LEADING ARTICLE

Drug Safety 2002; 25 (6): 381-392
0114-5916/02/0006-0381/$25.00/0

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

Ana Szarfman, Stella G. Machado and *Robert T. O'Neill*

Office of Biostatistics, Center for Drug Evaluation and Research, Food and Drug Administration, Rockville, Maryland, USA

Abstract

Since 1998, the US Food and Drug Administration (FDA) has been exploring new automated and rapid Bayesian data mining techniques. These techniques have been used to systematically screen the FDA's huge MedWatch database of voluntary reports of adverse drug events for possible events of concern.

The data mining method currently being used is the Multi-Item Gamma Poisson Shrinker (MGPS) program that replaced the Gamma Poisson Shrinker (GPS) program we originally used with the legacy database. The MGPS algorithm, the technical aspects of which are summarised in this paper, computes signal scores for pairs, and for higher-order (e.g. triplet, quadruplet) combinations of drugs and events that are significantly more frequent than their pair-wise associations would predict. MGPS generates consistent, redundant, and replicable signals while minimising random patterns. Signals are generated without using external exposure data, adverse event background information, or medical information on adverse drug reactions. The MGPS interface streamlines multiple input-output processes that previously had been manually integrated. The system, however, cannot distinguish between already-known associations and new associations, so the reviewers must filter these events.

In addition to detecting possible serious single-drug adverse event problems, MGPS is currently being evaluated to detect possible synergistic interactions between drugs (drug interactions) and adverse events (syndromes), and to detect differences among subgroups defined by gender and by age, such as paediatrics and geriatrics.

In the current data, only 3.4% of all 1.2 million drug-event pairs ever reported (with frequencies ≥1) generate signals [lower 95% confidence interval limit of the adjusted ratios of the observed counts over expected (O/E) counts (denoted EB05) of ≥2]. The total frequency count that contributed to signals comprised

23% (2.4 million) of the total number, 10.4 million of drug-event pairs reported, greatly facilitating a more focused follow-up and evaluation.

The algorithm provides an objective, systematic view of the data alerting reviewers to critically important, new safety signals. The study of signals detected by current methods, signals stored in the Center for Drug Evaluation and Research's Monitoring Adverse Reports Tracking System, and the signals regarding cerivastatin, a cholesterol-lowering drug voluntarily withdrawn from the market in August 2001, exemplify the potential of data mining to improve early signal detection. The operating characteristics of data mining in detecting early safety signals, exemplified by studying a drug recently well characterised by large clinical trials confirms our experience that the signals generated by data mining have high enough specificity to deserve further investigation. The application of these tools may ultimately improve usage recommendations.

The large database of voluntary adverse event reports of the US Food and Drug Administration (FDA) is the primary data resource for the study and identification of adverse reactions to regulated drugs and biological products in the US. This database contains 34.5 years of data and over 2 million reports across all marketed products in the US and increases in size each year by more than 300 000 reports.[1-3] These data have provided critical evidence about known and unknown harms associated with single or combination drug treatments.

There are well-known inherent problems in systematically analysing and interpreting voluntarily submitted data involving multiple drugs, medical conditions, and events per report, without the benefit of a research protocol, randomisation, and a control group of persons not taking the drug. Other difficulties include chronic under-reporting, occasional publicity-driven and litigation-driven episodes of over-reporting and misreporting, incomplete and missing information, and inconsistencies and changes over time in reporting and naming/coding practices. In addition, there is considerable uncertainty regarding the quality and completeness of the information contained in each data field, including dosage, formulation type, timing of exposure, and length of exposure and follow-up and in estimating the corresponding product exposure and background rate of adverse events. The extraction of useful information from this database presents multiple challenges, including managing, storing, and analysing such a large amount of data, and resolving event and drug dictionary problems and data miscoding.

There is a need for analytical methods that are capable of systematically screening this database to identify potential serious adverse events of concern in such a noisy background that properly balance the concerns for excessive signalling and accounting for background noise.

We describe some of the characteristics of the new automated and rapid Bayesian data mining techniques coupled with interactive data visualisation programs that we use to systematically identify adverse drug events in the FDA MedWatch database of voluntary reports.

1. The Data Mining System

The FDA's Bayesian data mining system consists of three components:

- The FDA MedWatch database of voluntary adverse drug event reports from drug manufacturers, healthcare practitioners, and consumers. This is a passive surveillance tool.
- The statistical algorithm for detecting unexpectedly frequent combinations of drugs and events. The data mining method currently being used is the Multi-Item Gamma Poisson Shrinker (MGPS) program[4] that replaced the Gamma Poisson Shrinker (GPS) program[5-11] we originally used with the legacy database. MGPS ad-

justs for the multiplicity of drugs and events per record, which is an important feature to have for the data collected after October 1997 (prior to November 1997, reports were limited to a maximum of four Event Codes and five drugs; however current reports have no such limits). MGPS 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.

- Specialised graphic visualisation tools to help organise, identify, and interpret patterns in the data.[12,13]

In the last 4 years of investigation the experience with these tools has grown. We have:

- identified that it is feasible to apply data mining to obtain signals from the 'noisy' postmarketing database of voluntary reports of adverse drug events
- performed extensive analyses to validate the signal scores generated by data mining
- taken steps towards making data mining results more easily available to safety evaluators by prototyping an internal web site with data mining results stored as Acrobat PDF files to continue the validation process
- reported the results of this work at several public symposia, including the recent Workshop on Datamining with Applications in Genomics, Clinical Trials and Post-Marketing Drug Risk at the Harvard School of Public Health on May 31 to June 1, 2001,[14] the First International Signal Generation Symposium in Southampton on June 25 to 26, 2001, and the Drug Information Association 2001 Annual Meeting in Denver on July 8 to 10, 2001. In the last two meetings as well as at the 9th Merck-Temple Conference On Research topics in Pharmaceutical Statistics on October 5, 2001, scientists from pharmaceutical companies presented their results with GPS and MGPS including some preliminary validation results.

The Gamma Poisson Shrinker (GPS) data-mining algorithm is in the public domain,[15] a version of MGPS for academic/research purposes will soon be made publicly available, and the FDA adverse event data are made available to the public via the National Technical Information Service.[16] This access has facilitated pilot use by scientists outside the Agency including those at several pharmaceutical companies. Other centers at the FDA and the National Immunisation Program at the National Centers for Disease Control and Prevention are piloting the application of MGPS in-house.[17]

2. Database

Prior to November 1997, adverse events were coded using over 1200 event codes in COSTART (Coding Symbols for Thesaurus of Adverse Reaction Terms). Since November 1997, adverse event reports have been coded using the more granular MedDRA (Medical Dictionary for Regulatory Activities) with over 11 000 preferred term event codes, of which 7000 are in use in the current database. Currently, the sum of all frequencies of drug-event pairs is 10.4 million (50% from before 1997). This exceeds the total number of reports, since reports typically involve information on more than one drug and event. Prior to November 1997, reports were limited to a maximum of four event codes and five drugs on each submitted report form; however, current reports have no such limits.

3. The Empirical Bayesian Estimation Program

The primary output of this program (originally prototyped as an S-PLUS program and currently in C++) consists of a set of 'signal scores' that are interpreted as hypotheses regarding potential causal associations among drugs and events.[4]

Considered as a two-way (drug-by-event) table, the MedWatch database is quite sparsely populated – only 1.17 million (2%) of the over 56 million possible drug-event pairs of 7000 event codes and over 8000 decoded drug names in use have ever been reported, under 600 000 (1.0%) have been reported more than once, and 400 000 (0.7%) more than twice.

For an arbitrary itemset, it is desired to estimate the expectation $\lambda = E[N/E]$, where N is the observed frequency of the itemset, and E is a baseline or null hypotheis count, such as a count predicted from the assumption that items are independent. An itemset is defined by its members *i, j, k,* ..., which occur as subscripts to *N, E,* and other variables, so that, for example, N_{ij} is the number of reports or transactions involving both item *i* and item *j*, E_{ijk} is the baseline prediction for the number of reports including the itemset triple (*i, j, k*), etc.

A common model for computing baseline counts is the assumption of within-stratum independence, and when E is computed under this assumption we shall often denote it by $E0$. Assume that all reports are assigned to strata denoted by $s = 1, 2, ..., S$. Let

P_i^s = proportion of reports in stratum s that contain item i
n_s = total number of reports in stratum s

For pairs of items, baseline frequencies E_{ij} are defined under independence as

$$E0_{ij} = \Sigma_s \, n_s \, P_i^s \, P_j^s$$

For triples, baseline frequencies assuming independence are

$$E0_{ijk} = \Sigma_s \, n_s \, P_i^s \, P_j^s \, P_k^s$$

For itemsets of size 3 or more, an "all-2-factor" loglinear model can be defined as the frequencies $E2$ for the itemset that match all the estimated pairwise two-way marginal frequencies but contain no higher order dependencies. For triples, $E2_{ijk}$ agree with the estimates for the three pairs

$$\lambda_{ij} E0_{ij} \qquad \lambda_{ik} E0_{ik} \qquad \lambda_{jk} E0_{jk}$$

For 4-tuples $E2_{ijkl}$ agrees with 6 such pairs, etc.

Then for itemsets of size 3 or more we compare the estimated frequency to the all-2-factor prediction by simple subtraction. For example, in case of triples,

$$Excess2_{ijk} = \lambda_{ijk} E0_{ijk} - E2_{ijk}$$

The parameters λ above are estimated by the geometric means, denoted by *EBGM*, of their empirical Bayes posterior distributions. For simplicity, the formulae below use just two subscripts, for itemsets of size 2, such as the occurence of drug *i* and symptom *j* in a medical report. Estimates for other itemset sizes are computed analogously. Let

N_{ij} = the observed counts
E_{ij} = the expected (baseline) count
$RR_{ij} = N_{ij} / E_{ij}$ = ratio of observed to baseline

We estimate $\lambda_{ij} = \mu_{ij} / E_{ij}$, where $N_{ij} \sim$ Poisson(μ_{ij})

Assume a superpopulation model for λ_{ij} (prior distribution) based on a mixture of two gamma distributions (a convenient 5-parameter family of distributions that can fit almost any empirical distribution):

$$\pi(\lambda; \alpha_1, \beta_1, \alpha_2, \beta_2, P) = P \, g(\lambda; \alpha_1, \beta_1) + (1 - P) \, g(\lambda; \alpha_2, \beta_2)$$
$$g(\lambda; \alpha, \beta) = \beta^{\alpha} \lambda^{\alpha - 1} e^{-\beta\lambda} / \Gamma(\alpha)$$

Estimate the prior distribution from all the (N_{ij}, E_{ij}) pairs. Estimate the 5 hyperparameters

$$\theta = (\alpha_1, \beta_1, \alpha_2, \beta_2, P)$$

by maximizing the likelihood function $L(\theta)$ in 5 dimensions:

$$L(\theta) = \Pi_{i,j} \{P\, f(N_{ij};\, \alpha_1,\, \beta_1,\, E_{ij}) + (1 - P)\, f(N_{ij};\, \alpha_2,\, \beta_2,\, E_{ij})\}$$

where $f(n;\, \alpha,\, \beta,\, E) = (1 + \beta/E)^{-n}\, (1 + E/\beta)^{-\alpha}\, \Gamma(\alpha + n) / \Gamma(\alpha)\, n!$

If a threshold (minimum count) for the observed counts is used, these formulae are modified to condition on $N_{ij} \geq n^*$ (where n^* = the threshold count).

Given θ, the posterior distributions of each λ_{ij} are also a mixture of gamma distributions used to create "shrinkage" estimates. Assuming that θ and E are known, then the distribution of N is

$$\text{Prob}(N = n) = P\, f(n;\, \alpha_1,\, \beta_1,\, E) + (1 - P)\, f(n;\, \alpha_2,\, \beta_2,\, E)$$

Let Q_n be the posterior probability that λ came from the first component of the mixture, given $N = n$. From Bayes rule, the formula for Q_n is

$$Q_n = P\, f(n;\, \alpha_1,\, \beta_1,\, E) / [P\, f(n;\, \alpha_1, \beta_1, E) + (1 - P)\, f(n;\, \alpha_2,\, \beta_2,\, E)]$$

Then, the posterior distribution of λ, after observing $N = n$ can be represented as

$$\lambda | N = n \sim \pi(\lambda;\, \alpha_1 + n,\, \beta_1 + E,\, \alpha_2 + n,\, \beta_2 + E,\, Q_n)$$

where (as above)

$$\pi(\lambda;\, \alpha_1,\, \beta_1,\, \alpha_2,\, \beta_2,\, P) = P\, g(\lambda;\, \alpha_1,\, \beta_1) + (1 - P)\, g(\lambda;\, \alpha_2,\, \beta_2)$$

We will use the expectation value

$$E[\log(\lambda_{ij}) \mid N_{ij},\, \theta]$$

as a means of estimating the "true" value of λ_{ij}

To obtain a quantity on the same scale as *RR*, we define the Empirical Bayes Geometric Mean:

$$EBGM_{ij} = e^{E[\log(\lambda_{ij}) \mid N_{ij},\, \theta]}, \text{ where}$$

$$E[\lambda \mid N = n, \theta] = Q_n\, (\alpha_1 + n)/(\beta_1 + E)] + (1 - Q_n)\, (\alpha_2 + n)/(\beta_2 + E)$$

$$E[\log(\lambda) \mid N = n, \theta] = Q_n\, [\psi(\alpha_1 + n) - \log(\beta_1 + E)] + (1 - Q_n)\, [\psi(\alpha_2 + n) - \log(\beta_2 + E)]$$

where $\psi(x) = d(\log \Gamma(x))/dx$. The above expectations follow from known properties of gamma distributions. In the same way, the cumulative gamma distribution function can be used to obtain percentiles of the posterior distribution of λ. The 5th percentile of λ is denoted

$$EB05_{ij} = \text{Solution to: } \text{Prob}(\lambda < EB05 \mid N_{ij},\, \theta) = 0.05$$

and is interpreted as a lower 1-sided 95% confidence limit.

Fig 1. A technical overview of the Multi-Item Gamma Poisson Shrinker program.

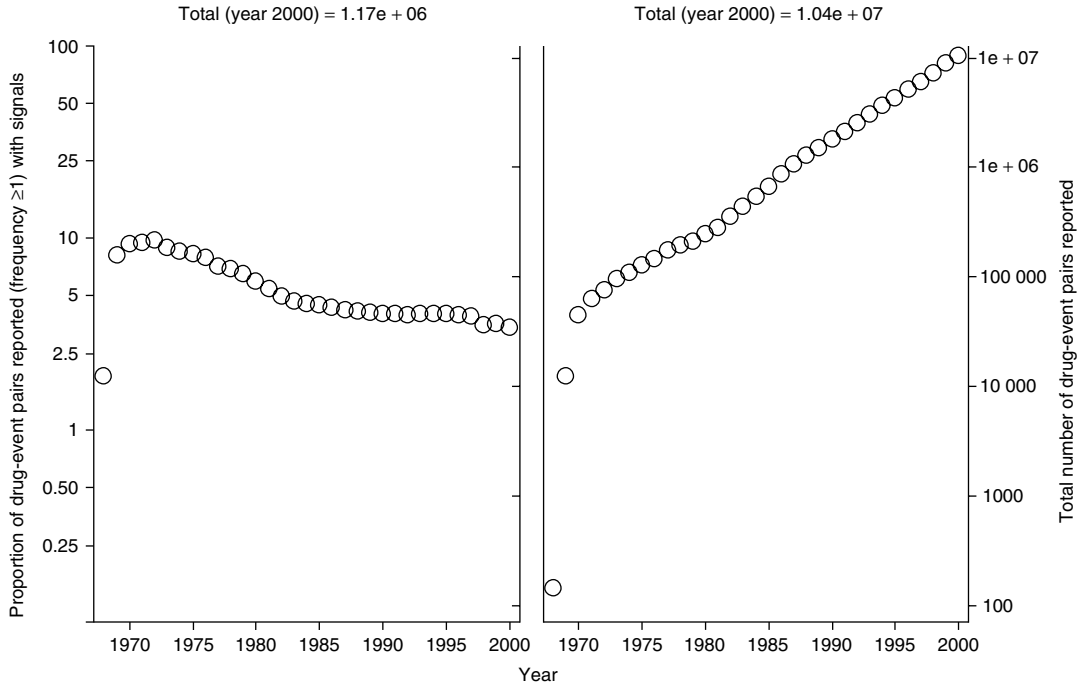

Fig. 2. Progression of the proportion of drug-event pairs with reported frequency (≥1) with signals (EB05 ≥2) (y1-axis) and of the total number of drug-event pairs reported (y2-axis) by year (x-axis) over the period 1968-2000. EB05 = lower 95% confidence interval limit of the Empirical Bayes Geometric Mean (EBGM). **e + 06** = 1 million; **e + 07** = 10 million.

To compute internal expected counts for every item set pair and higher multiples MGPS uses a stratified full independence model. The probabilities of individual item values within a stratum (year, age, and gender) are estimated from that stratum's marginal counts and total number of reports and added across strata to reduce confounding.

MGPS uses a model derived from the data to describe the distribution of the ratios of the observed counts over expected counts (O/E) in the data to shrink unreliably estimated O/E due to small expected counts, a frequent finding in this sparsely populated database. For greater than two-way associations MGPS can also identify the excess number of reports unexplained by pair-wise associations. Figure 1 gives a technical overview of MGPS.

4. Positive Signal Scores for Pair-Wise and Multiple-Item Associations

MGPS derives signal scores, adjusted ratios of the O/E counts [Empirical Bayes Geometric Mean (EBGM)] exclusively from the 'numerator' data.

We use the lower 95% confidence interval limit of the EBGM (denoted as EB05) for pairs and higher multiples as the signal score. For higher multiples the number of reports unexplained by pair-wise occurrences are also analysed (figure 1).

The cut-off for identifying a signal score for pair-wise combinations as higher-than-expected is an EB05 ≥2. This criterion ensures with a high degree of confidence that, regardless of count size, the particular drug-event combination is being reported at least twice as often as it would be if there

were no association between the drug and the event. Our experience indicates that the signals generated using this cut-off have high enough specificity to deserve further investigation (see validation, section 6).

5. Number of Signals

We used MGPS to study both the progression of the proportion of signals within the populated drug-event pairs and within the total number of populated pairs over the period 1968 to 2000 (the 33-year span of this database).

In the current data, only 3.4% (<40 000) of all 1.17 million drug-event pairs ever reported (with frequencies ≥1) were signalled by data mining. This proportion peaked around 10% between 1969 to 1974 and decreased steadily over the following years (figure 2, left panel).

During the same 33-year period, the total number of drug-event pairs reported increased steadily from 5000 in 1969 to 10.4 million in 2000 (figure 2, right panel). In the year 2000, the total frequency count that contributed to the <40 000 signals comprises 23% (2.4 million) of the total 10.4 million drug-event associations reported. This proportion peaked around 40% in 1969 and decreased steadily over the following years (not shown).

Over the course of time the chance of an event being reported increases, more data are being accrued, more drugs are added, and the opportunity for a drug-event pair to be reported increases. Even so, as described above, a small proportion of signals (3.4%) captures a high proportion of the total number of drug-event associations reported (23%).

Because a clinically meaningful individual event typically corresponds to multiple event codes (between 5 to 10), the <40 000 positive signal scores associated with drug-event pairs may actually represent between 4000 to 8000 clinically distinct drug-event associations. A high proportion of these clinically meaningful events includes known reactions and indications. This decreases the number of potential signals even further.

6. Validation of the Signal Scores Generated by Data Mining

An overview of several validation activities including examining the correlation between internal counts of event frequencies with negative signals naming a drug (surrogate usage) with drug-exposure estimates obtained from an independent external source has been recently published.[11,18] Other activities include the study of the potential role of MGPS in paediatric safety assessment and early detection of adverse drug events,[18] of GPS in the early detection of intussusception after rotavirus vaccination[19] and the policy implications of data mining.[17]

The validation process of the consistency of scores started by defining what an ideal post-marketing signalling and alerting system should be doing. The goal of an ideal screening system is to generate accurate, systematic, objective, rapid, and reproducible signals.

Ideally we would like to show that data mining identifies true serious adverse event-drug combinations when used with the spontaneous reports database.

The validation of data mining in an absolute sense is an unreachable goal. There is no true gold standard for every adverse drug event against which the data mining systematic results can be linked and compared. We also need to consider the following:

- That there are different collections of serious adverse event-drug combinations: one in the patient population, a second in the voluntary reports, and a third in medical databases. Collections of medical knowledge about adverse events also exist in additional populations such as healthcare providers, medical reviewers/regulators, and in literature reports and product package inserts. These groups are not identical.
- That there are no complementary techniques in place to systematically and independently detect and analyse potential adverse drug event signals using other medical databases, such as health maintenance organisation records and

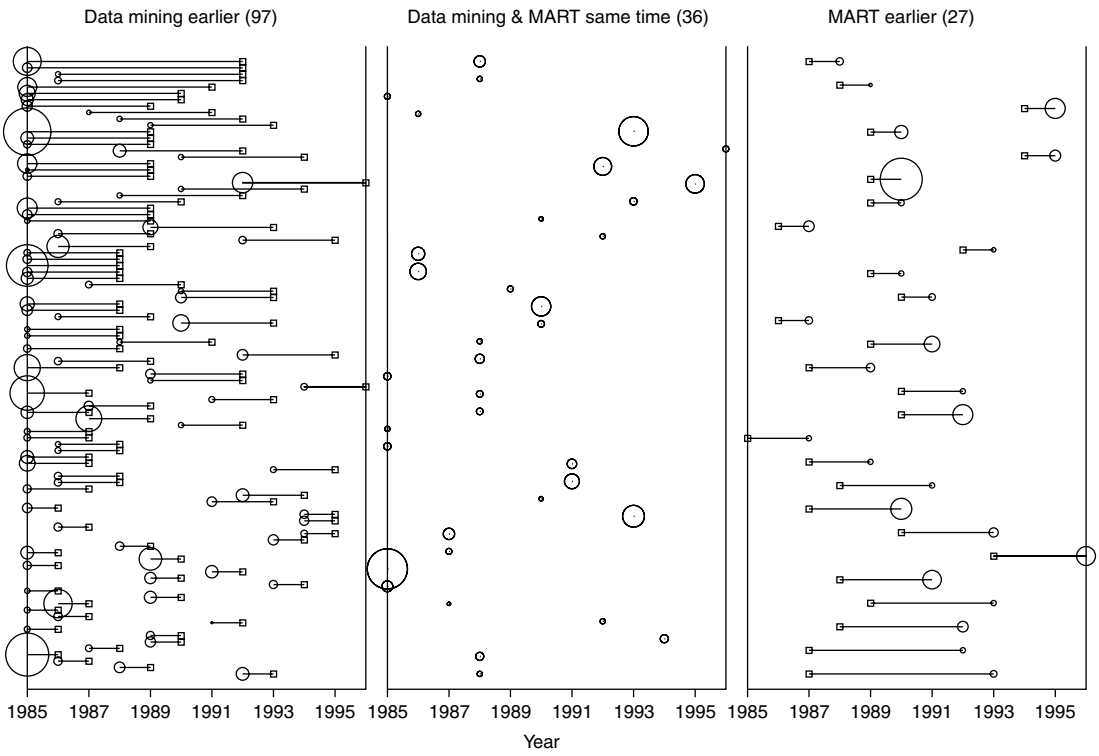

Fig. 3. Differences in time to detection of 160 different drug-event combinations coded as 'signals' (95 drugs) in the Monitoring Adverse Reports Tracking (MART) system by current and data mining method. The length of each line represents the difference in the time of detection between the two methods. The size of the circles is proportional to the number of reports when the first signal scores were detected. Squares represent the unidentified number of reports when signals were entered into MART.

military databases. These complementary techniques that would facilitate the systematic and independent validation of our data mining approach with spontaneous reports are not available. In addition, the medical knowledge related to known or suspected drug induced adverse events of healthcare providers, of medical reviewers/regulators, of literature reports, and product package inserts, are also difficult to access and link to in a systematic way.

- That a subset of serious adverse event-drug combinations in the patient population can be detected by the best data-scrutiny method using the voluntary reports. This subset can be detected by reviewers using current methods or by using the MGPS algorithm[4] and some specific threshold level. These last two groups are not identical.

Through retrospective studies, we have demonstrated that data mining performs well in early detection of simple drug-event associations that represent safety signals.

We 'rolled back' the database in time by extracting the data and performing analyses of signal scores associated with cumulative counts over time. We then looked to see what signals would have been present if we had applied the method at the end of cumulative time periods.

We started by using GPS to analyse the differences in year of detection of 30 known adverse

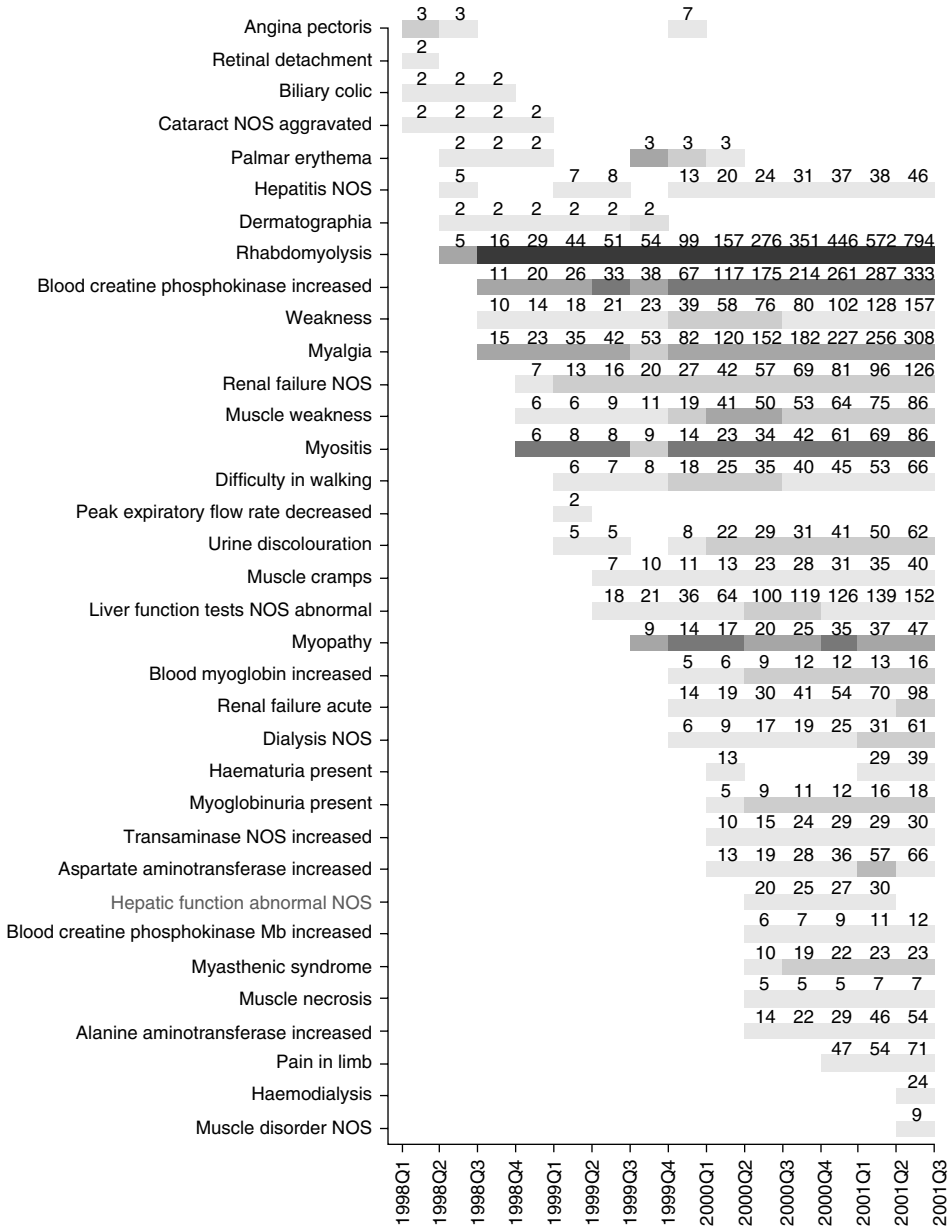

Fig. 4. Cumulative signal scores (grey shading) and number of reports (numbers) by year and quarter (x-axis) for cerivastatin after analysing only the data collected between first quarter of 1998 and the third quarter of 2001. Event codes are Medical Dictionary for Regulatory Activities (MedDRA) preferred terms (y-axis). For the grey shading, darker shading corresponds to stronger signals. **NOS** = not otherwise stated.

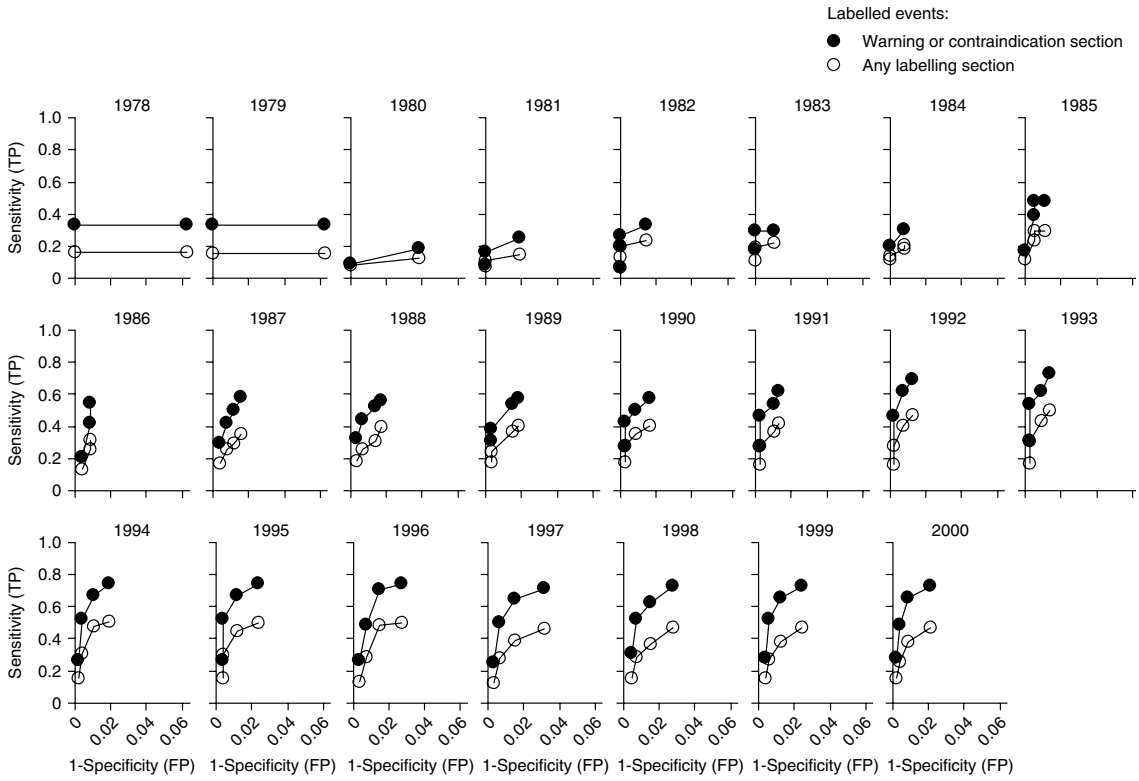

Fig. 5. Receiver operator curves describe the progression of the sensitivity [true positive (TP), y-axis] against 1-specificity [false positive (FP), x-axis] of the data mining approach for one drug well characterised by recent large clinical trials in the cumulative post-marketing annual adverse event data between 1978-2000.

event signals selected by two experienced epidemiologists at Center for Drug Evaluation and Research (CDER). Data mining identified as positive all 30 signals, 20 in the data collected 1 to 5 years before index cases were detected by standard methods, nine the same year and one signal 1-year after.

We then used GPS to analyse the differences in time of detection of 160 drug-event signals for 85 drugs detected between 1985 to 1996 and collected in CDER's Monitoring Adverse Reports Tracking (MART) System selected because the drug-event combinations names matched positive data mining drug-event combinations. The MART system collected adverse events manually detected by safety evaluators until November 1997. Ninety-seven positive data mining signals were detected in the data collected 1 to 4 years before they were entered as signals in the MART system, 36 the same year, and 27 of them 1 to 3 years later (figure 3). Half of the 27 signals detected later included severe liver events, Stevens Johnson's syndrome, aplastic anaemia, and anaphylaxis, that are easier to characterise with fewer reports by current methods.

Some additional examples using MGPS showing the early detection of important signals have been published.[18] A recent data mining analysis of cerivastatin is shown in figure 4 (cumulative signal plot), which indicates how the ongoing application of data mining can assist in the early iden-

tification of important drug safety problems (in this case a severe muscle adverse reaction, rhabdomyolysis and renal failure in 1998, and dialysis in 1999). Cerivastatin was voluntarily withdrawn from the US market on August 2001 because of reports of fatal rhabdomyolysis, renal failure, and other organ failure.[20]

The operating characteristics of MGPS have been studied through data mining runs using actual data. The sensitivity and specificity of data mining taking, as an example, a drug which has been well-characterised by recent large clinical trials, here called Drug A, were evaluated for four 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 (figure 5).

We contrasted the four separate EB05 cut points with events identified either in any labelling section or in the warnings or contraindications sections for Drug A. To compare labelled events and event codes we had to reconcile the fact that labelled events use their own terminology and do not correspond one-to-one with the event codes in the MedWatch database. There are multiple event codes that can provide evidence for an event at the granularity of the labelled event. We considered each labelled event as being 'derived from' a cluster of one or more event codes in spontaneous reports and used our experience analysing these reports to produce a 'mapping table' between this drug's labelled events and event codes. For event codes that overlapped labelled events, we applied our experience analysing spontaneous reports to assign them to a preferred labelled event. To reconcile label and coding terminology differences, we used a double criterion to analyse sensitivity and specificity. Sensitivity was estimated by counting current labelled events or current warnings and contraindications signalled by any mapped event code as the number of detected/(number of detected + undetected). Specificity was estimated by counting individual event codes that did not map to a labelled event as the number of true negative/(number of true negative + false positive).

The results showed a high degree of specificity for all thresholds used since the first signals were detected by MGPS. The sensitivity for warnings and contraindications was higher than the one observed for the analysis across all labelled events. Lowering the thresholds for important event codes or detecting higher order synergic associations between drugs and multiple events increase sensitivity (not shown).

However, using the label as gold standard for data mining has important limitations because appropriately labelled events may not be reported any more, spurious events may get into the label to avoid 'failure to warn' lawsuits, and the negative signals of today may appear as positive in the future.

7. Conclusions

Data mining of adverse event databases is a tool to help with the challenging task of systematically detecting signals among the over 300 000 MedWatch reports submitted annually to the FDA and is most usefully applied with full awareness of the limitations and circumstances of voluntary reporting, coding, database characteristics, etc. The data mining signals by themselves are not indicators of problems, but indicators of possible problems. Data mining is not intended to replace current pharmacovigilance techniques, but to enhance them.[18]

The MGPS algorithm appears to be quite robust in that it systematically identifies relatively subtle associations even when operating on a database that is known to contain considerable 'noise' (misreporting, duplications, coding errors).

Signals are generated for a relatively small proportion of all distinct drug-event pairs in the database. These signals capture a high proportion of the total number of drug-event pairs reported, greatly facilitating more focused follow-up and evaluation.

The development of complementary techniques to systematically and independently detect and analyse potential adverse drug event signals using other

medical databases, such as health maintenance organisation records and military databases will facilitate the systematic, independent validation of the Bayesian data mining program outputs.

It is our opinion that the development of better practical tools for safety data mining will make an important contribution to public health.

Acknowledgements

The datamining technology referred to in this article was developed with grants from the Office of Women's Health and the Center of Drug Evaluation and Research of the Food and Drug Administration and from an 'Unmet Needs' Grant from the National Centers for Disease Control and Prevention, United States Department of Health & Human Services.

We thank William DuMouchel of AT&T for developing the empirical Bayes data mining algorithms that we are applying to frequency counts; David Fram of Lincoln Technologies, Inc, Jeremy Pool, Ilya Yunus, and Ava-Robin Cohen of PPD Informatics™ for providing critical technical information development and implementation expertise; Diane Wysowski and Janos Bacsanyi from CDER for providing adverse event signals detected by current methods; Susan Ellenberg, Miles Braun, and Manette Niu from CBER, FDA and Henry Rolka from CDC for precious feedback and collaboration. We thank Phillip Perucci and Stacey Nichols from FDA for very valuable technical support.

References

1. Baum C, Kweder SL, Anello C. The spontaneous reporting system in the United States. In: Strom BL, editor. Pharmacoepidemiology. 2nd ed. New York; John Wiley & Sons, 1994: 125-37
2. 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
3. Trontell AE. How the US food and drug administration defines and detects adverse drug events. Curr Ther Res Clin Exp 2001; 62: 641-9
4. 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
5. 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
6. 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
7. 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
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
9. 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
10. 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
11. 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
12. 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
13. 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
14. 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 [Accessed 2002 May]
15. ftp://ftp.research.att.com/dist/gps [Accessed 2002 May]
16. US Department of Commerce National Technical Information Service (NTIS), http://www.ntis.gov [Accessed 2002 May]
17. 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
18. 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
19. 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
20. FDA Talk paper. Bayer voluntarily withdraws baycol. FDA talk paper no. T01-34. 2001 Aug 8

Correspondence and offprints: Dr *Ana Szarfman*, Office of Biostatistics, Center for Drug Evaluation and Research, Food and Drug Administration, 5600 Fishers Lane, Rockville, MD 20857, USA.
E-mail: SZARFMAN@cder.fda.gov