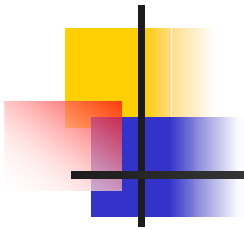
Bayesian Measurement of Associations in Adverse Drug Reaction Databases

William DuMouchel
Shannon Laboratory, AT&T Labs – Research
dumouchel@research.att.com

Joint Statistical Meetings, New York City, 13 August 2002



Data Mining of Spontaneous ADR Reports

- Databases of Adverse Drug Reaction Reports
  - Objectives and Limitations
- Drug – Event Counts as a Two-Way Table
  - Empirical Bayes Compared to Other Approaches
- Generalization to Data Mining *Market Basket Problem*
  - Models for Item Sets with 3 or More Items
- Guilty and Innocent Bystanders
  - Adjusting Drug-ADR Associations for Drug-Drug Associations
- Monitoring for Change over Time
  - Kalman Filter Model for Event Frequencies in Databases
- Discussion and Conclusion

Databases of Adverse Drug Reactions

- FDA Spontaneous Report System (SRS)
  - Post-Marketing Surveillance of all Drugs since 1969
  - Data in the Public Domain, Available from FDA
- FDA Adverse Event Reporting System (AERS)
  - Replaced SRS in 1997 with New AE Coding System
    - COSTART vs. MEDRA
- FDA/CDC Vaccine Adverse Events (VAERS)
  - Stricter Laws for Vaccine Adverse Event Reporting
- Other Databases for Medical Devices, etc.
- World Health Organization Collects Similar Data across Countries

Objectives and Limitations of Analysis

- 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
    - We 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
  - Substantial Under-Reporting to the FDA
  - No Certainty that a Reported Reaction Was Causal
  - Differential Report Rates of Adverse Events by Drug

Finding "Interestingly Large" Cell Counts in a Massive Frequency Table

- Large Two-Way Table with Possibly Millions of Cells
  - 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)
  - Analyzed SRS Database with 1398 Drugs and 952 AE Codes
  - N_{ij} = Count of Reports Containing Drug i and Event j
  - Only 386K out of 1331K Cells Have $N_{ij} > 0$
  - 174 Drug-Event Combinations Have $N_{ij} > 1000$
- Naïve Baseline Frequencies $E_{ij} = N_{i.} N_{.j} / N_{..}$
  - Extension to Stratification: Sum Independence Frequencies Defined Separately over Strata Based on Age, Sex, etc.

Empirical Bayes **G**amma-**P**oisson **S**hrinker

- Estimate $\lambda_{ij} = \mu_{ij}/E_{ij}$, where $N_{ij} \sim$ Poisson(μ_{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 = 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 ($N = 2000$, $E = 1000$)
- GPS Software Available ftp://ftp.research.att.com/dist/gps/
  - ML and EB Estimation, with Excel-Compatible Input/Output

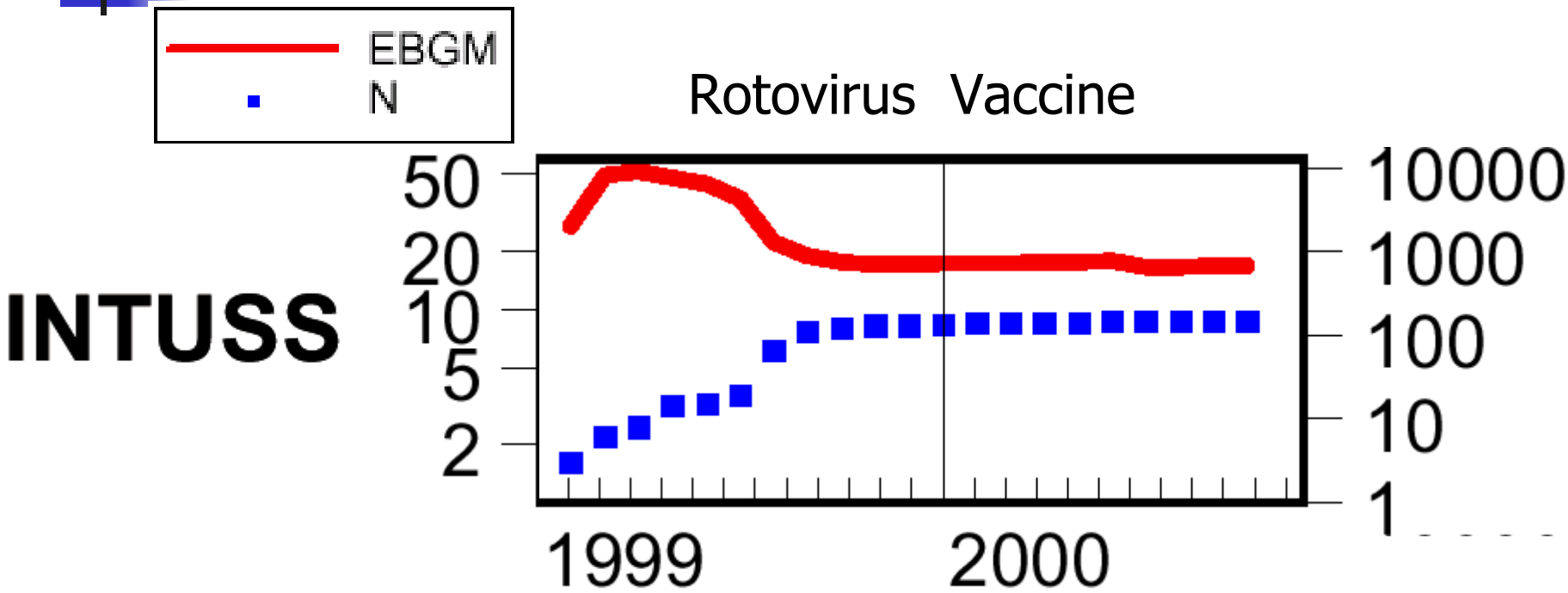
Alternative: Proportional Reporting Ratio

- Each Cell of Drug x Event Table Defines a 2 x 2 Table
  - Evans (Pharmacoepi. Drug Safety 10: 483-96, 2001)
  - Pool Counts Over All Other Drugs and All Other Events
  - $PRR_{ij} = [a_{ij}/(a_{ij} + b_{ij})] / [c_{ij}/(c_{ij} + d_{ij})]$
  - Reduce Variance by Requiring $N_{ij}=a_{ij}>2$ and $\chi^2>4$
- For N>20 or so, N/E = EBGM = PRR to a few percent
  - PRR Could Adjust for Stratification, but None Published
  - EB05, EB95 Provides Confidence Limits Not Available for PRR
  - EBGM and EB05 Available and Reliable for N = 1 or 2
  - Shrinkage Estimation Smoothing Provides Elegant Transition from N = 1 to Large N
  - Generalization: MGPS for Triples & Higher-Order Associations

Alternative: BCPNN

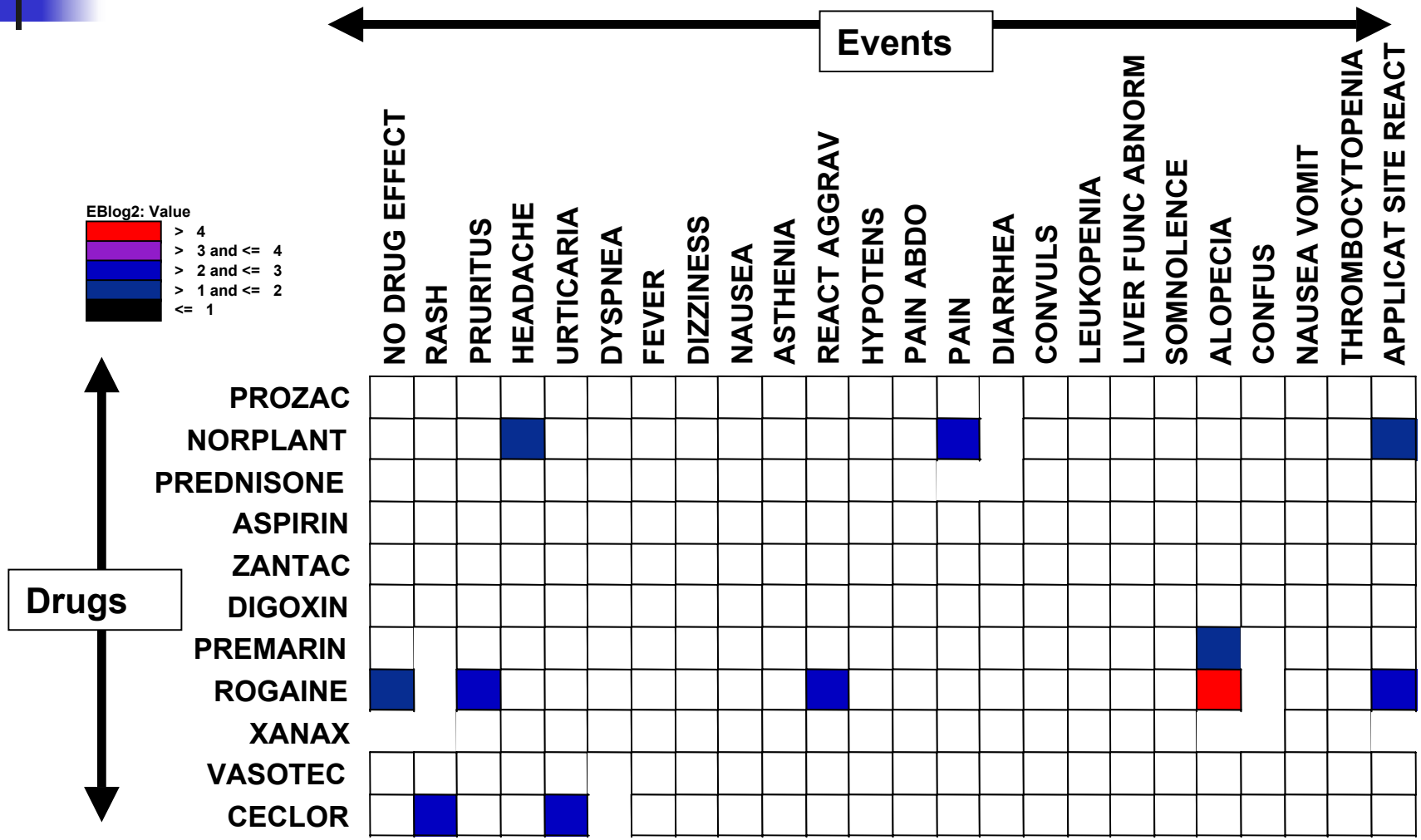
- Bayesian Confidence Propagation by Neural Network
  - Orre et al (Comput Stat Data Anal 3: 473-93, 2000)
  - Bayesian Shrinkage Model Based on Multinomial, not Poisson
  - Uses 2x2 Tables Based on Counting Reports, not Combinations
  - Computes Posterior Mean and Variance of IC = $\log_2(\lambda)$
    - Signal Score IC – $2*\sqrt{V}$ Similar in Concept to EB05
  - Bayesian Prior is Fixed in Advance, Not Estimated from Data
    - Results Very Similar to MGPS with Exponential Prior Dist.
- For N>20 or so, N/E = EBGM = 2^{IC}
  - Adjustment for Stratification Vars Not Available in BCPNN
  - Confidence Limits EB05 Do Not Depend on Normal Approx.
  - MGPS Generalization to Triples, etc., Better Developed

Example of Large Signal with Small N



- The RV Vaccine Was Used in U.S. in 1998-99 and Was Withdrawn from the Market when the Association with Intussusception, a Severe GI Condition, Was Confirmed.

Spatial Map Showing the Signal Scores (SS) for the Most Frequently Reported Events (columns) and Drugs (rows) by the Intensity of the SS (color) [Prepared by Ana Szarfman]



Multi-item Gamma Poisson Shrinker

- Extend GPS to Analyze Arbitrary Itemset Frequencies
  - E.g. Drug-Drug-Event, Drug-Event-Event, 4-tuples, etc.
  - "Market Basket Problem" in Data Mining Literature
  - Computational Challenge—Huge No. of Possible Itemsets
- EB Model Same as GPS—Baseline Freqs. E Change
  - P_{si} = Prop. of Stratum s Reports with Item i (Drug or Event)
  - P_{si} Small, but $\Sigma_i P_{si}$ (= Expected # Items/Report) > 1
  - For Triples, $E_{ijk} = \Sigma_s n_s P_{si} P_{sj} P_{sk}$ (n_s: #Reports in Stratum s)
  - Condition on $N_{ijk} \geq n^*$ to Reduce Counting and EB Calculations
    - We Choose Smaller n^* than in Market Basket Literature
  - Interpretation of EBGM & EB05 Same as for GPS
- MGPS Extensions: Different Definitions of Baseline
  - Compare 2 Populations: $F_{ijk} = E_{ijk}$ * (EBGM from Elsewhere)

Multi-Item Associations vs. Pairwise Associations

- Suppose Itemset (Drug A, Drug B, C = Kidney Failure) Is Unusually Frequent
  - Are merely the Pairs AB, AC, BC Frequent, or Does AB Cause C (Drug Interaction)
- Comparison of EB Estimate to the Predictions of All-2-Factor Interaction Log-Linear Model
  - $EBGMdiff = EBGM - E_{All2F}/E$
    - E is the Expected Count from Independence
    - Compute E_{All2F} with Shrinkage Estimates of Pairwise Counts
- Alternate Model: Define $\lambda = \mu / E_{All2F}$ and Shrink Counts toward the All-2-Factor Model Directly
  - In MGPS, define Baseline as E_{All2F}
  - Resulting $EBGM > 1$ Indicates Possible 3-Factor Interaction

Guilty and Innocent Bystanders

- GPS, PRR and Similar Methods Don't Account for Effect of Drug-Drug Assocs. on Drug-Event Assocs.
  - Toy Example: DI=Drug of Interest, GB=Guilty Bystander Drug IB=Innocent Bystander Drug, AE=Adverse Event [All 0-1 Vars]
  - P(DI=1)=.5, P(GB=1|DI)=.75-DI/2, P(IB=1|DI)=.25+DI/2, P(AE=1|DI,GB,IB) = .25+(DI+GB)/4

All 16 Jt. Probs

Prob x 128	DI=0				DI=1			
	IB=0		IB=1		IB=0		IB=1	
	GB=0	GB=1	GB=0	GB=1	GB=0	GB=1	GB=0	GB=1
AE=0	9	18	3	6	6	1	18	3
AE=1	3	18	1	6	6	3	18	9

Note Bias in Odds Ratios

Prob x128	GB=0	GB=1	IB=0	IB=1	DI=0	DI=1	GB=0		GB=1	
							DI=0	DI=1	DI=0	DI=1
AE=0	36	28	34	30	36	28	12	24	24	4
AE=1	28	36	30	34	28	36	4	24	24	12
OR	1.65		1.28		1.65		3		3	

Detecting Bystander Bias

- Loglinear Models or Logistic Regression
  - Note this Bias Is Distinct from 3-Factor Interaction
  - All-2-Factor Model Can Detect Bystander Bias
  - Practical Limit of About 25 Items to Fit All-2-Factor Model
  - Logistic Regression of Each AE on a Few Hundred Drugs Might Be Feasible
- Example: Drugs for Type 2 Diabetes/Hprtn/Hi Chol. in AERS (1997-2001)
  - LACTIC.ACIDOSIS [*OR* = Odds Ratios]

	N	*E*	*N/E*	*OR.1*	*OR.9*	*tstat*
ATORVASTATIN	39	54.8	0.7	0.7	0.3	-6.4
ENALAPRIL	39	24.0	1.6	1.6	0.9	-0.9
FUROSEMIDE	148	69.6	2.1	2.3	1.5	4.4
GLIPIZIDE	78	21.7	3.6	3.8	0.5	-5.9
HYDROCHLOROTHIAZIDE	20	20.4	1.0	1.0	0.6	-2.2
LISINOPRIL	62	36.3	1.7	1.8	0.7	-2.2
METFORMIN	685	31.7	21.6	44.9	56.4	71.1
PIOGLITAZONE	9	10.1	0.9	0.9	0.2	-5.5
PRAVASTATIN	11	16.2	0.7	0.7	0.4	-3.1

  - OR.1: Logistic Regression on 1 Drug + 162 Strata; OR.9: Use all 9 Drugs

Screening for Bystander Effects

- Generic Search with No Prior Specification of Hypotheses
- Naïve Bayes Model Using Drug1-Drug2-AE Triples
  - DI, AE, $\{D_j, j = 1, \ldots, J\}$ (D_j: Potentially Confounding Drugs)
  - Assume $P(\{D_1,\ldots,D_J\}|DI, AE) = \Pi_j P(D_j \mid DI, AE)$, then:
  - $OR(DI,AE|D_1=0,\ldots,D_J=0) = OR(DI,AE)\ \Pi_j[OR(DI,AE|D_j=0)/OR(DI,AE)]$
  - $EBGM(DI,AE|D_1=0,\ldots,D_J=0) \approx EBGM(DI,AE)\ \Pi_j[EBGM(DI,AE|D_j=0)/EBGM(DI,AE)]$
- For each DI-D_j-AE Triple, Compare DI-AE Overall and w/ $D_j=0$
  - Product of Ratios Above Is "Bystander Bias Adjustment Factor"
  - Interpreted as Extrapolating to Situation w/ No Concomitant Drugs
    - Sensitive to DI – D_J Drug Interactions as well as Confounding Effects
  - Repeat this Analysis for ALL Combinations of DI-AE
    - Take Most Frequent 548 Drugs and 688 AEs from Post-1997 AERS: 177,020 Observed Drug-AE Pairs, Potentially 103M Drug1-Drug2-AE Triples
    - Example with Restriction to 691,722 D1-D2-AE Triples Appearing in 5+ Reports
    - Frequent-Triple Restriction Reduces Interpretability of Bias Adjustment Factor
    - Assume Restricted Factor Is Useful as a Relative Indicator of Bystander Bias

Glipizide – Lactic Acidosis Revisited

	DRUG	Adverse.Event	N	E	EBGM	#CONCOM.DRUGS	logBias	adjEBGM
0	GLIPIZIDE	Lactic Acidosis	78	21.74	3.40	20	-4.18	0.052

	CONCOMITANT	N.Triple	E.Triple	NwoCONCOM	EwoCONCOM	EBGMwoCONC	EBGMratio
1	AMLODIPINE	8	0.97	70	20.77	3.17	0.933
2	ASPIRIN	12	1.70	66	20.04	3.09	0.910
3	BENAZEPRIL	5	0.11	73	21.62	3.18	0.936
4	DIGOXIN	10	0.69	68	21.04	3.04	0.895
5	FUROSEMIDE	18	1.11	60	20.63	2.73	0.804
6	GEMFIBROZIL	6	0.19	72	21.55	3.15	0.928
7	INSULIN	5	0.85	73	20.89	3.29	0.969
8	ISOSORBIDE	10	0.39	68	21.35	3.00	0.883
9	LEVOTHYROXINE	10	0.92	68	20.82	3.07	0.904
10	LISINOPRIL	8	0.60	70	21.13	3.12	0.919
11	METFORMIN	74	0.51	4	21.23	0.23	0.068
12	METOPROLOL	5	0.58	73	21.16	3.25	0.957
13	NIFEDIPINE	5	0.39	73	21.34	3.23	0.951
14	PAROXETINE	5	0.39	73	21.35	3.22	0.948
15	QUINAPRIL	5	0.16	73	21.57	3.19	0.939
16	RANITIDINE	5	0.51	73	21.23	3.24	0.954
17	SIMVASTATIN	9	0.63	69	21.10	3.08	0.907
18	VITAMIN	5	0.81	73	20.93	3.29	0.969
19	VITAMIN_D	5	0.09	73	21.65	3.18	0.936
20	WARFARIN	7	0.76	71	20.98	3.19	0.939

logBias = sum(log(*EBGMratio*))

More Results from Naïve Bayes Model

- Largest 15 Bias Adjustments for Drug-AE Pairs Having EBGM>10, N>100

	DRUG	Adverse.Event	N	E	EBGM	#ConcDrugs	logBias
2	METAMIZOLE	Blister	116	1.8	60.3	107	-24.1
8	VANCOMYCIN	Blister	149	14.0	10.5	104	-18.1
15	DEXTROAMPHETAMINE	Cerebrovascular Accident Nos	110	7.8	13.9	6	-8.9
14	DEXTROAMPHETAMINE	Injury Nos	122	8.9	13.4	9	-10.1
9	AMPHOTERICIN B	Multi-Organ Failure	141	10.5	13.2	82	-13.5
11	DOPAMINE	Multi-Organ Failure	105	7.9	13.0	77	-12.2
13	VANCOMYCIN	Multi-Organ Failure	188	18.4	10.1	87	-10.8
12	DOPAMINE	Shock	111	10.8	10.2	87	-11.6
1	AMPHOTERICIN B	Stevens Johnson Syndrome	108	9.2	11.6	91	-27.1
3	METAMIZOLE	Stevens Johnson Syndrome	131	2.0	62.6	113	-23.7
6	VANCOMYCIN	Stevens Johnson Syndrome	154	14.9	10.2	106	-19.4
4	METAMIZOLE	Toxic Epidermal Necrolysis	137	1.4	91.3	117	-22.1
5	CEFTAZIDIME	Toxic Epidermal Necrolysis	114	4.5	24.4	104	-19.8
7	VANCOMYCIN	Toxic Epidermal Necrolysis	171	10.5	16.0	120	-18.7
10	ANTIHYPERTENSIVE	Vulvovaginal Discomfort	112	10.6	10.4	6	-13.0

Investigate for Possible Interactions or Confounding with Indications or Other Drugs

Monitoring for Change Over Time

- Suppose a Database of Reports Is Replaced Regularly
  - E.g. Examine all New Reports Every Month or Quarter
  - Millions of Event Frequencies Being Monitored for Change
    - Almost All Counts 0 or Small
    - Comparison to Independence Not an Issue, but Comparison to the Recent Past Is
    - May Want to Detect Significant Decreases as well as Increases
- KFGPS: Method to Smooth Event Count Time Series
  - Detect Which Ones Have Shown Sudden Frequency Shifts
  - Shrinkage Estimates Discount Poisson-Level Variations
    - Adaptation of Well Known Kalman Filter Methodology
    - Bayesian Estimates Allow Posterior Selection of Largest Shifts
  - Updating Scheme Requires Storage of Just Last Period Data
  - Baseline Frequency this Period Is Posterior Estimate from Last

Future Work

- Graphical Exploration of the Thousands of Empirical Bayes Estimates Generated
- Use of Demographic Variables as Items
  - Associations with Dummy Variables for Age, Sex, etc.
  - Compare with Stratification by such Variables
- Analysis of other Types of Clinical Databases
  - Adverse Events from Collections of Clinical Trials
  - Associations from HMO-style Databases

Preliminary Work: Insurance Claims Data

- MarketScan 1998 Database from MEDSTAT Group (Thomson Corporation)
  - Longitudinal Histories of Inpatient, Outpatient and Prescription Drug Experience for Millions of Covered Lives
    - Private, Medicare and Medicaid Eligible Individuals
- Goal: Use MGPS to Detect ADR Associations
- Challenge: Vast Majority of Drug-Diagnosis Signals Relate Drugs to Primary Symptoms and Co-Morbidities of Diseases They Are Intended to Treat
  - Eliminate ICD9 Codes w/ No Corresponding MEDRA Term
  - Temporal Information: Symptom Occurs *After* Drug Prescription
- Limited success: Sensitivity Seems Good but many False Positives from Drug Indications

MGPS Model and Algorithm Seem to Perform Well on the Association Problem

- Estimate Interestingness Measure: Frequency Ratio vs. Independence or any other Baseline Model
- Empirical Bayes Shrinkage for Bias-Variance Tradeoff
- Reliable Estimation for much Lower Values of N than Previous Market Basket Literature
- Use of All-Two-Factor Log-Linear Model Allows Sophisticated Analyses of Larger Item Sets
- Ongoing Use and Validation by FDA Researchers
- Detection of Drug-Drug Confounding and Interactions
  - Logistic Regression and Naïve Bayes Models Are Useful
- Time Series Kalman Filter Model for Event Frequencies
  - State Space Model Provides Efficient Summary of Past History
  - Incorporates Separate Model for Analysis of First-Time Event Counts

References and Acknowledgements

- References to EB Measurement of Associations
  - DuMouchel W, Friedman C, Hripcsak G, Johnson S, Clayton P (1996) Two applications of statistical modeling to natural language processing. *AI and Statistics* V, ch. 39, edited by D. Fisher and H. Lenz, Springer-Verlag.
  - DuMouchel W (1999) Bayesian data mining in large frequency tables, with an application to the FDA Spontaneous Reporting System (with discussion), *The American Statistician*, 53:177-202.
  - DuMouchel W, Pregibon D (2001) Empirical Bayes Screening for Multi-Item Associations *Proc. ACMSIGKDD 2001 Intl. Conf. on Knowledge Discovery from Data.* [Best Applications Paper Award]
- Acknowledgement and Thanks for Data, Comments and Suggestions
  - Ana Szarfman, CDER, US FDA
  - David Fram, Lincoln Technologies
  - June Almenoff, GlaxoSmithKline