

**Ana Szarfman, MD, PhD, Medical Officer, QMR, OB,
OPaSS, CDER, FDA, szarfman@cder.fda.gov,
301-827-3209**

Syndromic Surveillance and Risk Management using MGPS

**National Syndromic Surveillance Conference
The New York Academy of Medicine
September 23-24th, 2002**

Overview

- **Explanation for non-statisticians**
- **New Multiple Item Empirical Bayesian Gamma-Poisson Shrinker (MGPS) Program**
- **Number of Signals**
- **Applications**
- **Validation**
- **Some additional characteristics**
- **Public access**
- **URLs to learn more about our work**
- **References**

What does safety data mining do?

- **Automates the systematic detection of 'higher than expected' signals without using external exposure data or adverse event background information**
- **Signal scores (adjusted O/E) are derived from application of a statistical model**
- **Operates on a database that is known to contain considerable noise (misreporting, duplications, coding errors)**
- **Uses stratification to control for potential confounding**
- **Allows comparison with other drugs of a similar class or used for a similar indication and with other events for a drug**
- **The algorithm generates reliable signals without linkage to external medical information on adverse drug reactions**
- **Provides an objective and systematic view of the data, reminding reviewers of what they already know and alerting them to critically important new safety signals**

What does safety data mining do? (cont)

- Identifies signs of potential *drug-drug interactions* and broadens the understanding of *signals of drug-associated syndromes* with existing and new drugs quickly and at minimal cost
- Alerts reviewers from FDA and the pharmaceutical industry early to potential problems with new and existing drugs, including complications when prescribing multiple drugs to a single patient, or when prescribing certain drugs to specific population subgroups, for example, the elderly, or women, or people who already present medical problems
- Points out the need for additional controlled studies of certain drugs, used either as monotherapy or in conjunction with other drugs
- The system does not automatically hide already-known associations from raw data mining results so until this function is developed reviewers have to filter these known events manually
- Adding context information will broaden the understanding of signals by by drug class, indication, chemical structure and function, and clinical pharmacology information

New Screening Algorithms

(written by William DuMouchel from AT&T Labs, Ana Szarfman, principal investigator)

- **The GPS (Empirical Bayesian Gamma-Poisson Shrinker) program**
 - Computes signal scores for drug-event pairs by event, drug, gender, age, such as pediatrics and geriatrics
- **The MGPS (Multiple GPS) program**
 - Computes signal scores for pairs, and higher-order (e.g., triplet, quadruplet) combinations of drugs and events that are significantly more frequent than their pairwise associations would predict
 - Adjusts for the multiplicity of drugs and events per record, an important feature to have for the post October 1997 data
 - It replaced GPS

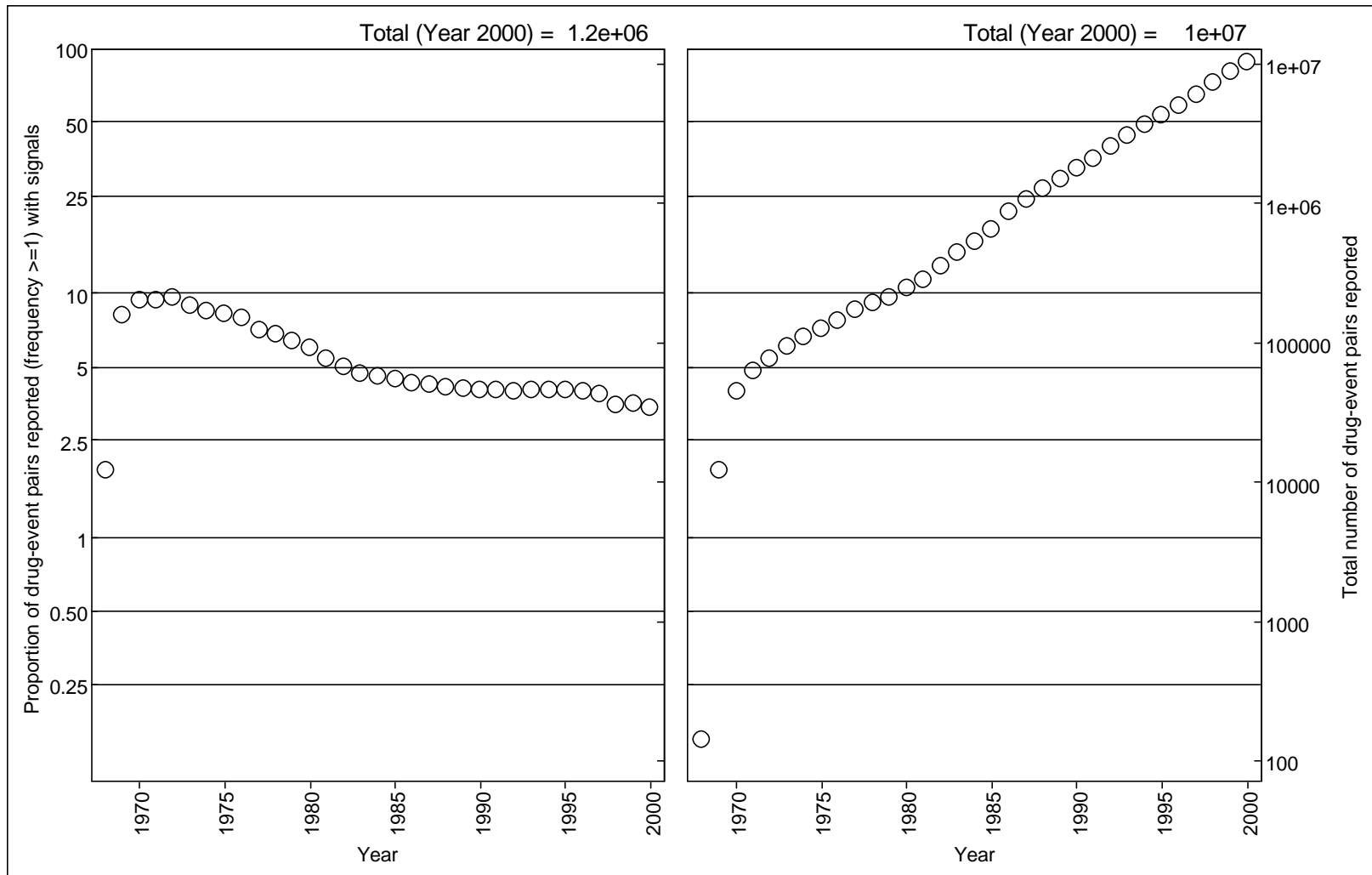
Input to system: the AERS MedWatch database

- **Over 8,000 drugs: *Representing all drugs and biologics in the US market***
- **Over 7,000 Event Codes**
- **Over 2 million reports**
- **35 years of data stored in a standardized structure**
- **Over 56 million possible drug-event pairs**

Number of Drug-Event Pairs with Signals

- In the current data, only 3.4% of all 1.2 million drug-event pairs ever reported (with frequencies ≥ 1) are signaled (next page, left panel)
- The total number of drug-event pairs reported increased steadily from 5,000 in 1969 to 10.4 million in 2000 (next page, right panel)
- More drugs have been added, and the opportunity of a drug-event pair to exist has increased
- In 2000, the total frequency count that contributed to the signals comprises 23% (2.4 million) of the total 10.4 million drug-event associations reported (not shown)
- A small proportion of signals (3.4%) captures a high proportion of the total number of drug-event associations reported (23%)
- A clinically meaningful individual event typically corresponds to multiple Event Codes (between 5-10). The signals include known adverse reactions and conditions being treated coded as events. This decreases the number of potential signals even further

Progression of the proportion of drug-event pairs with reported frequency (≥ 1) with signals (EB05 ≥ 2) (y1-axis) and of the total number of drug-event pairs reported (y2-axis) by year (x-axis) over the period 1978-2000



MGPS: Multiple Item GPS program

- **Definitions**

- **Shrinkage towards assuming independence between all items**
- **EB05 is the lower 5% bound for the ratio of adjusted O/E**
- **Both Excess and Excess2 are counts of reports instead of ratios of O/E**
- **Excess is the $E * (EB05-1)$**
- **Excess2 is the estimated number of reports unexplained by pair-wise occurrences**
- **$EBGMDiff=EBGM-E2/E$**

Data mining quadruple 'Event-Event-Event-Event' associations (syndromes) in single reports. Deaths: Top ranking EB05 quadruple event combinations having **rhabdomyolysis** showing the syndromes with only a small number of cases that cannot be explained by pair-wise interactions

		% Cutpoints:						
		Min.	0% - 10%	10% - 25%	25% - 75%	75% - 90%	90% - 100%	Max.
EB05		1276	1598	2029	8394	16086	53919	

Color-coded signal scores (EB05)

Numbers: first, number of reports, second, excess2

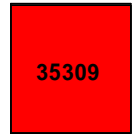
E_Blood Creatine Phosphokinase Increased/ E_Blood Myoglobin Increased/ E_Myoglobinuria Present/ E_Rhabdomyolysis



53919

5/ -1

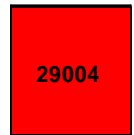
E_Blood Creatine Phosphokinase Increased/ E_Myoglobinuria Present/ E_Rhabdomyolysis/ E_Urine Discolouration



35309

4/ 0

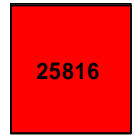
E_Blood Myoglobin Increased/ E_Condition Aggravated/ E_Myoglobinuria Present/ E_Rhabdomyolysis



29004

4/ 1

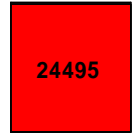
E_Blood Creatine Phosphokinase Increased/ E_Blood Myoglobin Increased/ E_Renal Failure Nos/ E_Rhabdomyolysis



25816

7/ 2

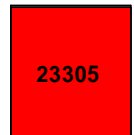
E_Blood Creatine Phosphokinase Increased/ E_Myoglobinuria Present/ E_Renal Failure Nos/ E_Rhabdomyolysis



24495

9/ 3

E_Blood Creatine Phosphokinase Increased/ E_Deceased Activity/ E_Myoglobinuria Present/ E_Rhabdomyolysis



23305

3/ 0

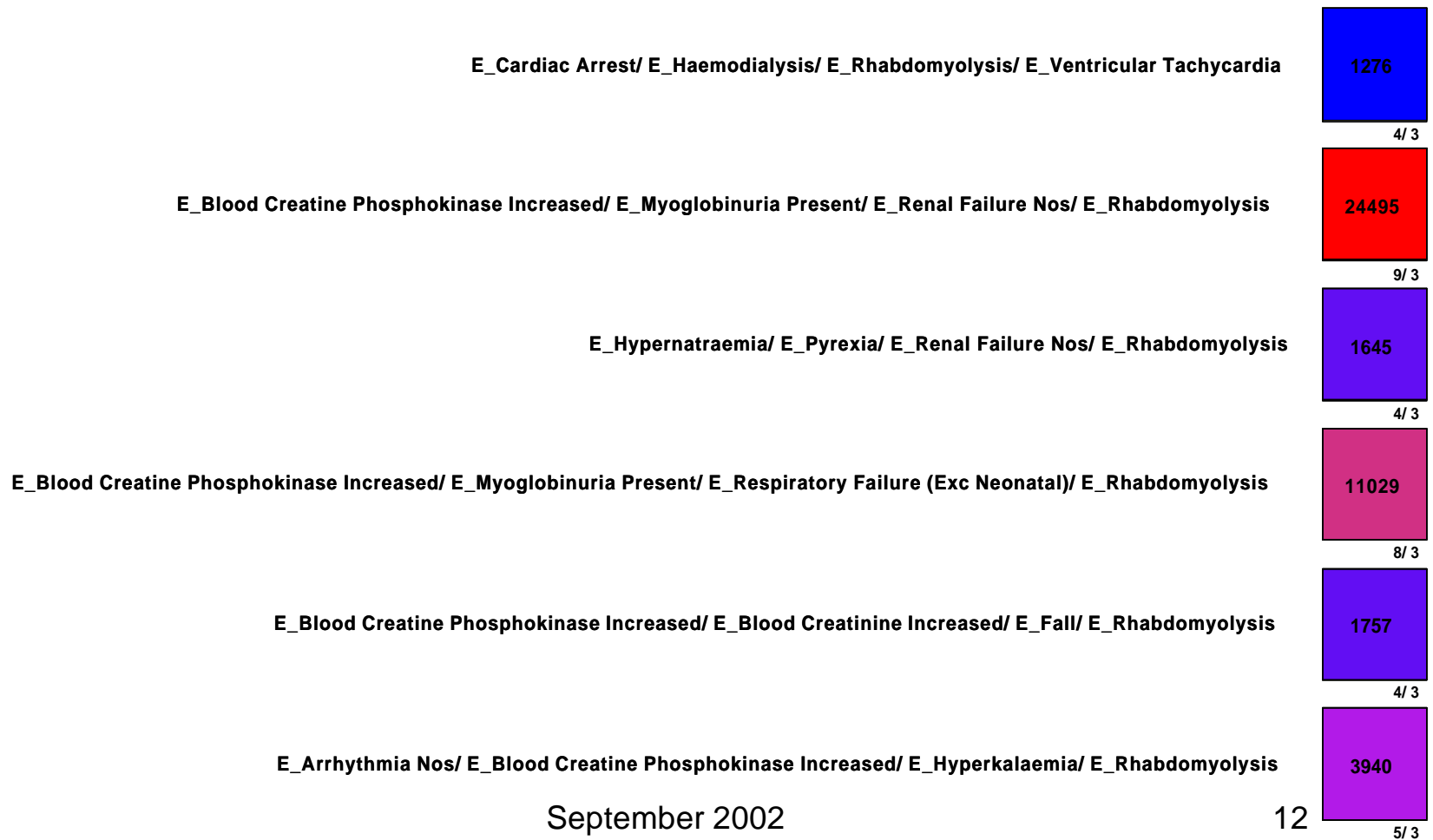
September 2002

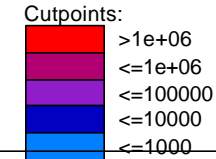
Same as in previous slide showing the rhabdomyolysis syndromes having a larger number of cases that cannot be explained by pair-wise interactions (top Excess2)

		% Cutpoints:						
		Min.	0% - 10%	10% - 25%	25% - 75%	75% - 90%	90% - 100%	Max.
EB05		1276	1598	2029	8394	16086	53919	

Color-coded signal scores (EB05)

Numbers: first, Number of reports, second, Excess2





Pairs, triples, and quadruples: EB05 for the corresponding components of pairs, triples, and quadruples associated with a quadruple having 'QT Prolongation' or 'Torsades de Pointes'

Pairs

PAIR / E_Electrocardiogram Qt Prolonged E_Fear, Focus Nec	216/ 183
PAIR / E_Electrocardiogram Qt Prolonged E_Ventricular Fibrillation	299/ 260
PAIR / E_Electrocardiogram Qt Prolonged E_Ventricular Tachycardia	362/ 320

TUPLES/CONC

Triples

TRIPLE / E_Electrocardiogram Qt Prolonged E_Fear, Focus Nec E_Ventricular Fibrillation	213/ 4
TRIPLE / E_Electrocardiogram Qt Prolonged E_Fear, Focus Nec E_Ventricular Tachycardia	205/ 3
TRIPLE / E_Electrocardiogram Qt Prolonged E_Ventricular Fibrillation E_Ventricular Tachycardia	236/ -52

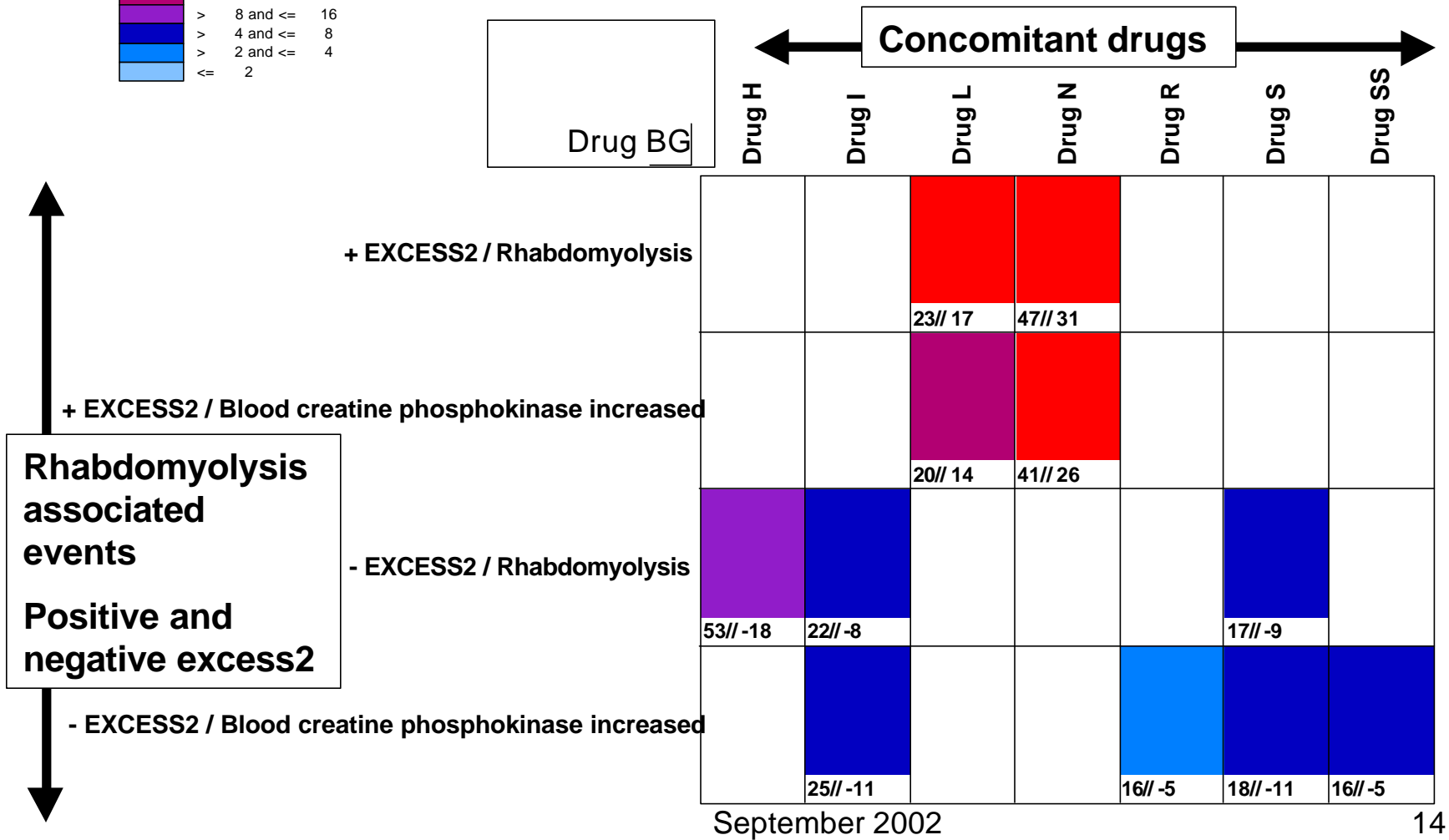
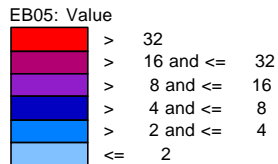
Quadruple

QUADRUPLE / E_Electrocardiogram Qt Prolonged E_Fear, Focus Nec E_Ventricular Fibrillation E_Ventricular Tachycardia	203/ 2
---	--------

SZARFMAN@CDER.FDA.GOV data JAN2001, page 1 - run dated 23APR2001, EB05 >2

Color-coded signal scores (EB05)
Numbers: first, Number of reports, second, Excess2

**Triplets: Synergic Drug-Drug Interactions with Drug BG.
Some signals cannot be explained by two-way interactions
(Positive Excess2)**



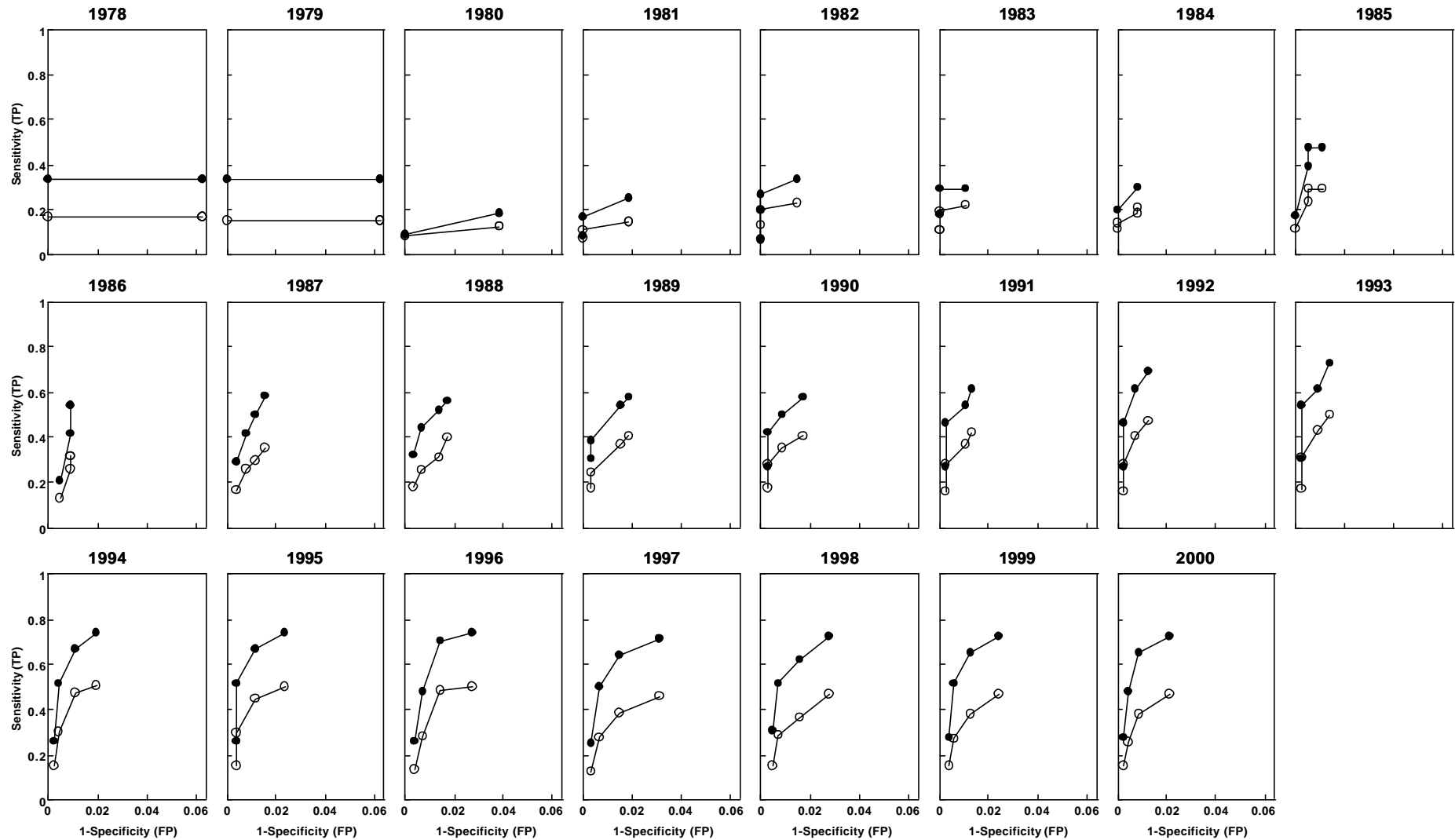
Process of validation

- **Data mining confirmed known positive and negative adverse event-drug associations across hundreds of drugs**
- **Detected many important signals not previously recognized**
- **Has shown that adverse events could have been signaled earlier for many drugs and in many drug classes, and for subpopulations, including *pediatric patients, and men and women***

Operating characteristics of MGPS

- **The sensitivity and specificity for tamoxifen, a drug well-characterized by recent large clinical trials were evaluated for 4 different EB05 cutoff points (1.5, 2, 4, and 8), each requiring a greater ratio of higher than expected to signal**
- **The information required in the current drug label was used as a gold standard of a true signal, across all years of marketing**

Tamoxifen: Receiver Operator curves for analysis of specificity/sensitivity across 23 years of marketing using EB05 cut-offs of 1.5, 2, 4, 8. Note the False Positive rate of <math><0.06</math> since marketing started



Labeled Events:
 ● Warning or Contraindication Section
 ○ Any Labeling Section

September 2002

Some additional characteristics

- **MGPS correctly identifies known signals and new problems and provides objective and systematic way of summarizing signals to help prioritize reviews and manage risk**
- **The system can be trained to recognize and suppress outputs with known disease related/labeled events**
- **Time sequences help detect publicity effect**
- **We have demonstrated that we can predict useful external denominators from “numerators”**

Public access

- **The FDA's post-marketing database: U.S. Department of Commerce National Technical Information Service (NTIS) <http://www.ntis.gov/>**
- **The C++ GPS program for calculating signals: <ftp://ftp.research.att.com/dist/gps>**
- **This access simplifies independent re-examination and validation and facilitates use by groups managing drug risks**

URLs to learn more about our work

- **Video Clips. Workshop on datamining with applications in genomics, clinical trials and post-marketing drug risk. Schering-Plough Workshop 2000-2001. Harvard School of Public Health. Available from URL: <http://www.biostat.harvard.edu/events/schering-plough/old/agenda2000-01.html>**
- **Web page prototype with over 11,000 drug profiles for dissemination of information to specific CDER and CBER reviewers has been populated**
- **DuMouchel W.: <ftp://ftp.research.att.com/dist/gps>**
- **Szarfman A. New methods for signal detection. <http://www.fda.gov/cder/present/ispe-1999/default.htm>**

Uncertainties in interpretation of FDA's post-marketing safety database

- No research protocol that can control for
 - Selection bias
 - Under-reporting
 - Reports enriched in response to publicity or “Dear Doctor” letters
 - Variable historical data
- Some of the data may be invalid: duplications, coding errors, poor quality of information

- Impossible to review the over 2 million records collected and the over 0.3 million new records per year using standard methods

- Impossible to “clean” all these records prior to applying algorithms

REFERENCES:

- Szarfman A, Machado SG, O'Neill RT. 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; 25:381-392
- FDA Talk paper. Bayer voluntarily withdraws baycol. FDA talk paper no. T01-34. 2001 Aug 8
- DuMouchel W, O'Neill RT, Szarfman A. Video Clips. Workshop on datamining with applications in genomics, clinical trials and post-marketing drug risk. Schering-Plough Workshop 2000-2001. Harvard School of Public Health. Available from URL: <http://www.biostat.harvard.edu/events/schering-plough/old/agenda2000-01.html>
- DuMouchel W, Pregibon D. Empirical bayes screening for multi-item associations. Proceedings of the conference on knowledge discovery and data; 2001 Aug 26-29; San Diego (CA): ACM Press: 67-76
- O'Neill RT, Szarfman A. Some FDA perspectives on data mining for pediatric safety assessment. Workshop on Adverse Drug Events in Pediatrics. *Curr Ther Res Clin Exp* 2001; 62:650-63
- Niu MT, Erwin DE, Braun MM. Data mining in the US vaccine adverse event reporting system (VAERS): early detection of intussusception and other events after rotavirus vaccination. *Vaccine* 2001; 19: 4627-34
- Rolka H, Barker L, Cadwel B, et al. Data mining for post-licensure vaccine safety and policy implications for using results. 2001 Proceedings of the Section on Health Policy Statistics, American Statistical Association. In press
- Trontell AE. How the US food and drug administration defines and detects adverse drug events. *Curr Ther Res Clin Exp* 2001; 62: 641-9
- Szarfman A. The application of bayesian data mining and graphic visualization tools to screen FDA's spontaneous reporting system database. Proceedings of the Section on Bayesian Statistical Science, 2000. American Statistical Association, 2000: 67-71
- Graham D, Waller P, Kurz X. A view from regulatory agencies. In: Strom BL, editor. *Pharmacoepidemiology*, 3rd ed. New York: John Wiley & Sons, 2000: 109-24
- DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *The American Statistician* 1999; 53: 177-90
- O'Neill RT, Szarfman A. Discussion: Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *The American Statistician* 1999; 53: 190-6
- Louis TA, Shen W. Discussion: Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system by William DuMouchel. *The American Statistician* 1999; 53: 196-8
- Madigan D. Discussion: Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system by William DuMouchel. *The American Statistician* 1999; 53: 198-200
- DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. Reply. *The American Statistician* 1999; 53: 201-2
- Szarfman A. Discussion: a report on the activities of the adverse events working groups: focus on improving the detection of rare but serious events. Proceedings of the Biopharmaceutical Section, 1999. Alexandria (VA): American Statistical Association: 12-4
- Szarfman A. New Methods for Signal Detection. 15th International Conference on Pharmacoepidemiology 1999. URL: <http://www.fda.gov/cder/present/ispe-1999/ispe-ana.pdf>
- Szarfman A, Talarico L, Levine JG. Analysis and risk assessment of hematological data from clinical trials: toxicology of the hematopoietic system. In: Sipes IG, McQueen CA, Gandolfi AJ. *Comprehensive toxicology*. Vol. 4. New York; Elsevier Science Inc.: 1997: 363-79
- Levine JG, Szarfman A. Standardised data structures and visualisation tools: a way to accelerate the regulatory review of the integrated summary of safety of new drug applications. *Biopharmaceutical Report* 1996; 4 (3): 12-7