

Manfred Hauben · Lester Reich · Stephanie Chung

Postmarketing surveillance of potentially fatal reactions to oncology drugs: potential utility of two signal-detection algorithms

Received: 28 May 2004 / Accepted: 27 August 2004 / Published online: 17 November 2004
© Springer-Verlag 2004

Abstract *Purpose:* Several data mining algorithms (DMAs) are being studied in hopes of enhancing screening of large post-marketing safety databases for signals of novel adverse events (AEs). The objective of this study was to apply two DMAs to the United States FDA Adverse Event Reporting System (AERS) database to see whether signals of potentially fatal AEs with cancer drugs might have been identified earlier than with traditional methods.

Methods: Screening algorithms used for analysis were the multi-item gamma Poisson shrinker (MGPS) and proportional reporting ratios (PRRs). Data mining was performed on data from the FDA AERS database. When a signal was identified, it was compared with that in the year in which the event was added to package insert and/or the year a “case series” was published. A recent publication summarizing the time of dissemination of information on potentially fatal AEs to cancer drugs provided the data set for analysis.

Results: The peer-reviewed published analysis contained 21 drugs and 26 drug–event combinations (DECs) that were considered sufficiently specific for data mining. Twenty-four of the DECs generated a signal of disproportionate reporting with PRRs (6 at 1 year and 16 from 2 years to 18 years prior to either a published “case series” or a package insert change) and 20 with MGPS (3 at 1 year and 11 from 2 years to 16 years prior to

either a published “case series” or a package insert change). Two DECs did not signal with either DMA.

Conclusion: At least one commonly cited DMA generated a signal of disproportionate reporting for 24 of 26 DECs for selected cancer drugs. For 16 DECs, one could conclude that a signal was generated well in advance (≥ 2 years) of standard techniques in use with at least one DMA. DMAs might be useful in supplementing traditional surveillance strategies with oncology drugs and other drugs with similar features. (i.e., drugs that may be approved on an accelerated basis, are known to have serious toxicity, are administered to patients with substantial and complicated comorbid illness, are not available to the general medical community, and may have a high frequency of “off-label” use).

Purpose

A compilation of potentially fatal adverse drug reactions for cancer drugs previously published in an oncology journal concluded that these events may be discovered as many as 36 years after a drug receives United States Food and Drug Administration (FDA) approval [8]. This suggested to the authors that there was a need for continued vigilance for such reactions with cancer drugs. In hopes of enhancing the ability to screen large databases of adverse event (AE) reports, several computer-assisted statistical signal detection algorithms—also known as data mining algorithms (DMAs)—are being studied in hopes of improving safety surveillance [1–7, 9–11]. These algorithms might be especially useful for cancer drugs that (1) can be approved on an accelerated basis, (2) are known to have serious toxicity, (3) are administered to patients with substantial and complicated comorbid illness, (4) are not available to the general medical community, and (5) may have a high frequency of “off-label” use. In this study, we applied the DMAs known as the multi-item gamma-Poisson shrinker (MGPS) and proportional reporting ratios

M. Hauben · L. Reich (✉) · S. Chung
Pfizer Inc, 150 E. 42nd Street, New York, NY 10017, USA
E-mail: lester.reich@pfizer.com
Tel.: +1-212-7336143
Fax: +1-212-8088679

M. Hauben
Department of Medicine,
New York University School of Medicine,
New York, NY, USA

M. Hauben
Departments of Pharmacology and Community
and Preventive Medicine, New York Medical College,
Valhalla, NY 10595, USA

(PRRs) to the FDA Adverse Event Reporting System (AERS) database to determine whether these methods would have flagged a sample of serious AEs obtained from the aforementioned published compilation of cancer AEs in advance of the “traditional” methods that were in operation at the time.

Methods

The FDA AERS database is a computerized information database for post-approval safety surveillance. It functions as an early warning system for adverse drug reactions not detected during pre-approval testing. It contains AE reports with approved drugs and therapeutic biological products submitted in accordance with mandatory reporting obligations by pharmaceutical companies and voluntarily by health care professionals and consumers. AEs are submitted on MedWatch forms. AE reports are reviewed and coded for data entry in accordance with the standardized terminology of the Medical Dictionary for Medical Regulatory Activities (MedDRA). Quarterly extracts are available through the National Technical Information Service (NTIS). These quarterly updates are subjected to extensive cleaning (i.e., removal of redundant drug nomenclature and duplicate reports) prior to data mining. The data extract used for the current analysis was through the first quarter of 2003 [12].

The two DMAs chosen for this analysis were PRRs [2] and the empirical Bayesian MGPS, (Lincoln Technologies, Wellesley Hills, MA, USA) [11].

The PRR is a simple metric relating the proportional representation of an event of interest with a drug of interest compared with the proportional representation of that event among all other drugs in the database (Table 1). For this analysis, a $PRR > 2$ with an associated $\chi^2 > 4$ (with Yates correction) was considered a “signal” of disproportionate reporting (SDR), which has been frequently cited in published studies of data mining in conjunction with case count thresholds (e.g., $n > 2$) [2]. We also examined whether imposing a case count threshold ($n > 2$) in series with PRRs affected the results.

The theoretical basis of MGPS has been described in detail elsewhere [1, 11] but briefly is as follows. Expected counts for item sets (i.e., drug–event combinations or DEC)s are based on the product of the marginal probabilities of each item (drug and event) in the database.

Table 1 Proportional reporting ratios

	Drug of interest	All other drugs	Total
Reaction(s) of interest	A	B	A + B
All other reactions	C	D	C + D
Total	A + C	B + D	A + B + C + D

$$PRRs = [A(A + C)]/[B(B + D)]$$

The observed to expected (O/E) ratio is initially calculated as a crude disproportionality metric. Since the same ratio could be obtained from cell counts (frequencies) of markedly different sizes (O/E ratios based on smaller cell counts being considered more variable or imprecise) further modeling using maximum likelihood estimation and Bayesian inference are used to adjust the crude O/E ratios based on the respective cell counts. Each cell is considered to represent a Poisson process in which the Poisson parameter distribution is related to a mixture of two gamma distributions. The prior probability distribution of the gamma parameters are obtained by applying an interactive maximum likelihood algorithm to a negative binomial mixture likelihood. Posterior estimates of the gamma parameters are obtained by updating the prior with the individual cell counts via Bayes theorem.

Using logarithmic transformations or taking the lower 5% cut-off of the posterior distribution ($EB_{0.05}$), an expectation value that adjusts for the variability by down weighting or “shrinking” the parameters associated with low cell counts is obtained. These metrics are known as the empirical Bayes geometric mean (EBGM) and the $EB_{0.05}$. An $EB_{0.05}$ of 8 may therefore be interpreted to mean that reports of the particular DEC occur in the database eight times more frequently than would be expected if drug and event were independently distributed in the database. The signal metric used for a threshold in the current analysis was the frequently cited lower 5% cut-off of the EBGM greater than two ($EB_{0.05} > 2$) [1]. The developers of MGPS have stated that for $EB_{0.05} \geq 2$, “our experience indicates that the signals using this cutoff have high enough specificity to deserve further investigations.”

A variety of data mining options and parameters exist including basic covariate adjustment (stratification by age, gender, and year of report) and cumulative subsetting. Stratification tends to reduce spurious associations due to confounding and markedly decreases the volume of disproportionalities [1, 5].

For the present analysis, the data mining was performed on suspect drug–AE pairs using stratification by age, gender, and FDA year of report with cumulative subsetting by year.

A recent peer-reviewed publication summarizing all AEs associated with oncology drugs reported from 2000 to 2002 using a pre-defined search strategy provided the sample of DEC)s for this analysis.¹ In addition to the drug and event, this publication provided the year the drug was approved, the data source for the AE (e.g., FDA MedWatch Program, investigative team’s comprehensive cancer center) and the time interval between approval and package insert revision. It also supplied the year and reference in which ten or more cases (i.e., “case series”) of the AE were described in a single article. The year in which the signal metric exceeded the specified thresholds was compared with the year of the package insert revision and/or publication of a “case series.” For each specific AE or group of AEs, the ver-

batim term(s) from the paper and all MedDRA Preferred Terms that were considered clinically equivalent or closely related to the verbatim term as determined by an experienced reviewer were used for data mining. A second experienced reviewer reviewed and verified the initial findings.

Results

The peer-reviewed published analysis contained 21 drugs and 26 DECAs that were considered sufficiently specific for data mining. Prior to 1995, 10 of the drugs were approved in the United States. Of the 11 drugs approved in 1995 or later, 6 underwent standard approval process and the remaining 5 drugs underwent an accelerated approval process. Of the DECAs, 24 generated a signal of disproportionate reporting with PRR (6 at 1 year and 16 from 2 years to 18 years prior to either a published “case series” or a package insert change) and 20 with MGPS (3 at 1 year and 11 from 2 years to 16 years prior to either a published “case series” or a package insert change). One DECA did not generate a signal with either algorithm. With PRRs, 18 DECAs (7 involving drugs that

underwent accelerated approval) would have been highlighted in the year the first MedWatch form (range for number of reports received in first year: 1–14) was received. With MGPS, 6 DECAs (4 involving drugs that underwent accelerated approval) would have been highlighted in the year the first MedWatch form (range 5–24) was received. These and other findings are provided in Table 2. Imposing a case count threshold of $n > 2$ in series with a PRR threshold affected timing to initial SDR for three DECAs. For only one of these DECAs (gemtuzumab-hepatic venoocclusive disease) did this threshold effect timing of SDR in relation to package insert change/“case series” (i.e., SDR occurred one year after instead of the same year).

Discussion and conclusion

At least one DMA appeared to generate a signal of disproportionate reporting for 22 of 26 DECAs 1 year or more prior to publication of a case series and/or a package insert change for selected cancer drugs. The temporal resolution of 1 year is a limitation to this study, which makes it difficult to determine with any

Table 2 Data mining on 26 potentially fatal drug–event combinations with 21 oncology drugs

Drug (approval date [month/year]) ^a : event	Timing (year)				
	“Case series” ^a	1st MedWatch	1st year PRR > 2 ($\chi^2 > 4$)	1st year EB05 > 2	Year of package insert change ^a
Drugs approved before 1995					
Vinblastine (11–65): fatalities with intrathecal use		1969	1991	–	2000
Tamoxifen (12/77): endometrial carcinoma	2000	1983	1984	1984	2002
Tamoxifen (12/77): uterine sarcoma	2000	1989	1989	1993	2002
Immune globulin (3/81): stroke and other thrombotic episodes	2000	1985	1999	2002	–
Interferon alpha-2a (6/86): neuropsychiatric events	–	1986	1989	1990	2001
Interferon alpha-2b (6/86): neuropsychiatric events	–	1987	1992	1993	2001
Flutamide (1/89): interstitial pneumonitis	2002	1990	1999	2001	–
Erythropoietin (6/89): pure red cell aplasia	2002	1999	1999	2000	2002
Erythropoietin (6/89): serratia infection	2001	–	NA	NA	–
Fludarabine (4/91): pancytopenia	–	1992	1992	1992	2002
Pamidronate (10/91): severe renal toxicity	2001	1983	1983	1996	2002
Vinorelbine (12/94): interstitial pneumonitis	1996	1995	1995	2002	2000
Drugs approved after 1995 (standard approval)					
Bicalutamide (10/95): interstitial pneumonitis	2002	1996	1996	–	2001
Gemcitabine (5/96): hemolytic uremic syndrome	1999	1996	1996	1997	2000
Oprelvekin (11–97): papilledema	–	1999	1999	2001	2001
Rituxamab (11/97): severe infusion reactions	2000	1998	1998	1998	1999
Trastuzumab (9/98): severe infusion reactions	1999	1998	1998	1999	2000
Arsenic trioxide (9/00): torsades/sudden death	2001	2002	2002	–	In 1st label
Drugs approved after 1995 (accelerated approval)					
Irinotecan (6/96): treatment-related fatalities	2001	1996	1996	1996	2002
Capecitabine (4/98): warfarin interaction (PT increased)	–	1998	1998	1998	2001
Capecitabine (4/98): renal failure	–	1999	–	–	1999
Capecitabine (4/98): neutropenia in patients with renal insufficiency	–	1998	1998	2000	2000
Thalidomide (7/98) DVT and PE	2002	1999	1999	1999	–
Gemtuzumab (5/00): hepatic veno-occlusive disease	2001	2001	2001	2002	2001
Gemtuzumab (5/00): severe infusion reactions	2001	2000	2000	2000	2001
Imatinib (5/01): tumor/GI bleeding	2002	2001	2001	–	2002

^aAs stated in journal article and rounded off to the nearest year except where noted

degree of certainty whether the signal from the DMA truly predated the initial signal identified in the cases in which the difference was only 1 year. Nevertheless, for 16 DEC with PRRs and 11 DEC with MGPS, one could conclude that a SDR was significant in that it would have been generated well enough in advance (≥ 2 years) of standard techniques in use at the time. Therefore, both methods could have had potential utility in this setting. PRRs have certain advantages (e.g., they are simpler and more “sensitive”), but a potential drawback of such methods is reduced “specificity,” which could result in an overabundance of signals, including “false-positive” signals that would be difficult to manage in general pharmacovigilance settings. However, in the specific situation studied here, the enhanced “sensitivity” may justify the reduced “specificity.” Going forward, any performance differentials between PRR and MGPS similar to what we observed are likely to be significantly mitigated when these methods are used as one element of a comprehensive pharmacovigilance program that utilizes multiple approaches to signal detection. Additionally, given that currently cited thresholds are unvalidated, somewhat subjective, and adjustable, performance gradients would be mitigated by titrating or optimizing thresholds. Noteworthy is the finding that performance differentials seemed to narrow with the most recently approved drugs, especially those that went with accelerated approval (e.g., four of the eight listed DEC in this category “signaled” in the same year, which was the first year of reporting for the relevant DEC). One possible explanation is that there might have been more intensive reporting for these drugs because of proactive targeted surveillance by the manufacturer. It should be noted that our retrospective analysis may not reflect “real-life” pharmacovigilance in that we preselected our event terms and did not do “open-ended” (i.e., all events exceeding designated thresholds were reviewed for relevance) data mining and that a SDR reflects reporting behavior and may or may not correlate with causality. Currently, drugs in this class may be approved on an accelerated basis, are known to have serious toxicity, are administered to patients with substantial and compli-

cated comorbid illness, are not available to the general medical community and may have a high frequency of “off-label” use. Automated methods of drug surveillance might usefully supplement traditional surveillance strategies for oncology drugs and drugs marketed under similar circumstances.

References

1. DuMouchel W (1999) Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *Am Stat* 53(3):177–190
2. Evans SJW, Waller PC, Davis S (2001) Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf* 10(6):483–486
3. Follman M, Michel A (2003) Proportional reporting ratios for signal detection in the drug safety database of a pharmaceutical company. *Pharmacoepidemiol Drug Saf* 12(suppl 1):S171
4. Gould AL (2003) Practical pharmacovigilance analysis strategies. *Pharmacoepidemiol Drug Saf* 12(7):559–574
5. Hauben M (2003) A brief primer on automated signal detection. *Ann Pharmacother* 37(7/8):1117–1123
6. Kawabe E (2003) Comparison of methodologies for signal detection using Japanese spontaneous reports. *Pharmacoepidemiol Drug Saf* 12(suppl 1):S169
7. Kilgour-Christie J, Czarnecki A, Simmons V (2001) Reporting ratio calculations from a pharmaceutical company database. *Pharmacoepidemiol Drug Saf* 10:S159
8. Ladewski LA, Belknap SM, Nebeker JR, et al (2003) Dissemination of information of potentially fatal adverse drug reactions for cancer drugs from 2000 to 2002: first results from the Research on Adverse Drug Events and Reports Project. *J Clin Oncol* 21:3859–3866
9. Lindquist M, Stahl M, Bate A (2002) A retrospective evaluation of a data mining approach to aid finding new adverse drug reaction signals in the WHO international database. *Drug Saf* 23(6):533–542
10. Roux E, Tubert-Bitter P, Thiessard F, Fourrier A, et al (2003) Evaluation of data mining methods in pharmacovigilance using simulated data sets. *Pharmacoepidemiol Drug Saf* 12(suppl 1):S183
11. Szarfman A., Machado SG, O’Neill RT (2002) 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 Saf* 25(6):381–392
12. U.S. Food and Drug Administration. Center for Drug Evaluation and Research. Adverse Events Reporting System (AERS). <http://www.fda.gov/cder/aers/default.htm>