Use of data mining at the Food and Drug Administration

RECEIVED 8 December 2014
REVISED 25 April 2015
ACCEPTED 10 May 2015
PUBLISHED ONLINE FIRST 23 July 2015

Hesha J Duggirala[1,*], Joseph M Tonning[2], Ella Smith[3], Roselie A Bright[4], John D Baker[1], Robert Ball[5], Carlos Bell[2], Susan J Bright-Ponte[1], Taxiarchis Botsis[5], Khaled Bouri[4], Marc Boyer[3], Keith Burkhart[2], G Steven Condrey[6], James J Chen[7], Stuart Chirtel[3], Ross W Filice[4], Henry Francis[2], Hongying Jiang[8], Jonathan Levine[4], David Martin[5], Taiye Oladipo[3], Rene O'Neill[8], Lee Anne M. Palmer[1], Antonio Paredes[9], George Rochester[9], Deborah Sholtes[9], Ana Szarfman[2], Hui-Lee Wong[8], Zhiheng Xu[8], Taha Kass-Hout[4]

AMIA INFORMATICS PROFESSIONALS. LEADING THE WAY.
OXFORD UNIVERSITY PRESS

ABSTRACT

Objectives This article summarizes past and current data mining activities at the United States Food and Drug Administration (FDA).
Target audience We address data miners in all sectors, anyone interested in the safety of products regulated by the FDA (predominantly medical products, food, veterinary products and nutrition, and tobacco products), and those interested in FDA activities.
Scope Topics include routine and developmental data mining activities, short descriptions of mined FDA data, advantages and challenges of data mining at the FDA, and future directions of data mining at the FDA.

Keywords: data mining, pharmacovigilance, disproportionality analysis

INTRODUCTION

The diverse products regulated by the United States Food and Drug Administration (FDA) represent approximately 25% of the US economy, are used daily, and affect the health of many millions of people and animals. Beyond food and drugs, these products include nutritional supplements, genetically engineered foods, vaccines, artificial hearts, surgical lasers, devices used to administer drugs and biologics, gene therapies, veterinary drugs, pet food, tobacco products, and many others. Adverse events associated with these products are responsible for tremendous public health impacts and financial costs. Adverse-event-related costs impact healthcare product development, health insurance premiums, and healthcare services (eg, hospitalizations), all of which lead to long-term societal losses, such as permanent disability and death.[1] Ensuring the safety of the manifold products that fall under the FDA's regulation is a formidable challenge.

The FDA collects and maintains sets of data that provide safety information for its regulated products. The annual number of such reports received by the FDA has steadily increased over the decades, due to factors such as increases in population, the number and type of FDA-regulated products, awareness of the importance of reporting potentially product-related problems, and the increased ease of submitting reports (eg, online reporting tools). The FDA currently receives more than 2 million adverse event, use error, and product complaint reports each year from consumers, healthcare professionals, manufacturers, and others. These reports are entered into various databases maintained by the FDA so that the agency can perform analyses to identify potential safety issues and enhance understanding of those issues. Since the 1990s, the FDA has been exploring and expanding its use of data mining to:

- analyze increasing numbers of reports;
- speed identification of potential safety issues;
- aid in prioritizing potential safety issues; and
- free personnel to devote a higher proportion of their time to tasks that are not yet readily assisted by machines.

As basic data mining methods have become routine for more and more safety report databases, the FDA has recommended the use of data mining to the drug industry[2] and FDA data mining experts have begun developing more sophisticated methods and applying data mining to other types of product safety FDA and non-FDA databases.

In this paper, we summarize the data mining tools and methods that the FDA currently uses to identify safety signals. We also address the expansion of data mining to include new types of methods and to address additional databases. Because the data and processes depend on regulatory authorities that vary by the type of product being analyzed and undergo constant modification, detailed descriptions of the various data and data mining methods are provided on a new FDA webpage.[3] Both the present article and the FDA webpage incorporate input from all the FDA regulatory centers in addition to the FDA's Office of the Commissioner, which serves all the centers. More details about the FDA's organization are available on its website.[4]

DATA MINING METHODS APPLIED TO SAFETY REPORTS

The FDA's safety reports databases are analyzed with routine and prototype data mining methods and tools, which are described in detail below.

Disproportionality Methods

The FDA largely utilizes disproportionality methods to identify statistical associations between products and events. Such methods compare the observed count of a product-event combination with an "expected" count. Unexpectedly high reporting associations "signal"[5] that there may be a causal association between a particular adverse event and a product. Identified safety signals are referred to as disproportionately reported combinations.

The proportional reporting ratio (PRR) is the foundational concept of many disproportionality methods.[6,7] The PRR is the degree of disproportionate reporting of an adverse event for a product of interest compared to the reporting of this same adverse event for all other products in the database. Thus, the entire database is used as a background

*Correspondence to Hesha Duggirala, 7519 Standish Place, HFV-200, Rockville, Maryland, USA; Tel: +240-402-6218; Hesha.Duggirala@fda.hhs.gov
 For numbered affiliations see end of article

REVIEW

"expected" count. However, because disproportionality methods do not adjust for small observed or expected numbers of reports of the product-event pair of interest, other, more advanced statistical methods are also employed, such as the Multi-Item Gamma Poisson Shrinker (MGPS).[2,8,9]

Various commercially available software programs generate PRR and/or MGPS scores (eg, Empirica Signal™, PV Analyser™, Molecular Analysis of Side Effects [MASE™], and Statistical Analysis Systems™ [SAS™]).

Change-Point Analysis

Change-point analysis (CPA) is a statistical method for determining whether a change in either the slope[10–15] or variability[16] in a time series or sequence in very large databases has taken place. As a complimentary tool to the signal detection efforts at the FDA, CPA could be critical for public health regulation, the surveillance of adverse events and recalls, and regulators' understanding of the longitudinal effect of adverse events that result from the use of regulated products.

Text Mining

Text mining is of interest because a large volume of "unstructured" data (eg, narratives, event descriptions) is submitted as part of adverse event reporting.

The FDA recently developed the Vaccine *a*dverse *e*vent Text Mining (V*ae*TM) text mining system from the FDA vaccines adverse event reports database. V*ae*TM currently extracts diagnostic, treatment, and various assessment information using rules.[17] The newest version of the system (released in summer 2014) includes laboratory test results and temporal information modules; the latter associates the above features on a time axis and provides a critical overview of the adverse events following the administration of not only vaccines but also drugs.

Incorporation of Reference Data Into Data Mining

For drugs, the FDA is evaluating and advising on product development for a proprietary software tool called Molecular Analysis of Side Effects (MASE)™.[18] MASE™ integrates publicly available adverse drug event reports data with various chemical and biological data sources in a drug-centric manner. This software tool is being utilized to assess the biological plausibility of safety signals. The program can identify targets, enzymes, and transporters that are disproportionately associated with drugs and events. This "mechanism mining" tool generates enzymatic, pathway, and molecular target hypotheses that warrant further evaluation. The program was recently used to study infusion reactions.[18]

Beyond the FDA's experiences with geographical information systems (GIS) technology to manage product quality threats resulting from natural disasters,[19] the agency is also exploring GIS technology to enable safety data analysis for routine circumstances. Product surveillance using GIS will allow analysts to capture, store, retrieve, analyze, manage, and display safety data geographically and/or temporally. Tracking potential safety signals in this manner can provide new opportunities for real-time interventions and the identification of:

- populations at risk (eg, those with genetic predispositions to specific adverse events);
- identification of patterns related to intentional or unintentional product contamination; and
- identification of areas where public health education and assistance may be appropriate.

Visualization Tools

Regardless of the analytical tools used, the visualization of data is paramount. Graphical tools that the FDA uses to visualize large and complex volumes of data include heat maps[20] and sector maps.[21] Other visualization tools that allow researchers to closely compare related products and outcomes[21] and that can display contrasting sub-groups[20] are very valuable.

The FDA also uses the network analysis technique, which incorporates automated pattern recognition and has been applied to Vaccine Adverse Event Reporting System (VAERS) data.[22] Another prototype tool, the Adverse Event Network Analyzer,[23] incorporates various algorithms to identify patterns in VAERS data and can support processing other types of data as well.

SAFETY REPORT DATABASES AT THE FDA

Due to the unique analytic needs stemming from both product type characteristics and product type-specific regulatory authorities, no single adverse event database for all products exists at the FDA. Table 1 summarizes those FDA safety report databases for which data mining is used.

The Place of Data Mining in Safety Report Assessment

Data mining analyses are used to detect potential signals and generate hypotheses related to those signals, but cannot be used in isolation to establish causality between an adverse event and a product. There are many possible reasons other than a direct causal relationship for there to be a statistical association between a product and an event[24] (eg, the recent questions about Pradaxa®25). Hands-on case reviews, analysis of other data sources (eg, FDA regulatory databases, the World Health Organization drug safety report database,[26] public scientific literature, and public knowledge databases[27–29]), and further epidemiologic assessments[25,30] are necessary to characterize the clinical and public health significance of signals generated by data mining analyses.[2]

When the evidence for a new safety issue is compelling, the FDA may take regulatory action (such as issuing a product recall or changing a product labeling) and is responsible for informing the public of these actions, along with any firm-initiated communications.

PAST SUCCESSFUL MINING OF SAFETY REPORT DATABASES

Mining the FDA's safety report databases has identified important safety issues in recent years.

The first vaccine safety signal detected with the use of MGPS alone was an association between febrile seizures and Fluzone® 2010–2011 influenza vaccine administration in young children.[31] The signaling threshold, database restrictions, adjustment, and baseline data mining were strategies adopted a priori to enhance the specificity of the data mining analyses of the 2010–2011 influenza vaccine data.

Data mining has assisted in the evaluation of many important drug safety signals, including associations between pituitary tumors and atypical antipsychotics,[32] pathological gambling and Parkinsonian therapy,[21] as well as pancreatitis and atypical antipsychotics and valproic acid.[9] Even data for older drugs may contain hidden signals of toxicity that can be elicited by data mining, as was the case for the association between hepatotoxicity and propylthiouriacil.[20] The importance of evaluating other types of data in conjunction with signals identified by data mining was exemplified in the evaluation of amyotrophic lateral sclerosis's association with statins.[33]

Mining dietary supplement safety report data identified that unusual levels of liver toxicity were associated with Hydroxycut®, a weight-loss dietary supplement. Further investigation of the clinical records of the patients with liver damage who took Hydroxycut® confirmed that the relative timing of Hydroxycut® use and liver damage was consistent with causality, and in most cases, no other cause of liver damage could be found.[34] Hydroxycut® was voluntarily recalled from the market in May 2009 due to hepatic toxicity.[35] Hydroxycut® was subsequently reformulated and remarketed.

Table 1: Data mining of safety reports (reports of adverse events, injuries, death, use errors, and hazardous product qualities) received by the FDA, by type of product, database characteristics, and data mining method

Product type	Database features as of spring 2014			Data mining method	
	Current number of reports received annually	Database start date	Cumulative number of reports	Stage of use	Method or tool
Drugs	770 000 in 2013	1968	>7 000 000	Routine	MGPS with Empirica Signal™
				Developmental	*Vae*™
				Developmental	MGPS with MASE™
				Developmental	GIS
Medical devices	670 000 in 2013	1991	3 300 000	Developmental	CPA
				Developmental	GIS
Vaccines	35 000–40 000	1990	>4 50000	Developmental	*Vae*™
				Routine	MGPS with Empirica Signal™
				Developmental	Adverse Event Network Analyzer
				Developmental	GIS
Foods, cosmetics, and dietary supplements	6000 in 2013	2002	40 500	Routine	MGPS with Empirica Signal™
				Developmental	GIS
Animal drugs and devices	75 000	1991	400 000	Developmental	PRR and MGPS with PV Analyser™
				Developmental	GIS

Notes: CPA, change-point analysis; GIS, geographical information systems; MASE™, Molecular Analysis of Side Effects™; MGPS, Multi-Item Gamma Poisson Shrinker; PRR, proportional reporting ratio; *Vae*™, Vaccine *a*dverse *e*vent Text Mining. Databases that are too small for data mining were excluded. The "Drugs" category includes the following products intended for human use: prescription drugs, over-the-counter drugs, homeopathic drugs, human cellular products, blood derivatives, and products that are a combination medical device and drug. The "Medical Devices" category includes products that are a combination of medical device and drug that are not in the "Drugs" category.

Retrospective data mining of the Manufacturer and User Facility Device Experience database showed that safety signals associated with an implantable cardioverter defibrillator could have been detected as early as March 2006.[36] Using traditional methods, the association between lead fracture and inappropriate shock events and Sprint Fidelis® leads was, instead, detected 10 months later, in January 2007. The manufacturer announced a voluntary market withdrawal in October 2007.

These examples highlight the important role that data mining has played in product safety report surveillance at the FDA.

DATA MINING METHODS APPLIED TO OTHER TYPES OF DATA

Encouraged by the success of using data mining methods for the analysis of safety report data, FDA experts have started to apply data mining techniques to other types of data, as summarized in Table 2.

Disproportionality Analysis of Published Literature

The FDA has partnered with the National Library of Medicine (NLM) to identify disproportionate reporting of drug-adverse event pairs in MEDLINE®, the NLM's publicly available database of over 20 million biomedical abstracted articles and citations. Experts in cognitive science and linguistics from the NLM have mapped the medical subject headings (MeSH) terms[37] used for indexing of citations in MEDLINE® with adverse events terminology in the Medical Dictionary for Regulatory Activities dictionary.[38] MeSH terms related to drug names have been mapped to the Anatomical Therapeutic Chemical Classification System and RxNorm.[28]

The FDA has applied Empirica Study™ and other software packages, such as SAS JMP™ and JReview™, to analyze clinical trial drug data in either new drug applications or supplemental applications. Empirica Study™ interfaces with data that conforms to the standardized Study Data Tabulation Model of the Clinical Data Interchange Standards Consortium data standards to create a broad set of automatically generated analytical outputs and tailor-made, reusable tables and graphs. These outputs have helped reviewers more efficiently analyze potential safety issues in clinical trial data on drugs approved by the FDA.[39]

Text Mining

The FDA has also explored text mining using Linguamatics™ I2E software to study clinical safety based on chemical structure information contained in medical literature. Linguamatics™ I2E enables custom searches using natural language processing to interpret unstructured text. The ability to predict the clinical safety of a drug based on chemical structures is becoming increasingly important, especially when adequate safety data are absent or equivocal.[40]

A semantic text mining tool is currently being researched, with a view to creating a scalable, secure, industrial-scale, and flexible framework for the widest possible variety of text mining applications to reside upon. The Search and Retrieval Framework (SARF), which was developed by the FDA, is now able to both search within any number of available repositories and screen for massive lists of items within those repositories. SARF includes state-of-the-art ontologies maintained by the NLM and the FDA along with general-purpose dictionaries. Additionally, any number of new dictionaries can be added and selected by the user.

For vaccines, the FDA is working with the Innovation Center for Biomedical Informatics at Georgetown University on the development of Georgetown Vaccine Information and Safety Resource (G-VISR) tool.

Table 2: Types of data and the data mining methods used for them at the FDA

Type of data	Stage of use of data mining	Data mining method or tool	Data mining purpose
MEDLINE®	Developmental	Disproportionality analysis	Find drug-adverse event signal pairs
Medical literature	Developmental	Linguamatics I2E natural language processing; using chemical structure information from the medical literature	Study clinical safety
		Georgetown Vaccine Information and Safety Resource tool	Collect molecular and adverse event information
Medical device documents	Developmental	Search and Retrieval Framework semantic text mining	Search within any number of repositories. Screen for massive lists of items within repositories
Clinical study data in drug applications	Routine	Empirica Study™ creation of a wide set of automatically generated analytical outputs and tailor-made, reusable tables and graphs	Save reviewers from having to create the tables and graphs
Social media	Developmental	MedWatcher Social; uses standard product and adverse event dictionaries	Detect adverse events related to medical products
Tobacco documents	Developmental	Topic modeling methods	Characterize documents and estimate topics covered by the documents
Questions received at the Center for Food Safety and Nutrition call center	Developmental	SAS™ data step programming and SAS™ text mining node	Categorize and group the predominant types of questions
	Routine	SAS Enterprise Miner™	Maintain standardized data fields

Note: SAS™, Statistical Analysis Systems™.

G-VISR mines biomedical literature and existing databases to collect molecular and adverse event information related to individual vaccines.

The FDA is also studying the utility of detecting signals from social media data. MedWatcher Social is an exploratory data mining tool that can detect adverse events related to medical products using publicly available data on social media platforms (eg, Twitter, Facebook, health-related web blogs) to curate and map health information to standard product and adverse event dictionaries.[41] MedWatcher Social has the potential to be able to incorporate logarithmic internet search terms in the near future.

The SAS Enterprise Miner™ specialized text mining software package was recently used to perform text mining of consumer, industry, and governmental questions received by the Center for Food Safety and Nutrition's call center. The combination of SAS™ data step programming and the SAS™ text mining node was useful in categorizing and grouping the predominant types of inquiries received.

Text mining also plays an important role in maintaining standardized data fields at the Center for Food Safety and Nutrition call center.

The FDA is also developing behavioral linguistic methods for medical device documents to analyze free text fields and extract manufacturer reporting patterns as well as vector, matrix, and free-space approaches to text association.

Topic Modeling

Topic modeling can be a useful methodology for characterizing document content based on key terms and estimating topics contained within documents. It can also be used to estimate and identify topics from the document, word and phrase content, and cluster documents. For example, documents associated with the topic "menthol" would comprise one cluster. Documents on menthol that describe usage patterns among "youth" would then be a subset of this more general cluster. Specific techniques being explored for tobacco documents include:

- Latent Dirichlet allocation, which identifies topics contained in disparate text. (It is currently being programmed for use.);
- k-methods (k-means and k-nearest neighbor);
- hierarchical clustering;
- latent variable latent semantic analysis; and
- Probabilistic latent semantic analysis.

ADVANTAGES AND CHALLENGES OF DATA MINING

Advantages of Mining Safety Report Databases

The FDA has noted the following advantages of data mining:

- **Standard Processes.** Historically, manual analyses (whether for generating a specific hypothesis, selecting an event codes to analyze, or selecting a case series or cohort by chart review) raised concerns regarding the accuracy, subjectivity, reproducibility, and interpretation of the data used for conducting the analyses. In contrast, because data mining is automated, the outputs produced are systematic and statistically "objective," given the limitations of the data.
- **Simultaneous Analysis**. Data mining calculations are made without a priori hypotheses for every product-event combination across an entire database at once.
- **Efficiency**. The signal scores for all the product-event pairs are computed in minutes, which is much faster than manually requesting traditional computerized exploratory analyses.
- **Prioritization of Investigating Signals**. Data mining enables much easier prioritization of investigating signals based on the seriousness of the event; the magnitude of the data mining scores; the redundancy of clusters of patterns for the product,

product class, and/or indication; and the number of collateral (similar) adverse event terms.

- **Automated Support of Further Signal Investigation:**
  - "Drill down" capabilities enhance manual exploration.
  - Stratification and sub-setting.
  - Observation of signals over time.
  - Identification of complex interdependent factors (eg, concomitant products and/or diseases).
  - Facilitation of the study of product interactions by automatically calculating unusual reporting patterns for patients using multiple products (eg, a drug for hypertension and a pacemaker).
  - Transparency, replication, and collaboration are fostered by detailed audit trails.
  - Identification and correction of data errors.
  - Facilitation of planning database and analytic improvements.
  - Support for understanding the biological plausibility of signals by incorporating reference datasets regarding chemistry and physiology.

Challenges and Data-Mining Mitigations Related to Safety Report Databases

Challenges inherent in safety report databases that limit the interpretability of signals have already been discussed elsewhere[2,42–52] and include:

- missing, incorrect, or vague information;
- separate reports about the same incident (eg, duplicate reports submitted by patient, physician, etc.);
- event may be due to the treated condition, another condition, or another product;
- underreporting due to lack of recognition of a possible product-event association, lack of awareness of reporting expectations or process, fear of litigation, or reporting to another public health organization instead of the FDA;
- over-reporting due to media publicity or litigation; and
- timeliness of reporting and processing.

The FDA's disproportionality and CPA tools work best on databases that use standard terms for the product, event, and co-variables, such as age. Although much of the standardization is done manually, text mining's potential use as a tool to assist with standardization and to use text fields to enhance the content of coded fields is being investigated. The joint use of clustering apparently related products, apparently related events, and standard references for products and events has helped analysts address incorrect or vague information in reports.

Challenges Related to Applying Data Mining to Safety Reports

Specific data mining methodologies and the interpretation of signals requires database-specific understanding of:

- acceptance of foreign (vs. domestic) reports, with different reporting requirements;
- changing reporting requirements over time;
- changing coding dictionaries for products and events resulting in discrepant product names and/or events;
- changing data entry and coding processes;
- inconsistent database structure architecture; and
- malicious reporting and spam.

Signal thresholds are adjusted to account for the severity of the adverse event related to the product and the severity of the condition for which the product is being used. For example, the threshold for evaluating a safety issue for a drug used to treat cancer would be different than the threshold for a drug used to treat acne.[2,9,54]

Additional challenges specific to interpreting signals generated from data mining safety reports include:

- All of the reports represent a reporter's concern that there is a product-event relationship; signals do not reflect actual rates of events per product use.
- The signals are database-specific. The contents of each database are functions of separate regulatory authorities, rather than simple inherent affinities.

Advantages and Challenges of Using Data Mining for Other Data Types

Data mining of other sources, such as medical literature, electronic health records, and social media, shares many of the challenges related to safety reports data. The quality of data in these sources can be better or worse, depending on the structure of the database and the training of those who enter the data, varying from presumably high-quality MeSH® indexes of MEDLINE®[37] to social media (eg, Facebook)[41] and web blogs.

THE FUTURE OF DATA MINING AT THE FDA

Analytic challenges will continue to grow with the addition of new surveillance data sources and the development of new methods of submitting spontaneous reports, such as web-based and mobile applications. It will be important for the FDA to structure its information technology systems so that data can be submitted, retrieved, processed, and evaluated in a standardized manner.

There will be vast increases and changes in surveillance data that are available in the near future. These include electronic health records,[56] personal health records, claims,[57] standards for health data,[55] data from federal and private sector mobile devices for tracking health,[58–60] and data from social websites (blogs, patient advocacy group sites, and search term logs).[41,61]

Outside research has shown that these data sources can be of value in post-market safety surveillance and other related fields;[41,52,56,61] the FDA would like to validate their utility for the surveillance of FDA-regulated products.

Further development and implementation of an advanced and integrated safety data mining system supported by appropriately experienced personnel will be essential for better informed decision making and risk management of product safety issues in real time. Specific desired data mining capabilities include:

- scalability to accommodate growing databases;
- further advanced natural language processing and text mining to automatically and accurately extract meaning from narratives in all sorts of databases;
- data processing that is very quick or methods that require less data processing, to move surveillance closer to real time;
- complete reference databases for topics including:
  - product characteristics;
  - event characteristics;
  - physiology; and
  - toxicology.
- additional advanced visual analytics with more advanced drill down functions coupled to context information across multiple data resources;

- More transparent human-readable audit trails to enable analysts to more efficiently communicate and validate each other's selection criteria, results, and interpretation.

As a result of these developments, researchers and policy makers will be better equipped to understand the limitations and biases of the data, leading to more informed decisions regarding FDA-regulated product safety.

Data mining holds promise for other FDA work, including:

- FDA field work. Potential uses include exploring trends in safety, inspection, and recalls data so that field managers can more effectively align available personnel and resources to have the greatest impact on public health. Data mining could also assist in enforcement coordination among district and headquarters personnel.
- Pre-approval safety reviews and efficacy evaluations of products.
- Information contained in tobacco health documents, including legal documents and research reports on a range of topics, including:
  - dose response relationships;
  - chemosensory effects;
  - neurobiology of dependence;
  - menthol–nicotine interactions;
  - product-related interactions;
  - advertising-related perceptions;
  - marketing strategies;
  - switching rates; and
  - initiation and cessation rates.

The FDA Data Mining Council, composed of the authors and other interested staff from across the FDA, promotes the improvement of data mining to support the FDA's mission of protecting and promoting public health. We advocate for sharing expertise among other government agencies, academia, and private sector companies to increase knowledge about data mining and improve data analysis.

CONTRIBUTORS

H.J.D. and J.M.T. originally conceived of the manuscript and identified programs. H.J.D., J.M.T., and E.S. combined individual sections into one coherent manuscript. R.A.B. significantly revised draft paper. All authors provided substantial contributions to the conception of the work. All authors participated in drafting the work or revising it critically for important intellectual content. All authors provided final approval of the version to be published. All authors agree to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

FUNDING

The research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

COMPETING INTERESTS

None.

REFERENCES

1. Reducing and Preventing Adverse Drug Events To Decrease Hospital Costs: Research in Action, Issue 1. March 2001. Agency for Healthcare Research and Quality, Rockville, MD. http://www.ahrq.gov/legacy/qual/aderia/aderia.htm. Accessed December 1, 2014.
2. Guidance for Industry. Good Pharmacovigilance Practices and Pharmacoepidemiologic Assessment. Food and Drug Administration, US Department of Health and Human Services, March, 2005. http://www.fda.gov/downloads/regulatoryinformation/guidances/ucm126834.pdf. Accessed December 2014.
3. Data Mining at FDA. http://www.fda.gov/datamining. Accessed June 10, 2015.
4. FDA Organization. http://www.fda.gov/AboutFDA/CentersOffices/default.htm. November 4, 2014. Accessed February 15, 2015.
5. Waller PC, Evans SJ. A model for the future conduct of pharmacovigilance. *Pharnaccoepidemiol Drug Saf.* 2003;12(1):17–29.
6. Evans SJ, Waller PC, Davis S. Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol Drug Saf.* 2001;10(6):483–486.
7. Finney DJ. Systemic signalling of adverse reactions to drugs. *Methods Inf Med.* 1974;13(1):1–10.
8. Bate A, Evans, S. Quantitative signal detection using spontaneous ADR reporting. *Pharmacoepidemiol and Drug Saf.* 2009;18(6):427–436.
9. Szarfman A, Tonning JM, Doraiswamy PM. Pharmacovigilance in the 21st century: new systematic tools for an old problem. *Pharmacotherapy.* 2004;24(9):1099–1104.
10. Kass-Hout TA, Xu Z, McMurray P, *et al.* Application of change point analysis to daily influenza-like illness emergency department visits. *J Am Med Inform Assoc.* 2012;19(6):1075–1081.
11. Kass-Hout TA, Xu Z. Change point analysis. https://sites.google.com/site/changepointanalysis. Accessed February 15, 2015.
12. Edwards AW, Cavalli-Sforza LL. A method for cluster analysis. *Biometrics.* 1965;21:362–375.
13. Auger IE, Lawrence CE. Algorithms for the optimal identification of segment neighborhoods. *Bull Math Biol.* 1989;51(1):39–54.
14. Bai J, Perron P. Estimating and testing linear models with multiple structural changes. *Econometrica.* 1998;66(1):47–78.
15. Killick R, Fearnhead P, Eckley IA. Optimal detection of changepoints with a linear computational cost. *JASA.* 2012;107(500):1590–1598.
16. Killick R, Eckley IA. changepoint: an R package for changepoint analysis. 2012. http://www.lancs.ac.uk/∼killick/Pub/KillickEckley2011.pdf. Accessed February 14, 2015.
17. Botsis T, Buttolph T, Nguyen M, *et al.* Vaccine adverse event text mining system for extracting features from vaccine safety reports. *J Am Med Inform Assoc.* 2012;19(6):1011–1018.
18. Moore PW, Burkhart KK, Jackson D. Drugs highly associated with infusion reactions reported using two different data-mining methodologies. *J Blood Disorders Transf.* 2014;5:195.
19. FDA's Geographic Information System. http://www.fda.gov/AboutFDA/CentersOffices/OC/OfficeoftheCounselortotheCommissioner/ucm227114.htm. Last updated April 2, 2012. Accessed February 15, 2015.
20. Rivkees SA, Szarfman A. Dissimilar hepatotoxicity profiles of propylthiouracil and methimazole in children. *J Clin Endocrinol Metab.* 2010;95(7):3260–3267.
21. Szarfman A, Doraiswamy PM, Tonning JM, *et al.* Association between pathologic gambling and Parkinsonian therapy as detected in the Food and Drug Administration Adverse Event database. *Arch Neurol.* 2006;63(2):299–300.
22. Ball R, Botsis T. Can network analysis improve pattern recognition among adverse events following immunization reported to VAERS? *Clin Pharmacol Ther. 2011*;90(2):271–278.
23. Botsis T, Scott J, Goud R, *et al.* Novel algorithms for improved pattern recognition using the US FDA adverse event network analyzer. *Stud Health Technol Inform.* 2014;205:1178–1182.
24. Almenoff J, Tonning JM, Gould AL, *et al.* Perspectives on the use of data mining in pharmacovigilance. *Drug Saf.* 2005;28(11):981–1007.
25. Pradaxa (dabigatran): Drug Safety Communication - Lower Risk for Stroke and Death, but Higher Risk for GI Bleeding Compared to Warfarin. Posted May 13, 2014. http://www.fda.gov/safety/medwatch/safetyinformation/safetyalertsforhumanmedicalproducts/ucm397179.htm. Accessed January 14, 2015.
26. Vigibase. World Health Organization. http://who-umc.org/DynPage.aspx?id=98082&mn1=7347&mn2=7252&mn3=7322&mn4=7326. Last updated December 19, 2014. Accessed February 15, 2015.
27. TOXNET databases. U.S. National Library of Medicine. http://toxnet.nlm.nih.gov/. Accessed February 15, 2015.

28. Medical terminologies at NLM. U.S. National Library of Medicine. http://www.nlm.nih.gov/medical-terms.html. Last reviewed December 2, 2013. Accessed February 15, 2015.
29. DAILYMED. U.S. National Library of Medicine. http://dailymed.nlm.nih.gov/dailymed/index.cfm. Accessed February 15, 2015.
30. Welcome to Mini-Sentinel. Food and Drug Administration. http://www.mini-sentinel.org/. Last updated October 15, 2014. Accessed February 15, 2015.
31. Martin D, Menschik M, Bryant-Genevier M, *et al.* Data mining for prospective early detection of safety signals in the Vaccine Adverse Event Reporting System (VAERS): a case study of febrile seizures after a 2010-2011 seasonal influenza virus vaccine. *Drug Saf.* 2013;36(7):547–556.
32. Szarfman A, Tonning JM, Levine JG, *et al.* Atypical antipsychotics and pituitary tumors: a pharmacovigilance study. *Pharmacotherapy.* 2006;26(6):748–758.
33. Colman E, Szarfman A, Wyeth J, *et al.* An evaluation of a data mining signal for amyotrophic lateral sclerosis and statins detected in FDA's spontaneous adverse event reporting system. *Pharmacoepidemiol Drug Saf.* 2008; 17(11):1068–1076.
34. Fong T-L, Klontz KC, Canas-Coto A, *et al.* Hepatoxicity due to Hydroxycut®: a case series. *Am J Gastroenterol.* 2010;105(7):1561–1566.
35. Warning on Hydroxycut. FDA. January 20, 2015 http://www.fda.gov/ForConsumers/ConsumerUpdates/ucm152152.htm. Accessed February 15, 2015.
36. Duggirala HJ, Herz ND, Caños DA, *et al.* Disproportionality analysis for signal detection of implantable cardioverter-defibrillator-related adverse events in the Food and Drug Administration Medical Device Reporting System. *Pharmacoepidemiol Drug Saf.* 2012;21(1):87–93.
37. Fact Sheet: Medical Subject Headings (MeSH®). U.S. National Library of Medicine. http://www.nlm.nih.gov/pubs/factsheets/mesh.html September 12, 2013. Accessed February 15, 2015.
38. Welcome to MedDRA. ICH Steering Committee. http://www.meddra.org/. Accessed February 15, 2015.
39. Ana Szarfman. Medical Officer's Consultative Reanalysis of the Febrile Neutropenia Studies of NDA 50-679. http://www.fda.gov/downloads/Drugs/DrugSafety/PostmarketDrugSafetyInformationforPatientsandProviders/DrugSafetyInformationforHeathcareProfessionals/UCM201520.pdf. Accessed December 1, 2014.
40. Botsis T, Ball R. Automating case definitions using literature-based reasoning. *Appl Clin Inform.* 2013;4(4):515–527.
41. Freifeld CC, Brownstein JS, Menone CM, *et al.* Digital drug safety surveillance: monitoring pharmaceutical products in twitter. *Drug Saf.* 2014;37(5): 343–350.
42. Tsong Y. Comparing reporting rates of adverse events between drugs with adjustment for year of marketing and secular trends in total reporting. *J. Biopharm. Stat.* 1995;5:95–114.
43. Hazell L, Shakir SA. Under-reporting of adverse drug reactions: a systematic review. *Drug Saf.* 2006;29(5):385–396.
44. Waller PC. Measuring the frequency of adverse drug reactions. *Br J Clin Pharmac.* 1992;33(3):249–252.
45. Meinzinger MM, Barry WS. Prospective study of the influence of the media on reporting medical events. *Ther Innov Regul Sci.* 1990;24(3): 575–577.
46. McAdams M, Staffa J, Dal Pan G. Estimating the extent of reporting to FDA: a case study of statin-associated rhabdomyolysis. *Pharmacoepidemiol Drug Saf.* 2008,17(3):229–239.
47. Graham DJ, Campen D, Hui R, *et al.* Risk of acute myocardial infarction and sudden cardiac death in patients treated with cyclo-oxygenase 2 selective and non-selective non-steroidal anti-inflammatory drugs: nested case-control study. *Lancet.* 2005;365(9458):475–481.
48. Bright RA. Surveillance of adverse medical device events. In: Brown SL, Bright RA, Tavris DR, eds. *Medical Device Epidemiology and Surveillance.* London, UK: John Wiley & Sons, Ltd.; 2007:43–61.
49. Balka E, Doyle-Waters M, Lecznarowicz D, *et al.* Technology, governance, and patient safety: Systems issues in technology and patient safety. *Int J Med Inform.* 2007:76 (Suppl 1):S35–S47.
50. Samore MH, Evans RS, Lassen A, *et al.* Surveillance of medical device-related hazards and adverse events in hospitalized patients. *JAMA.* 2004; 291(3):325–334.
51. Medical devices: early warning of problems is hampered by sever underreporting. US General Accounting Office. GAO/PEMD 87-1; 1987.
52. Hefflin B, Gross T, Schroeder T. Estimates of medical device-associated adverse events from emergency departments. *Am J Prev Med.* 2004;27(3): 246–253.
53. Bright RA, Nelson RC. Automated support for pharmacovigilance: a proposed system. *Pharmacoepidemiol Drug Saf.* 2002;11(2):121–125.
54. Update on the adoption of health information technology and related efforts to facilitate the electronic use and exchange of health information. Report to Congress. Office of the National Coordinator for Health Information Technology, US Department of Health and Human Services. October 2014. http://www.healthit.gov/sites/default/files/rtc_adoption_and_exchange9302014.pdf. Accessed February 1, 2015.
55. Zhan C, Kaczmarek R, Loyo-Berrios N, *et al.* Incidence and short-term outcomes of primary and revision hip replacement in the United States. *J Bone Joint Surg Am.* 2007;89(3):526–533.
56. Deering MJ. Issue brief: patient-generated health data and health IT. Office of the National Coordinator for Health Information Technology, US Department of Health and Human Services. http://www.healthit.gov/sites/default/files/pghd_brief_final122013.pdf. December 20, 2013. Accessed February 15, 2015.
57. Sands DZ, Wald JS. Transforming health care delivery through consumer engagement, health data transparency, and patient-generated health information. *Yearb Med Inform.* 2014;9(1):170–176.
58. HealthData.gov. US Department of Health and Human Services. http://www.healthdata.gov/. Accessed February 15, 2015.
59. National Electronic Injury Surveillance System (NEISS). US Consumer Product Safety Commission. http://www.cpsc.gov/en/Research–Statistics/NEISS-Injury-Data/. Accessed February 15, 2015.
60. Ginsberg J, Mohebbi MH, Patel RS, *et al.* Detecting influenza epidemics using search engine query data. *Nature.* 2009;457(7232): 1012–1014.
61. White RW, Harpaz R, Shah NH, *et al.* Toward enhanced pharmacovigilance using patient-generated data on the internet. *Clin Pharmacol Ther.* 2014; 96(2):239–246.

AUTHOR AFFILIATIONS

[1]Center for Veterinary Medicine, FDA
[2]Center for Drug Evaluation and Research, FDA
[3]Center for Food Safety and Applied Nutrition, FDA
[4]Office of the Commissioner, FDA
[5]Center for Biologics Evaluation and Research, FDA
[6]Office of Regulatory Affairs, FDA
[7]National Center for Toxicological Research, FDA
[8]Center for Devices and Radiological Health, FDA
[9]Center for Tobacco Products, FDA