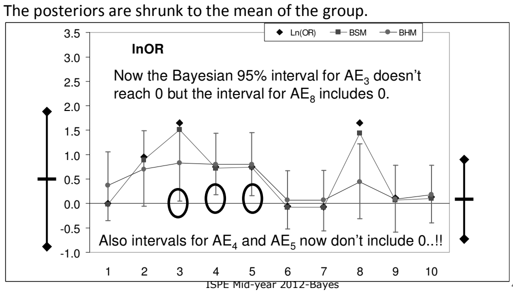
Bayesian Hierarchical Model (BHM): Imagine a distribution inside each group

- We use the distribution of AEs in each group as a PRIOR for each AE in the group.
- The posteriors are shrunk to the mean of the group.



Bayesian Hierarchical Model (BHM): Imagine a distribution inside each group

- We use the distribution of AEs in each group as a PRIOR for each AE in the group.
- The posteriors are shrunk to the mean of the group.



Applications of Bayesian methods

- BHM in trials, observational studies, spontaneous reports
  {Prieto-Merino D et al. (2011) Pharm Stat;10:554-9. Crooks C et al. Drug Saf. 2012;35:61-78}
- Missing Data & Unmeasured confounding
- {McCandless, L.C. et al Int J Biostat, 2010; 6:Article 16}
- {McCandless, L.C. et al Stat Med. 2007; 26:2331--47.}
- {McCandless, L.C. et al Stat Med. 2012;31:383-96.}
- Bayesian Propensity Scores (modelling uncertainty in PS)
- {McCandless, L.C. et al Stat Med 28:94-112}

New paper by Bill DuMouchel

www.imstat.org/sts/future_papers.html

Multivariate Bayesian Logistic Regression for Analysis of Clinical Trial Safety Issues

+ Discussion/commentaries

May well have application in observational studies as well

Bayesian methods have continuing innovation