

Data Mining of Report Databases: Adverse Drug Reactions vs. Automobile Safety Defects

William DuMouchel
Oracle Health Sciences

Committee for a Study of Electronic Vehicle
Controls and Unintended Acceleration

Public Session—15 November 2010

Spontaneous Report Database

- Collection of reports without a designed surveillance system
- Problems arise among a population of millions of users, some of whose reports filter back to a common authority
- Numerator-only analysis problem
 - Event counts without good measure of exposure for the different product categories
 - Variable problem reporting rates across product categories
- Data may have severe quality issues
 - Missing fields, incorrect coding, duplicate reports, etc.
- FDA AERS— >3M reports of drug adverse events (AEs)
 - About 5,000 generic drugs, 10,000 AE codes, over 40 report years

Searching for Associations

- No denominator to help measure absolute measure of risk
- Define measures of association or correlation among product categories and problem types
- Extremely helpful to have structured vocabularies
 - Products (drugs, vehicle characteristics)
 - Problems (adverse events, accidents or safety issues)
 - Reports often come in with free text narratives only
- Focused analysis vs screening for associations
 - Pre-specified problems with hypothesized causal mechanisms
 - Search for unanticipated relationships or product-problem associations
 - Statistical issues of multiple comparisons

Objectives and Limitations of Analyses of Spontaneous Reports

■ Explore for Drug-Event Associations

- Estimate a Measure of Association for every Combination
- How Can a Rate Be Defined without a Denominator?
 - Matching External Sales or Prescription Counts Not Feasible
 - Construct Internal Denominators from Independence Model
- Screening Objective – All Findings Require Follow-up

■ Severe Limitations of Data Reliability

- No Research Protocol
- Adverse Event Report Rates Vary from Year to Year
- Report Rates Vary by Drug and by ADR Type
- No Certainty that a Reported Reaction Was Causal

Two Data Cleaning Issues

■ Drug Name Standardization

- The AERS Database has over 300,000 “verbatim” drug names
 - Generic and Trade names
 - Misspellings
 - Dose included with drug name
- Now reduced to about 3000 ingredient names
 - Years of effort!

■ Duplicate Detection

- Same ADR Event often reported multiple times
 - By different manufacturers or other reporters
 - Follow-up reports not properly linked to earlier reports
- 3 Million reports have had about 200,000 duplicates removed
 - Undetected duplicates can severely bias estimated drug-event associations

Spontaneous Reports 2 x 2 Analyses

- For every $D_i E_j$ pair = (Drug of Interest, Event of Interest)
 - Use the database to tabulate a 2 x 2 table of report counts
 - Compute an *expected* or *baseline* count e from (a, b, c, d)
 - Based on assumption of no association between Drug and Event
 - $e = b(a + c)/(b + d)$ [Proportional Reporting Ratio method]
 - $e = bc/d$ [Reporting Odds Ratio method]
 - $e = (a + b)(a + c)/(a+b+c+d)$ [Relative Report Rate: MGPS method]
 - This method works best when adjusting for trend or demographic covariates in computation of e
 - n/e = Measure of Disproportionality for this Drug and Event

	Reports With Drug i	Reports W/O Drug i	Total
Reports With Event j	$n_{ij} = a$	b	$a + b$
Reports W/O Event j	c	d	$c + d$
Total	$a + c$	$b + d$	$a+b+c+d$

Disproportionality Analyses

- Although the idea of computing n/e ratios for all or some drug-event combinations is simple, its widespread use is very recent
 - Computer and database advances enabled ease of use and evaluation
- Biostatisticians were uncomfortable with performing formal analyses on tabulations of spontaneous reports
 - Unknown reporting mechanism can lead to reporting biases
 - Frequent noncausal associations with indications and comorbidities
 - All large values of n/e require follow-up for medical validity
- Small values of n and/or e require statistical sophistication
 - PRR requires threshold values of n and 2 x 2 table chi-squared value
 - Bayesian statistical methods produce “shrinkage” values of n/e
 - Help avoid the “multiple comparisons” fallacy
 - US FDA, UK MHRA and WHO UMC have each adopted Bayesian disproportionality methods

Adjusting for Patient Covariates

- Drug-Event Associations Can Be Induced by Confounders
 - Drugs predominately taken by one age group
 - Events predominately experienced by that age group
 - Example: Childhood vaccines and SIDS
- Stratification As an Adjustment Method
 - Divide the database into age-gender-report year strata
 - Compute the baseline expected values e for each stratum
 - Add the e 's across strata to compare to the total n
 - Mantel-Haentzel adjustment of 2x2 tables
- Other Potential Variables to Adjust For
 - Country of origin of the report
 - Role of reporter (consumer, manufacturer, health care provider, ...)

Combined Analysis of Drug-Event Counts in a Database

- Large Two-Way Table with Possibly Millions of Cells
 - One Column for each Drug, One Row for each Event
 - Rows and Columns May Have Thousands of Categories
 - Most Cells Are Empty, even though $N..$ Is very Large
- “Bayesian Data Mining in Large Frequency Tables”
 - *The American Statistician* (1999) (with Discussion)
 - Develops and Illustrates Bayesian Estimation Method “GPS”
- Whitfield & Whitfield, *Injury Prevention* 2004;10:88–92
 - References above paper and applies to tire and fire related crashes
 - Does NOT implement the recommended Bayesian methodology
 - Used alternative P-Value method that was not recommended

Association Measures of Rare Events

- Ratios of (Observed count)/(Expected count)
 - Easy to interpret as analogs of relative risk
 - Often too variable when Expected count is small
 - Observed = 3, Expected = 0.01 (300:1 risk ratio)
 - Compare to e.g. Observed = 300, Expected = 10 (30:1 risk ratio)
- Compute statistical significance levels (P-values)
 - Chi-squared of 2x2 tables, Poisson probability of Observed|Expected
 - Focuses too much attention on larger counts in the database
 - P-values themselves are not intuitive nor good for ranking associations
- Bayesian shrinkage estimates of Observed/Expected
 - Need parallel situation having very many things to estimate
 - Get more reliable estimates risk ratios

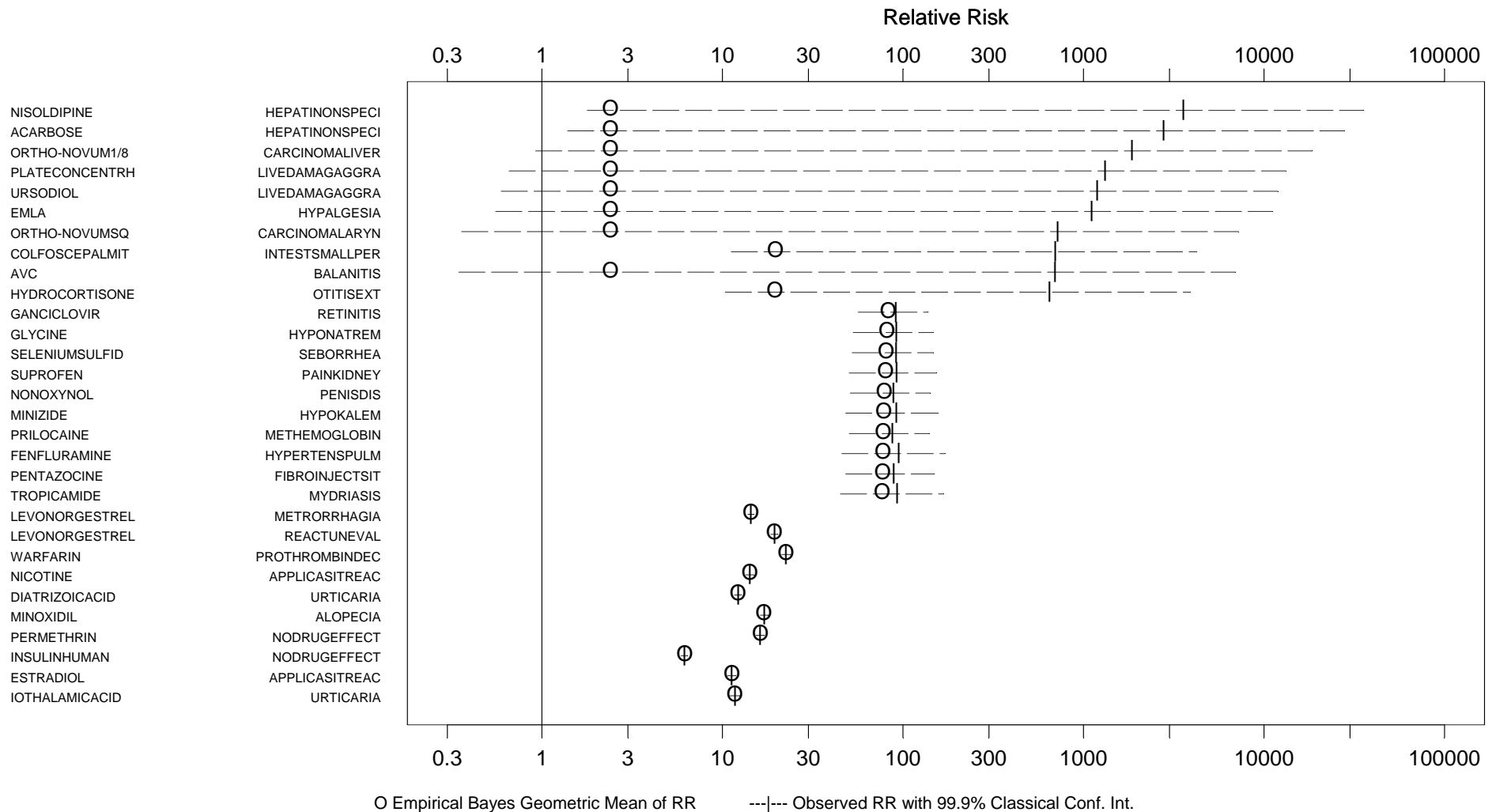
Bayesian Shrinkage Models

- Statistical validity of searching for extreme differences
 - Most significant adverse event or patient subgroup
- Classical approach to post-hoc interval estimates
 - Maintain centers of CI at observed differences
 - Expand widths of every CI
 - Expansion is greater the more differences you look at
 - If you look at too many, the CI's are too wide to be useful
- Bayesian approach
 - Requires a prior distribution for differences
 - Can estimate it from the multiple observed differences available
 - Centers of CI's are “shrunk” toward average or null difference
 - High-variance differences shrink the most
 - Widths of CI's usually shrink a little too
 - The more you look at, the better you can model the prior dist.

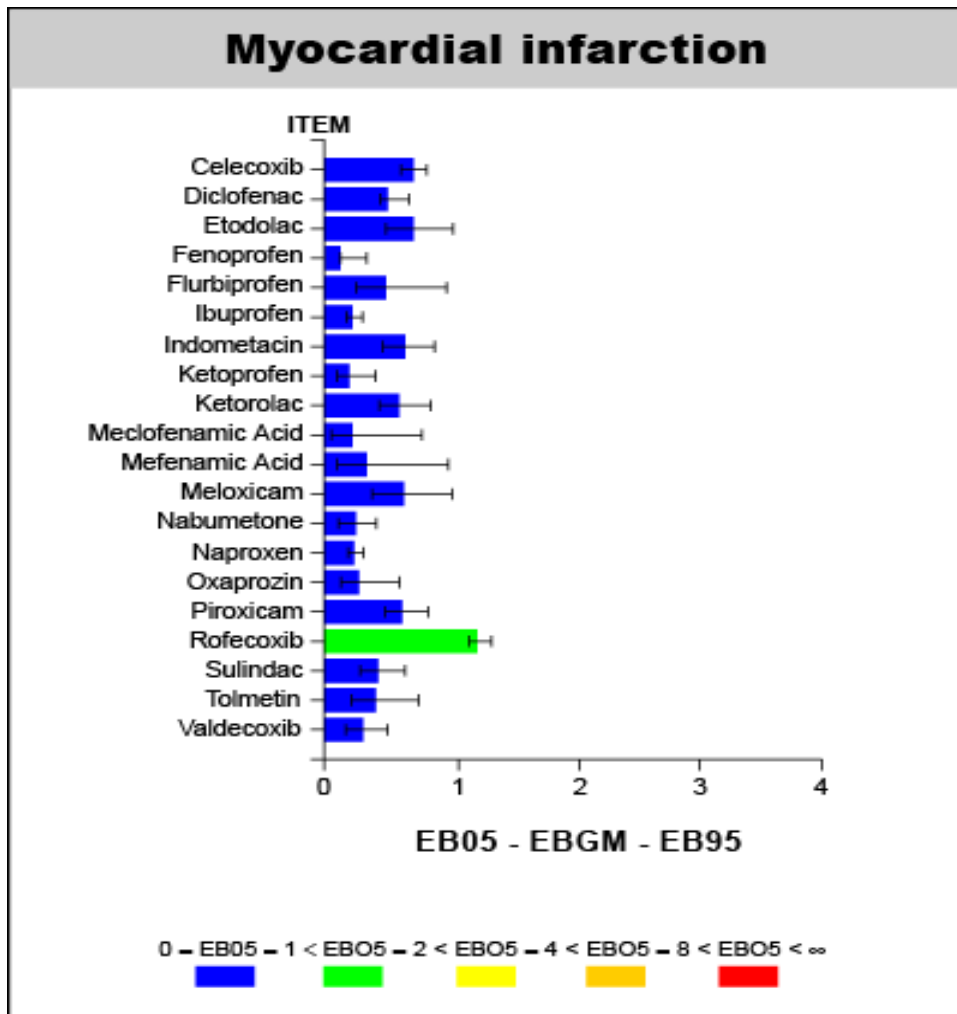
Empirical Bayes Gamma-Poisson Shrinker (GPS Method)

- Estimate $\lambda_{ij} = \mu_{ij}/E_{ij}$, where $N_{ij} \sim \text{Poisson}(\mu_{ij})$
- Assume Superpopulation Model for λ
 - Prior Distribution Is Mixture of 2 Gamma Distributions
 - Estimate the 5-Parameter Prior from All the (N_{ij}, E_{ij}) Pairs
- Posterior Distributions of each λ_{ij} Are Used to Create “Shrinkage” Estimates
 - EBGM = Empirical Bayes Geometric Mean of Posterior Dist.
 - Estimate of μ_{ij}/E_{ij} Has Smaller Variance than N_{ij}/E_{ij}
 - Rank Cells by $EB05_{ij}$ = Lower 5% Point of Posterior Dist.
 - More “Interesting” than Ranking Cells Based on “P-Values”
 - Compare $(N = 10, E = 0.1)$ to $(N = 2000, E = 1000)$

Plot of Classical Estimate with Conf. Int. and Bayesian "Shrinkage" Estimates [O]



Comparisons of NSAIDs in AERS



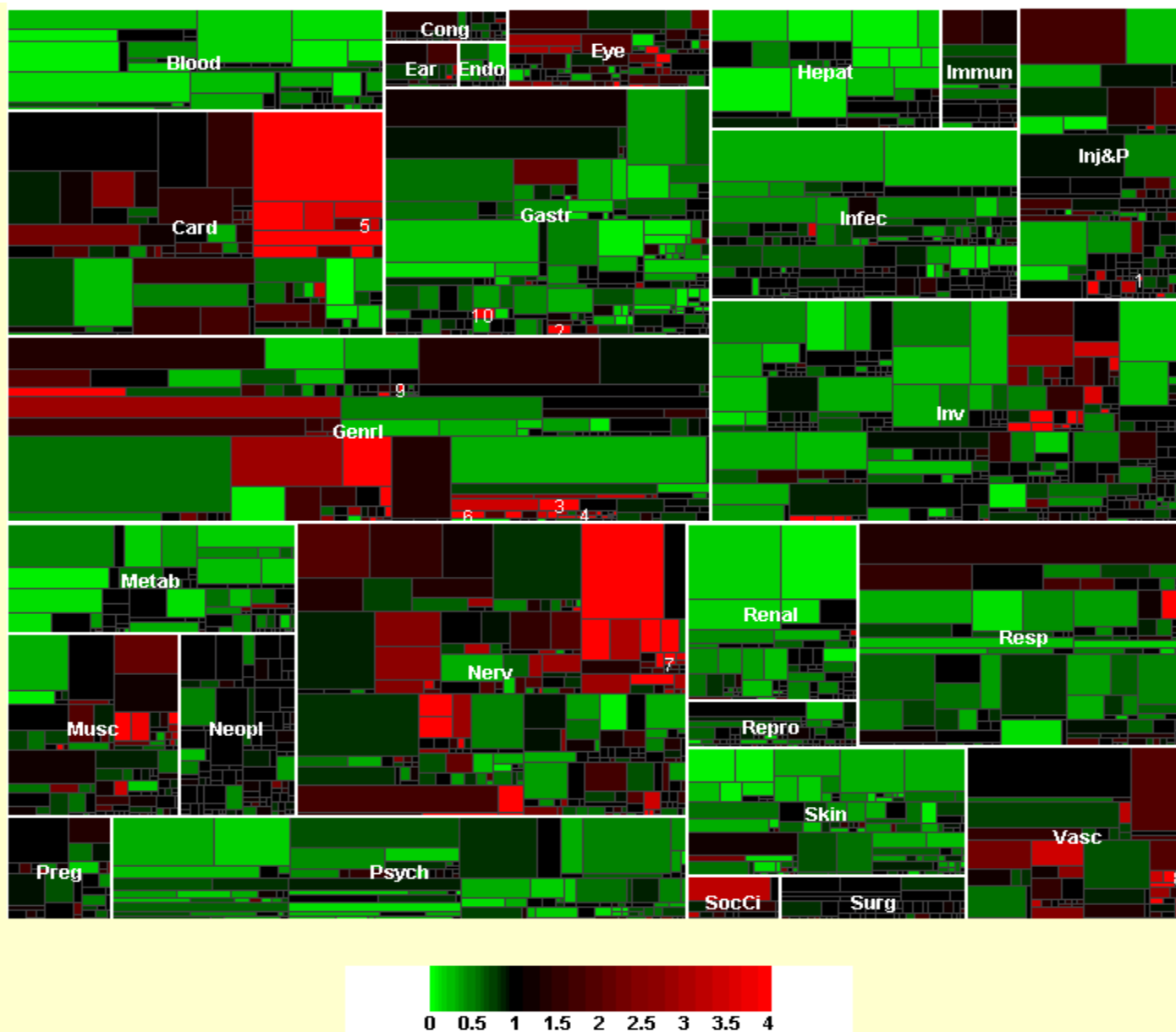
AERS to 3Q03 (Suspect drugs)

Blood pressure increased

Association	[1968-85] - [1999]	[1968-85] - [2000]	[1968-85] - [2001]	[1968-85] - [2002]	[1968-85] - [2003]	[1968-85] - [2004]	[1968-85] - [2005]
Rofecoxib	1.385	4.716	3.786	3.362	3.5	3.304	3.148
Valdecoxib				1.077	1.367	1.457	1.662
Celecoxib	0.416	0.706	0.899	0.964	1.034	1.113	1.177

$0 \leq EB05 \leq 1$ < $EB05 \leq 2$ < $EB05 \leq 4$ < $EB05 \leq 8$ < $EB05 < \infty$

Heat Map Profiling Spontaneous Reports for a Drug



Spontaneous Report 2x2 Analyses: Two Biases

■ Masking—Inappropriate Comparator Drugs

- “Other Drugs” often include ones that cause event of interest
 - Estimates for drug of interest will be biased downwards
- Need simultaneous estimates for all drugs w/ high associations

■ Confounding—Bias Due to Polypharmacy

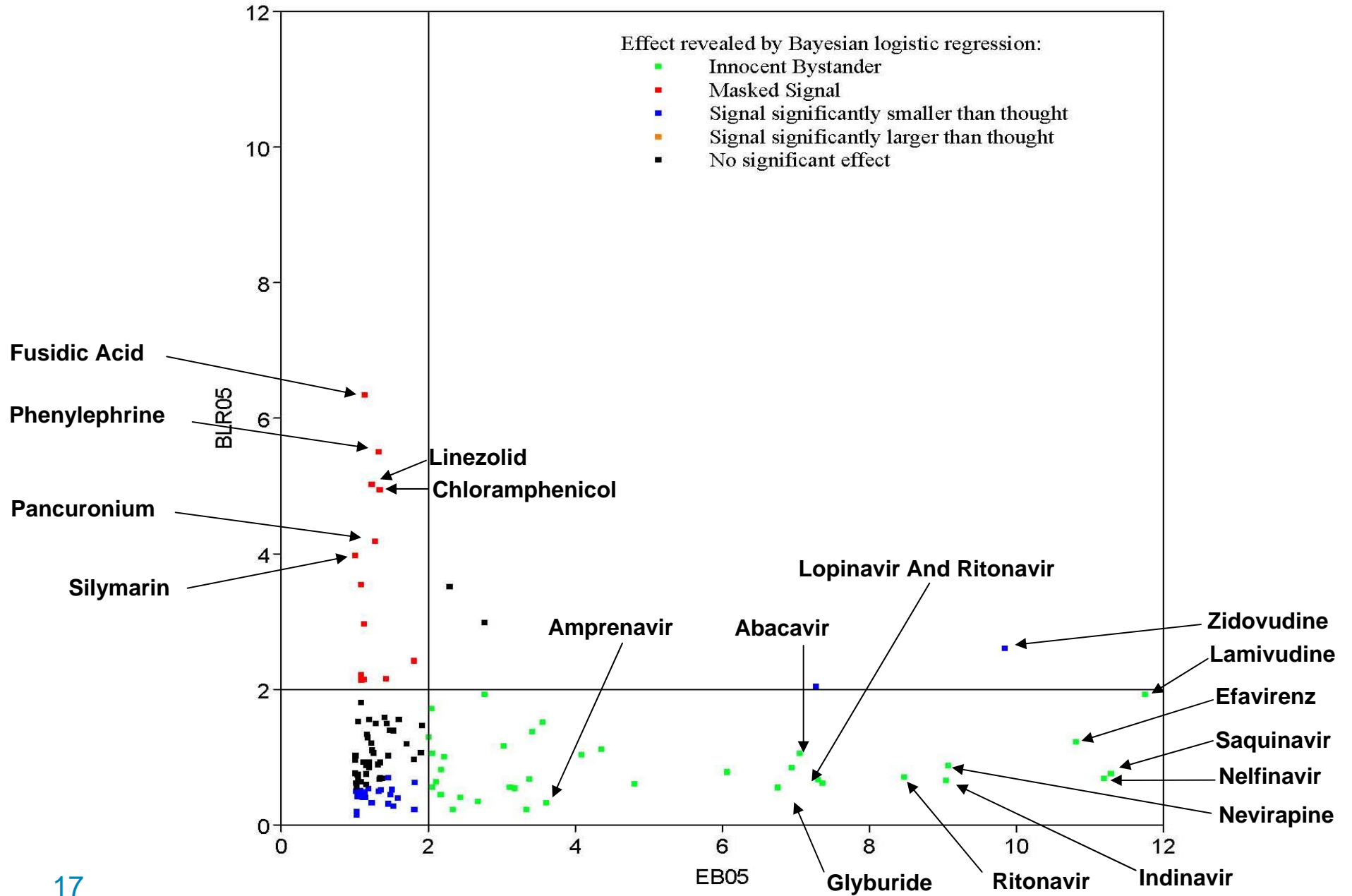
- Co-prescribed drugs partially inherit each other’s associations
- Synonymous terms: Signal leakage, Innocent bystander effect
- GPS, PRR and similar methods don’t account for effect of Drug-Drug associations on Drug-Event associations

■ Need a multivariate methodology

Logistic Regression

- Focus on specific events and drugs
 - Occurrence of event or problem in the report is the dependent var.
 - Presence/absence of potential causal factors are primary predictors
- Add covariates to the model as additional predictors
 - Dummy variables for age, gender, report year, etc.
- Add frequent concomitant drugs as more predictors
 - E.g., Drugs corresponding to top 100 Counts for response event
- Fit regression and convert to Odds Ratios and conf. limits
 - Non-overlapping confidence intervals worth investigating
 - Note patterns of agreement across events

PT = Lactic acidosis



Some References

2003	Almenoff, J.S., DuMouchel, W., Kindman, L.A., Yang, X., Fram, D.	Disproportionality analysis using empirical Bayes data mining: a tool for the evaluation of drug interactions in the post-marketing setting	Pharmacoepidemiology and Drug Safety 2003; 12: 517-521
2006	Almenoff, J.S., LaCroix, K.K., Yuen, N.A, Fram, D., DuMouchel, W.	Comparative Performance of Two Quantitative Safety Signalling Methods - Implications for Use in a Pharmacovigilance Department	Drug Safety 2006; 29 (10): 875-887
2005	Banks, D., Woo, E.J., Burwen, D.R., Perucci, P., Braun, M.M., Ball, R.	Comparing data mining methods on the VAERS database	Pharmacoepidemiology and Drug Safety 2005; 14: 601-609
2004	Kubota, K., Koide, D., Hirai, T.	Comparison of data mining methodologies using Japanese spontaneous reports	Pharmacoepidemiology and Drug Safety 2004; 13: 387-394
2000	Puijenbroek, E.P. van, Egberts, A.C.G., Heerdink, E.R., Leufkens, H.G.M.	Detecting drug-drug interactions using a database for spontaneous adverse drug reactions: an example with diuretics and non-steroidal anti-inflammatory drugs	Eur. J. Clin. Pharmacol. 2000; 56: 733-738
2002	Puijenbroek, E.P. van, Bate, A., Leufkens, H.G.M., Lindquist, M., Orre, R., Egberts, A.C.G.	A comparison of measures of disproportionality for signal detection in spontaneous reporting systems for adverse drug reactions	Pharmacoepidemiology and Drug Safety 2002; 11: 3-10
2005	Rolka, H., Bracy, D., Russell, C., Fram, D., Ball, R.	Using simulation to assess the sensitivity and specificity of a signal detection tool for multidimensional public health surveillance data	Statistics in Medicine 2005; 24: 551-562
2003	Roux, E., Thiessard, F., Fourrier, A., Bégaud, B., Tubert-Bitter, P.	Spontaneous Reporting System Modelling for Data Mining Methods Evaluation in Pharmacovigilance	Presented at the Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP) Workshop, AMIE-2003 Conference, Cyprus, October 2003
2006	Solomon, R., DuMouchel, W.	Contrast Media and Nephropathy: Findings from Systematic Analysis and Food and Drug Administration Reports of Adverse Effects	Investigative Radiology 2006; 41(8): 651-660
2002	Szarfman, A., Machado, S.G., O'Neill, R.	Use of Screening Algorithms and Computer Systems to Efficiently Signal Higher-Than-Expected Combinations of Drugs and Events in the US FDA's Spontaneous Reports Database	Drug Safety 2002; 25 (6): 381-392

Summary

■ Spontaneous Report Data Statistical Issues

- Extensive data cleaning necessary for millions of records
- Many non causal reasons for associations
- Poor design compared to clinical trial or cohort data
- Interpretation of comparator group is difficult
- Multiple comparison and *post-hoc* fallacies are endemic

■ But Systematic Analyses Can Be Fruitful

- About the only way to learn about very rare ADRs
- Hypothesis generation and/or a second data source for comparisons
- Bayesian approach to multiple comparisons helps assessment
- Computer tools essential for improved productivity
- Signal management (structured workflow) enables institutional “memory”