Toward Enhanced Pharmacovigilance using Patient-Generated Data on the Internet

Ryen W. White[*,1], **Rave Harpaz**[*,†,2], **Nigam H. Shah**[2], **William DuMouchel**[3,4], and **Eric Horvitz**[1]

[1]Microsoft Research, Redmond, WA

[2]Center for Biomedical Informatics Research, Stanford University, Stanford, CA

[3]Oracle Health Sciences, Burlington, MA

[4]Observational Medical Outcomes Partnership, Foundation for the National Institutes of Health, Bethesda, MD

Abstract

The promise of augmenting pharmacovigilance with patient-generated data drawn from the Internet was called out by a scientific committee charged with conducting a review of FDA's current and planned pharmacovigilance practices. To this end, we present a study on harnessing behavioral data drawn from Internet search logs to detect adverse drug reactions (ADRs). By analyzing search queries collected from 80 million consenting users and by using a widely recognized benchmark of ADRs, we find that the performance of ADR detection via search logs is comparable and complementary to detection based on FDA's adverse event reporting system (AERS). We show that by jointly leveraging data from AERS and search logs, the accuracy of ADR detection can be improved by 19% over the use of each data source independently. The results suggest that leveraging nontraditional sources, such as online search logs, could supplement existing pharmacovigilance approaches.

Keywords

drug safety; pharmacovigilance; adverse event reporting system; Internet search logs; patient-generated data; behavioral data

Introduction

Adverse drug reactions (ADRs) are the fourth leading cause of death in the United States, ahead of pulmonary disease, diabetes, HIV, and automobile accidents.[1-4] Beyond deaths,

adverse reactions cause millions of injuries across the world each year and billions of dollars in associated costs. Numerous ADRs could be prevented with more accurate and timely detection. Drug safety surveillance, or *pharmacovigilance*, targets the detection, assessment, and prevention of ADRs in the post-approval period. To date, pharmacovigilance programs, such as the US Food and Drug Administration (FDA) Adverse Event Reporting System (AERS), rely on spontaneous reporting. AERS pools reports of suspected ADRs collected from healthcare professionals, consumers, and pharmaceutical companies. The reports are used to identify and investigate safety concerns about drugs and to provide guidance for regulatory actions including issuing warnings, mandating label changes, and suspending the use of medications. Increasingly, statistical analyses of AERS data are being used to identify signals of potential ADRs[5]. While AERS has been invaluable and will continue to be a major source of information on adverse reactions, analysis of spontaneous reports is only one aspect of the developing science of pharmacovigilance. Recent high profile cases of ADRs, such as the overdue identification of heart attack risk associated with Vioxx and the inconsistent evidence that led to confusion over the safety of Avandia, along with recognized limitations of spontaneous reporting, highlight the need to devise more comprehensive approaches to pharmacovigilance, which would span and leverage scientific insights about ADRs from multiple complementary data sources[6-9].

Recent directions in pharmacovigilance focus on the expanded secondary use of electronic health records and medical insurance claims[6,8,10]. Ongoing efforts rely on analyzing clinical trials data, and the use of information related to mechanistic pharmacology and pharmacogenetics. In addition, there is a recognized need[11] to harness non-traditional resources that are generated by patients via the Internet, including online social media (e.g., patients' experiences with medications that are explicitly shared via online health forums and social networks[4,12,13]) and implicit health information contained in the logs of popular search engines.

Anonymized Internet search logs can serve as a planetary-scale sensor network for public health, identifying informative patterns of health information-seeking about medications, symptoms, and disorders. A study conducted in 2009 by the Centers for Disease Control and Prevention estimated that 61% of adults search the Web for health and medical related information[14]. Another study by the Pew Research Center in early 2013 reported that 72% of Internet users claimed to search online for health information, and that 8 in 10 online health inquiries start at a search engine[15]. Search logs are used in the Google Flu Trends project, demonstrating that statistics of influenza-related search terms recorded by search engines can be used to track rates of influenza[16]. Similarly, analyzing search queries about medications and medical conditions may provide early clues about ADRs.

The present work builds on an earlier study[17], demonstrating that large-scale analysis of Internet search queries can accurately signal drug interactions associated with hyperglycemia. The work has framed efforts to create a log-analysis tool at Microsoft Research, named the Behavioral Log-based Adverse Event Reporting System (BLAERS), a prototype system that can provide ongoing monitoring and exploration of ADRs from search logs.

Here, we present new findings and developments in the design, evaluation, and value of a surveillance system based on Internet search logs. We introduce a new approach for systematic signal detection using those logs, and further evaluate the potential value of search log data as a resource for generating early warnings about ADRs by using a large complement of drugs and outcomes.

The data used in this study comprises 18 months of Internet search logs from 2011 to 2013, collected from over 80 million users of a Web browser add-on from Microsoft. The logs were sourced from users who had consented to their collection and use when they installed browser software (an IRB was not required). The add-on recorded these users' search queries on the Google, Bing, and Yahoo! search engines and the URLs of the Web pages that they visited during this time period. An anonymous identifier, connected to the instance of the browser add-on, was used to track queries. All analyses were performed at the aggregate level across thousands of searchers, and no attempts were made to identify individual searchers from the logs.

The model of human behavior assumed in this work is that people search for information about drugs they are taking (or have been prescribed), and at a later time point search for symptoms or conditions they experience that may be linked to the drugs as potential adverse events. The links were inferred from longitudinal analysis of sequences of queried search terms corresponding to drugs, medical conditions, and their related symptoms. The inclusion of symptoms allows the identification of search behavior associated with conditions that could be related to drug consumption but where the user may not yet have been professionally diagnosed. Queried terms corresponding to the drugs, conditions, and symptoms of interest were identified using sets of synonyms automatically generated from medical ontologies and historical search-result click data. A methodology inspired by self-controlled study designs[18] was used to analyze the longitudinal sequences and estimate statistical associations between drugs and outcomes of interest. The associations were estimated by comparing aggregated query rates for a condition in a surveillance period after and before a drug was first queried for by each user. Signals were quantified by a statistic called the *query rate ratio* (*QRR*). We structured the observation period to raise the likelihood that terms associated with searches on symptoms and disorders are based in symptoms that have been experienced.

Signal detection accuracy was evaluated on the basis of correctly classifying 398 test cases (drug-outcome pairs) deemed as either true ADRs or negative controls (spurious ADRs) that comprise a recognized drug safety gold standard created by the Observational Medical Outcomes Partnership (OMOP)[8,19]. The gold standard includes 181 drugs covering non-steroidal anti-inflammatory drugs, antibiotics, antidepressants, ACE inhibitors, beta blockers, antiepileptics, and glucose lowering drugs, and is divided into four sets of test cases corresponding to one of four outcomes: *acute myocardial infarction*, *acute renal failure*, *acute liver injury*, and *upper gastrointestinal bleeding*, which represent four significant and actively monitored adverse events[20]. The results of this evaluation were compared with the accuracy of signal detection based on FDA's AERS. Last, we investigated the potential of Internet search logs to augment AERS-based surveillance by

evaluating a signal detection strategy that combines signals generated by jointly leveraging data from AERS and search logs.

Results

For the comparative evaluation, AERS signal scores (association statistics) for the same set of OMOP test cases were obtained from a recent study, characterizing the performance of signal detection based on AERS[5]. The study was based on almost the entire set of public domain AERS reports available to date (approximately five million reports).The AERS signals used for comparison in this study were generated by FDA's primary signal detection algorithm called the Multi-item Gamma Poisson Shrinker (MGPS)[21].

The association statistics used in the current evaluation are denoted by *EB05* and *QRR05*, and represent the lower 5th percentile of the observed-to-expected ratio distribution calculated by MGPS, and the lower 5th percentile of the *QRR* distribution respectively. The use of lower bound association statistics instead of point estimates is a recommended adjustment commonly applied by safety evaluators at the FDA[22] to reduce false signaling. In the case of AERS, this adjustment has been shown to provide greater accuracy than point estimates[5], and the same result was observed in this study for the *QRR* statistic.

Performance (signal detection accuracy) was measured based on the area under the receiver operating characteristic (ROC) curve (AUC). The evaluation and comparison was performed for each of the four OMOP outcomes separately. Of the original 398 OMOP test cases, the evaluation was restricted to a subset of 325 test cases (Table 1), for which there was at least one AERS report, and for which at least 50 distinct users queried for a given drug-outcome pair of interest (test case).

Table 2 and Figure 1 summarize the main results. Based on the 325 test cases, the performance of signal detection using search logs ranges from an AUC of 0.73 for acute myocardial infarction, to an AUC of 0.92 for upper gastrointestinal bleeding, with an average AUC of 0.83 for the four outcomes analyzed. The traditional analysis on AERS data attained an average AUC of 0.81. The relative AUC differences between the two data sources ranges from 4% in favor of AERS for acute renal failure, to 29% in favor of search logs for upper gastrointestinal bleeding, with an average relative difference of 11% in favor of search logs for the four outcomes investigated. The relative AUC difference is defined as the proportion of error reduction gained by using one data source over the other (formal definition in Methods).

The ROC curves of AERS and search logs (Figure 1) demonstrate that the two data sources have different operating characteristics, providing different tradeoffs in terms of sensitivity and specificity. Given that false alerts may compromise the value of a surveillance system, it has been advised that false positive rates (FPR) should be given key consideration in the assessment of a signal detection system[23-25]. Accordingly, partial-AUC analysis at 0.3 FPR (specificity>0.7), a suggested ROC region of clinical relevancy for signal detection assessment[26], shows (Table 2) that search logs generally perform better than AERS in this restricted ROC space and may improve upon AERS by an average of 12% for the four

outcomes analyzed. Establishing statistical significance of the *differences* in the observed AUC (see Methods) was not attainable (p>0.05). Thus, it can be argued that the accuracy of signals from traditional AERS analysis and search logs are comparable.

We explored the opportunity to harness analyses of search logs to complement and extend traditional AERS analysis. Table 3 shows that combining signals (association statistics) from AERS and search logs results in a substantial improvement in detection accuracy, averaging 19% (full-AUC) and 19% (partial-AUC), over the use of each source separately. In this case, the AUC improvements are statistically significant (p<0.05). The signals were combined through inverse variance weighting of signal-score point estimates (see Methods), and by using the lower 5th percentile of the weighted average distribution as a composite signal-score (denoted IVW05).

Supplementary Table S1 provides the signal statistics underlying the results of this study.

Discussion

It is widely acknowledged that no single data source or analytic approach would adequately address the need for more effective ADR detection. Progress in pharmacovigilance is likely to come via approaches that can effectively integrate safety evidence from multiple complementary data sources.

Search logs may provide early clues about ADRs as patients engage search engines to learn about medications that they are using and medical conditions they experience—effectively linking drugs and potential adverse events over time. The need to augment pharmacovigilance with safety evidence from search logs was recommended by a scientific committee reviewing the FDA's current and planned pharmacovigilance practices[11]. To this end, we present a study that informs the design of a signal detection system based on search logs and that systematically evaluates its potential value for use in pharmacovigilance.

Establishing baseline performance characteristics is essential for understanding how a surveillance system might perform in identifying future unknown ADRs. Our results suggest that a surveillance system based on Internet search logs can attain a relatively high degree of accuracy (average AUC of 0.83) in signaling true ADRs as well as differentiating them from likely spurious ones, with expected performance comparable to ADR detection based on AERS. The results also suggest that signals related to upper gastrointestinal bleeding and acute myocardial infarction can be detected more accurately through search logs than through AERS. Given the general consensus that AERS is better suited for surveillance of rare events than for events with a high background rate[5,27] such as myocardial infarction, the greater accuracy of search logs for detecting myocardial infarction further underscores the promise of using search logs for pharmacovigilance.

Supporting the vision of a computationally integrative approach to pharmacovigilance, we have shown that a systematic integration of signals from both AERS and search logs improves detection accuracy by an average of 19% over the use of each data source independently. Two earlier studies demonstrated similar potential by combining signals from AERS and observational data[28,29]. Despite these promising results, further research is

needed to understand the relative benefits and limitations of each data source, and to fully realize an integrative strategy to pharmacovigilance based on the fusion of multiple sources of data, including: clinical narratives[30], the biomedical literature[31], biological/chemical data[32,33], the social media[12], and search logs.

The OMOP gold standard is a widely acknowledged benchmark to systematically evaluate the accuracy of a pharmacovigilance signal detection system[5,34,35]. Albeit, the gold standard consists of test cases that were publically known during the time frame of our evaluation, and thus may be insufficient to evaluate real-world performance characteristics where emerging or unknown ADRs are targeted. Further, the public availability of knowledge about ADRs may affect reporting, search, and prescription patterns, which in turn could bias evaluation that is retrospective in nature. Consequently, and despite our efforts to mitigate this publicity bias (discussed below), the absolute performance metrics we report may be optimistic with respect to how we should anticipate performance for future safety issues. While a limited number of studies have proposed prospective evaluation strategies[27,36,37], which to some extent could address these issues, there are currently no established guidelines and appropriate benchmarks to do so with high fidelity. The lack of such benchmarks can partially be attributed to the challenges in ascertaining causality for relatively new associations and identifying the time in which they become publically known. Notwithstanding, these limitations are increasingly acknowledged by the drug safety community with several efforts (including the authors herein) to outline a comprehensive evaluation strategy, and a gold standard for that purpose. Relatedly, while seeking to characterize performance independent of a specific threshold implementation, we acknowledge that in real settings and in future evaluations a signaling threshold will need to be identified. Harpaz et al.[5] outline approaches for optimal threshold identification that depend on a stakeholder's tolerance for false positives, which we plan to pursue in future evaluations.

Detecting signals from search logs for pharmacovigilance requires consideration of biasing factors, noise, and uncertainties about such influences on queries as experiences, vocation, interests, and exposure to online content. Having each searcher serve essentially as their own control in our analyses mitigates certain confounding biases such as those associated with demographic factors, health status, and search habits. Users may search on medications, symptoms, and disorders for a variety of reasons, beyond the case where they are taking a medication and experiencing symptomatology. For example, healthcare professionals may routinely search for medical information. We developed a method for automatically identifying and excluding healthcare professionals (9% of the user population exhibiting index search events) from our analyses. There is also uncertainty about the alignment of the timing of a first search on a medication with its initial use. Users consuming a medication may experience symptoms before or after issuing search queries on the medication.

We took two steps to reduce the likelihood that reviewing online content influenced searches on adverse effects: (1) we enforce an additional gap between the last drug query and the first symptom/condition query, and (2) we ignore symptom/condition queries between the first and last drug query to remove instances of cycling between drugs and symptoms/conditions in exploratory searches. The exclusion period established by these two steps is mirrored

symmetrically to the time period before the first drug search, thereby ignoring symptom/condition queries appearing a short time before the drug appears in the logs.

To understand the potential confounding influence of exploratory searches on medications to subsequent exploratory (versus experiential) searches on symptoms or conditions, we sampled 1,000 online searchers and recorded the content of all Web pages that they visited during the exclusion period described above. We found that only 1.4% of searchers who later queried for a symptom/condition of interest had previously visited pages containing content on these symptoms or conditions, increasing our confidence that observed symptom/condition searching is related to experiencing a condition rather than motivated by the prior review of online content. More research is needed to understand the degree of influence of Web page content on search behavior. We are pursuing enhanced inference procedures to distinguish scenarios where users are experiencing adverse effects, versus performing more general explorations of conditions that have been linked to their medications, and stress that ascertaining drug exposure and outcome occurrence can at best be achieved by analyzing textual cues. Attempts to identify and contact users to validate exposure/outcome is prohibited per the terms under which the data is collected. Additional studies are also needed to understand the appearance and timing of search on medications to the time that patients have been prescribed medications and the influence of the appearance of symptomatology to the first and later searches on the medications.

Our assumptions about search behavior effectively set the stage for a longitudinal observational study, but may result in loss of valuable information contained in explicit searches for side effects, e.g., "piroxicam induced heart attack" or queries including both a drug and outcome, which fall into the exclusion period. Analyzing the gaps left by discarding explicit searches or developing a signaling strategy that leverages these searches merits further research.

The methods and results that we have described highlight the value of harnessing aggregations of online behavioral signals for pharmacovigilance. The wide dispersal and ubiquity of information seeking via online search provides a large-scale and anonymized sensor network for public health, with streams of data that complement the collection and analysis of spontaneous reports by the FDA. We believe continuing efforts on harnessing these and other non-traditional data streams will result in the earlier identification of adverse side effects of medications.

Methods

Concept definitions &term recognition

Each outcome in the OMOP gold standard is defined by a set of SNOMED CT[38] concept codes (definitions supplied by OMOP). Drugs in the gold standard are specified at the ingredient level by RxNorm[39] concept codes. An initial set of synonyms for each OMOP drug and outcome concept was obtained from BioPortal, a repository of over 300 biomedical ontologies that provides mappings among synonymous medical concepts[40].For each concept, the initial sets of synonyms were supplemented with consumer-oriented search terms derived from Bing's query-click logs. Additional terms to include were identified by

first identifying all results clicked for a certain query, and then identifying other queries that lead to the same pages (e.g., "bleeding stomach ulcers" for the concept, upper gastrointestinal bleeding)[41]. For each condition, we also identified a set of symptoms via a literature review (e.g., "tarry feces" for upper gastrointestinal bleeding) and used the processes described above to generate synonyms for the symptoms. Automated term recognition was then used to tag queried search terms associated with each of the OMOP drugs and condition concepts (and symptoms thereof). Figure 2 displays the top 10 queried terms associated with each of the four OMOP conditions. Although these terms were derived directly from the OMOP definitions, the choice of search terms used for signal detection may influence performance[16,17].

Excluded users

Users linked to 1000 or more search queries on any given day were classified as automated traffic (Internet bots) and removed. We found that the percentage of a user's queries containing a medical term within their first month of search activity could help identify healthcare professionals. We removed users with a percentage exceeding 20% as likely being healthcare professionals. The percentage threshold was derived from a predictive model (logistic regression) using various search statistics related to medical terms as potential predictors. Model selection and validation was based on 10-fold cross-validation using a manually-labeled sample of 170 users proportionally allocated to each of the percentage deciles. The model had an error rate of 20% in classifying healthcare professionals.

Search logs signal generation

Without loss of generality let D, C, S be the set of terms (synonyms) associated with a specific drug of interest, a specific condition of interest, and a symptom of the condition respectively. Denote by $q_i^{(t)}$ a queried search term issued by user i in time t. Let $T_i^0 = \min \{t | q_i^{(t)} \in D\}$ be the time of the first query for the drug of interest (time-zero), and $T_i^* = \max \{t | q_i^{(t)} \in D\}$ be the time of the last query for the drug. Let $\alpha = T_i^* - T_i^0$, β and γ be two pre-specified parameters, and $\theta = \alpha + \beta + \gamma$. The full surveillance period is then defined as $[T_i^0 - \theta, T_i^0 + \theta]$, and the exclusion period as $[T_i^0 - (\alpha+\beta), T_i^0 + (\alpha+\beta)]$ (Figure 3). The surveillance period restricts the length of observation for which a condition may be regarded as linked to a drug. The exclusion period is a time window in which we ignore all queries of conditions or symptoms to reduce the likelihood that these queries are part of an exploratory search or influenced by review of online content. α is assumed to be a time period where users may be cycling between drugs and symptoms/conditions in exploratory searches. The mean for α was 1.98 days, and its median was 1 day. β is an additional gap between T_i^* and the remaining observation period that reduces the likelihood that online information on adverse effects would have influenced follow-on searches. γ is the time duration beyond the exclusion period in which queries are included in the analysis. Notice that both periods are symmetric around time zero, the index event of the first drug search. We experimented with different values of the parameters β (1-10 days) and γ (30/60/90 days), and found that $\beta = 7$ days and $\gamma = 60$ days yield the best performance with respect to the four outcomes

analyzed. However, it is likely that different observation periods (defined by γ) would be required to detect other events, e.g., events with longer onset.

A statistical association between a drug-condition pair of interest is estimated by comparing the aggregate query rates for a condition or symptom of interest in the inclusion periods after and before time zero for the drug of interest (see Figure 3). Specifically, let

$$N_i^+ = \# \left\{ q_i^{(t)} | q_i^{(t)} \in C \cup S, T_i^0 + (\alpha + \beta) < t \leq T_i^0 + \theta \right\}$$

be the number of times user i queried for condition (or symptom) of interest in the inclusion period after T_i^0. Let

$$N_i^- = \# \left\{ q_i^{(t)} | q_i^{(t)} \in C \cup S, T_i^0 - \theta < t \leq T_i^0 - (\alpha + \beta) \right\}$$

be the equivalent quantity in the inclusion period prior to T_i^0. The *query rate ratio* that represents the association statistic for the drug-outcome pair of interest is given by

$$QRR = \frac{\sum_i N_i^+}{\sum_i N_i^-}$$

Figure 3 provides an illustration of the *QRR* calculation. It can be seen that in this analysis each user serves as their own control, forming the basis for a self-controlled study design. A similar method called Observational Screening was developed by the OMOP[19] to analyze ADRs in observational data.

The confidence interval for *QRR* (assuming a ratio of two Poisson rates) is given by[42]

$$\frac{2N^- N^+ + Z_{\alpha/2}^2 (N^- + N^+) \pm \sqrt{Z_{\alpha/2}^2 (N^- + N^+)(4N^- N^+ + Z_{\alpha/2}^2 (N^- + N^+))}}{2(N^-)^2}$$

where $N^+ = \sum_i N_i^+$, $N^- = \sum_i N_i^-$. The lower and upper bounds *QRR05, QRR95* are calculated by substituting $Z_{\alpha/2} = 1.64$.

Combining signals

Let $y_1 = \log QRR$, $s_1^2 = \mathrm{Var}(y_1)$, $y_2 = \log EBGM$, and $s_2^2 = \mathrm{Var}(y_2)$. The inverse variance weighted association statistic (*IVW*) for a given drug-outcome pair is given by

$$IVW = \frac{y_1/s_1^2 + y_2/s_2^2}{1/s_1^2 + 1/s_2^2}$$

where s_1^2 and s_2^2 are approximated by $\log(QRR95/QRR05)/2\,Z_{\alpha/2}$ and $\log(EB95/EB05)/2\,Z_{\alpha/2}$ respectively ($Z_{\alpha/2} = 1.64$). The lower 5% percentile of the *IVW* distribution is given by

$$IVW05 = IVW - Z_{\alpha/2}\sqrt{s_1^2 s_2^2/(s_1^2 + s_2^2)}$$

AUC statistics

The comparative statistic of the relative difference between the AUC of signals from search logs and the AUC of signals from AERS (used in Table 2) is defined by

$$\frac{\mathrm{AUC}(QRR05) - \mathrm{AUC}(EB05)}{\max(\mathrm{AUC}) - \mathrm{AUC}(EB05)}$$

where max(AUC)=1 for full-AUC analysis, and max(AUC)=0.3 for partial-AUC analysis at 0.3 FPR. The AUC difference (improvement) of the combined signals relative to either search logs or AERS (used in Table 3) is defined similarly by

$$\frac{\mathrm{AUC}(IVW05) - \max\left(\mathrm{AUC}(EB05), \mathrm{AUC}(QRR05)\right)}{\max(\mathrm{AUC}) - \max(\mathrm{AUC}(EB05), \mathrm{AUC}(QRR05))}$$

A two-sided test was applied to test whether the differences in the AUCs of search logs (based on *QRR05*) and AERS (based on *EB05*) were statistically significant. A one-sided test was applied to test whether the AUC of the combined signal-score (*IVW05*) represents a statistically significant improvement over the AUC of the individual sources. The tests were applied to the pooled set of signal scores representing all four outcomes in order to produce a single result (*p*-value). Statistical significance (*p*-values) was computed using stratified (by ground truth) bootstrapping of signal scores, available in the R package pROC[43]. Bootstrapping ensures that both independent and correlated AUCs (e.g., the combined versus individual signal scores) are appropriately tested.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies. JAMA. 1998; 279(15):1200–1205. 4/15/1998. [PubMed: 9555760]

2. Classen DC, Pestotnik SL, Evans RS, Lloyd JF, Burke JP. Adverse drug events in hospitalized patients. Excess length of stay, extra costs, and attributable mortality. JAMA. 1997; 277(4):301–306. 1/22/1997. [PubMed: 9002492]

3. Ahmad SR. Adverse drug event monitoring at the Food and Drug Administration - Your report can make a difference. Journal of General Internal Medicine. 2003; 18(1):57–60. 1/2003. [PubMed: 12534765]

4. Harpaz R, DuMouchel W, Shah NH, Madigan D, Ryan P, Friedman C. Novel data-mining methodologies for adverse drug event discovery and analysis. Clin Pharmacol Ther. 2012; 91(6): 1010–1021. 6/2012. [PubMed: 22549283]

5. Harpaz R, Dumouchel W, Lependu P, Bauer-Mehren A, Ryan P, Shah NH. Performance of pharmacovigilance signal-detection algorithms for the FDA adverse event reporting system. Clin Pharmacol Ther. Jun; 2013 93(6):539–546. [PubMed: 23571771]

6. Platt R, Wilson M, Chan KA, Benner JS, Marchibroda J, McClellan M. The New Sentinel Network - Improving the Evidence of Medical-Product Safety. New England Journal of Medicine. 2009; 361(7):645–647. 8/13/2009. [PubMed: 19635947]

7. Avorn J, Schneeweiss S. Managing Drug-Risk Information - What to Do with All Those New Numbers. New England Journal of Medicine. 2009; 361(7):647–649. 8/13/2009. [PubMed: 19635948]

8. Stang PE, Ryan PB, Racoosin JA, et al. Advancing the Science for Active Surveillance: Rationale and Design for the Observational Medical Outcomes Partnership. Annals of Internal Medicine. 2010; 153(9):600–W206. 11/2/2010. [PubMed: 21041580]

9. McClellan M. Drug Safety Reform at the FDA - Pendulum Swing or Systematic Improvement. N Engl J Med. 2007; 356:1700–1702. 2007. [PubMed: 17435081]

10. Coloma PM, Schuemie MJ, Trifiro G, et al. Combining electronic healthcare databases in Europe to allow for large-scale drug safety monitoring: the EU-ADR Project. Pharmacoepidemiol Drug Saf. 2011; 20(1):1–11. 1/2011. [PubMed: 21182150]

11. [Accessed Nov 2013] FDA Science Board Subcommittee: Review of the FDA/CDER Pharmacovigilance Program (Prepared for the FDA Science Board May 2011). http://www.fda.gov/downloads/AdvisoryCommittees/CommitteesMeetingMaterials/ScienceBoardtotheFoodandDrugAdministration/UCM276888.pdf

12. Leaman R, Wojtulewicz L, Sullivan R, Skariah A, Yang J, Gonzalez G. Towards Internet-Age Pharmacovigilance: Extracting Adverse Drug Reactions from User Posts in Health-Related Social Networks. Proceedings of the 2010 Workshop on Biomedical Natural Language Processing. 2010; 2010:117–125.

13. Wicks P, Vaughan TE, Massagli MP, Heywood J. Accelerated clinical discovery using self-reported patient data collected online and a patient-matching algorithm. Nat Biotech. 2011; 29(5): 411–414. 5/2011 print.

14. Centers for Disease Control and Prevention (CDC). [Accessed Nov 2013] Use of the Internet for Health Information: United States. 2009. http://www.cdc.gov/nchs/data/databriefs/db66.htm

15. Pew Research Center. [Accessed Nov 2013] Pew Internet & American Life Project Health Online. 2013. http://www.pewinternet.org/~/media/Files/Reports/2013/Pew%20Internet%20Health%20Online%20report.pdf

16. Ginsberg J, Mohebbi MH, Patel RS, Brammer L, Smolinski MS, Brilliant L. Detecting influenza epidemics using search engine query data. Nature. Feb 19; 2009 457(7232):1012–U1014. [PubMed: 19020500]

17. White RW, Tatonetti NP, Shah NH, Altman RB, Horvitz E. Web-scale pharmacovigilance: listening to signals from the crowd. Journal of the American Medical Informatics Association. 2013; 20(3):404–408. March 6, 2013. [PubMed: 23467469]

18. Louis TA, Lavori PW, Bailar JC, Polansky M. Crossover and Self-Controlled Designs in Clinical Research. New England Journal of Medicine. 1984; 310(1):24–31. [PubMed: 6689736]

19. Observational Medical Outcomes Partnership (OMOP). [Accessed Nov 2013] http://omop.org/

20. Trifiro G, Pariente A, Coloma PM, et al. Data mining on electronic health record databases for signal detection in pharmacovigilance: which events to monitor? Pharmacoepidemiology and Drug Safety. Dec; 2009 18(12):1176–1184. [PubMed: 19757412]

21. DuMouchel W, Pregibon D. Empirical bayes screening for multi-item associations. Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining. 2001; 2001:67–76.

22. Szarfman A, Machado SG, O'Neill RT. Use of screening algorithms and computer systems to efficiently signal higher-than-expected combinations of drugs and events in the US FDA's spontaneous reports database. Drug Saf. 2002; 25(6):381–392. 2002. [PubMed: 12071774]

23. Alvarez Y, Hidalgo A, Maignen F, Slattery J. Validation of statistical signal detection procedures in eudravigilance post-authorization data: a retrospective evaluation of the potential for earlier signalling. Drug Saf. 2010; 33(6):475–487. 6/1/2010. [PubMed: 20486730]

24. Almenoff JS, LaCroix KK, Yuen NA, Fram D, DuMouchel W. Comparative performance of two quantitative safety signalling methods: implications for use in a pharmacovigilance department. Drug Saf. 2006; 29(10):875–887. 2006. [PubMed: 16970511]

25. Berlin C, Blanch C, Lewis DJ, et al. Are all quantitative postmarketing signal detection methods equal? Performance characteristics of logistic regression and Multi-item Gamma Poisson Shrinker. Pharmacoepidemiol Drug Saf. 2012; 21(6):622–630. 6/2012. [PubMed: 21994119]

26. Ryan PB, Madigan D, Stang PE, Marc OJ, Racoosin JA, Hartzema AG. Empirical assessment of methods for risk identification in healthcare data: results from the experiments of the Observational Medical Outcomes Partnership. Stat Med. 2012 9/27/2012.

27. Hochberg AM, Reisinger SJ, Pearson RK, O'Hara DJ, Hall K. Using Data Mining to Predict Safety Actions from FDA Adverse Event Reporting System Data. Drug Information Journal. 2007; 41(5): 633–643. September 1, 2007.

28. Harpaz R, DuMouchel W, LePendu P, Shah NH. Empirical Bayes Model to Combine Signals of Adverse. Proc of 2013 ACM SIGKDD Int Conf on Knowledge Discovery and Data Mining (KDD'13).

29. Harpaz R, Vilar S, Dumouchel W, et al. Combing signals from spontaneous reports and electronic health records for detection of adverse drug reactions. Journal of the American Medical Informatics Association : JAMIA. May 1; 2013 20(3):413–419. [PubMed: 23118093]

30. LePendu P, Iyer SV, Bauer-Mehren A, et al. Pharmacovigilance Using Clinical Notes. Clin Pharmacol Ther. Jun; 2013 93(6):547–555. [PubMed: 23571773]

31. Shetty KD, Dalal SR. Using information mining of the medical literature to improve drug safety. J Am Med Inform Assoc. 2011; 18(5):668–674. 9/2011. [PubMed: 21546507]

32. Pouliot Y, Chiang AP, Butte AJ. Predicting adverse drug reactions using publicly available PubChem BioAssay data. Clin Pharmacol Ther. 2011; 90(1):90–99. 7/2011. [PubMed: 21613989]

33. Vilar S, Harpaz R, Chase HS, Costanzi S, Rabadan R, Friedman C. Facilitating adverse drug event detection in pharmacovigilance databases using molecular structure similarity: application to rhabdomyolysis. J Am Med Inform Assoc. 2011; 18(Suppl 1):i73–i80. 12/2011. [PubMed: 21946238]

34. Ryan PB, Schuemie MJ, Welebob E, Duke J, Valentine S, Hartzema AG. Defining a reference set to support methodological research in drug safety. Drug safety : an international journal of medical toxicology and drug experience. Oct; 2013 36(Suppl 1):S33–47. [PubMed: 24166222]

35. Ryan PB, Stang PE, Overhage JM, et al. A comparison of the empirical performance of methods for a risk identification system. Drug safety : an international journal of medical toxicology and drug experience. Oct; 2013 36(Suppl 1):S143–158. [PubMed: 24166231]

36. Caster O, Norén GN, Madigan D, Bate A. Large-scale regression-based pattern discovery: The example of screening the WHO global drug safety database. Statistical Analysis and Data Mining. 2010; 3(4):197–208.

37. Cami A, Arnold A, Manzi S, Reis B. Predicting Adverse Drug Events Using Pharmacological Network Models. Science Translational Medicine. 2011; 3(114):114ra127. December 21, 2011.

38. Systematized Nomenclature of Medicine-Clinical Terms (SNOMED CT). [Accessed Nov 2013] http://www.ihtsdo.org/snomed-ct/

39. [Accessed Nov 2013] RxNorm. http://www.nlm.nih.gov/research/umls/rxnorm/

40. Whetzel PL, Noy NF, Shah NH, et al. BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. Nucleic Acids Res. Jul.2011 39:W541–W545. [PubMed: 21672956]

41. Beeferman, D.; Berger, A. Agglomerative clustering of a search engine query log. Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining; Boston, Massachusetts, USA. 2000.

42. Graham PL, Mengersen K, Morton AP. Confidence limits for the ratio of two rates based on likelihood scores: non-iterative method. Stat Med. Jun 30; 2003 22(12):2071–2083. [PubMed: 12802823]

43. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S plus to analyze and compare ROC curves. Bmc Bioinformatics. Mar 17.2011 :12. [PubMed: 21219653]

Study Highlights

What is the current knowledge on the topic?

Augmenting pharmacovigilance with patient generated data on the Internet has been called out as a promising direction by an FDA working group.

What question did this study address?

Could harnessing behavioral data drawn from Internet search logs be used to detect adverse drug reactions (ADRs), and enhance current pharmacovigilance practices?

What this study adds to our knowledge?

A surveillance system based on Internet search logs can attain a relatively high degree of accuracy in identifying ADRs, with expected performance comparable or surpassing that based on FDA's adverse event reporting system (AERS). Jointly leveraging data from AERS and search logs can further improve detection accuracy by 19% over the use of each data source independently

How this might change clinical pharmacology and therapeutics?

This study informs the design, use, and potential value of a future working surveillance system based on Internet search logs to transform pharmacovigilance and support the vision of a more comprehensive approach.
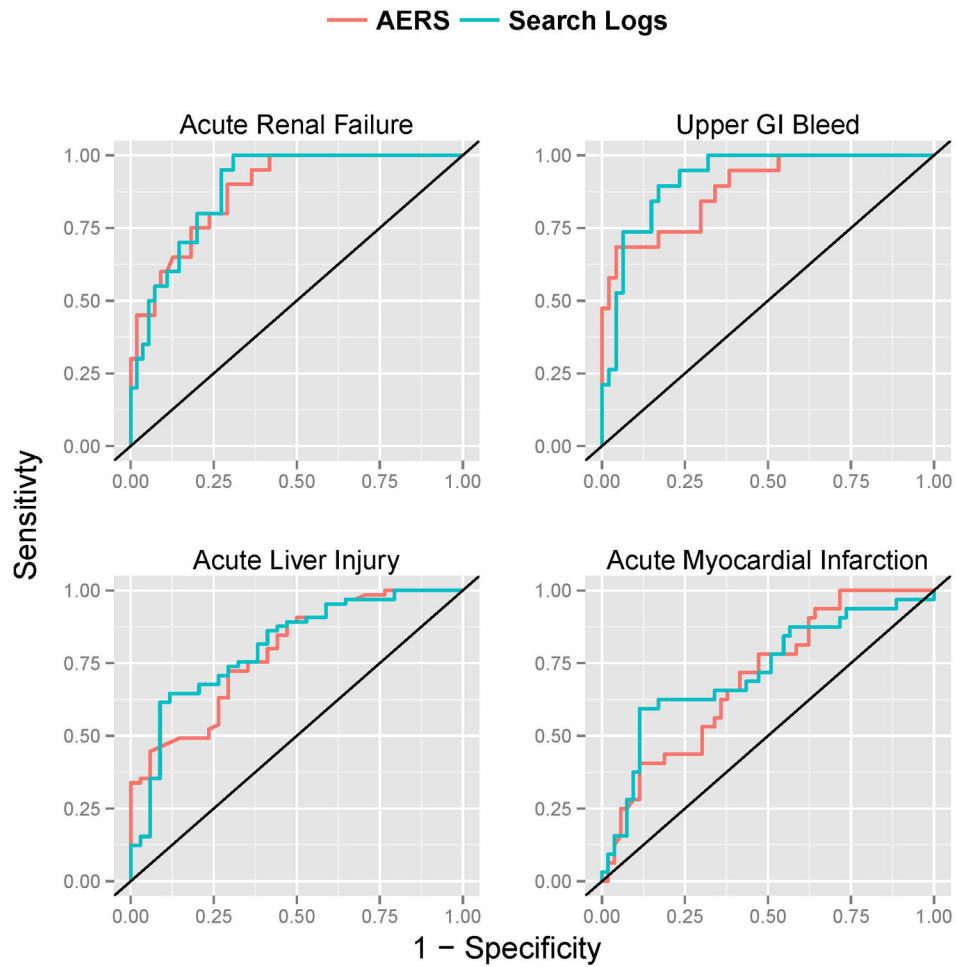
Figure 1.
ROC curves of signal detection using analyses of AERS data (red) and search logs (blue).
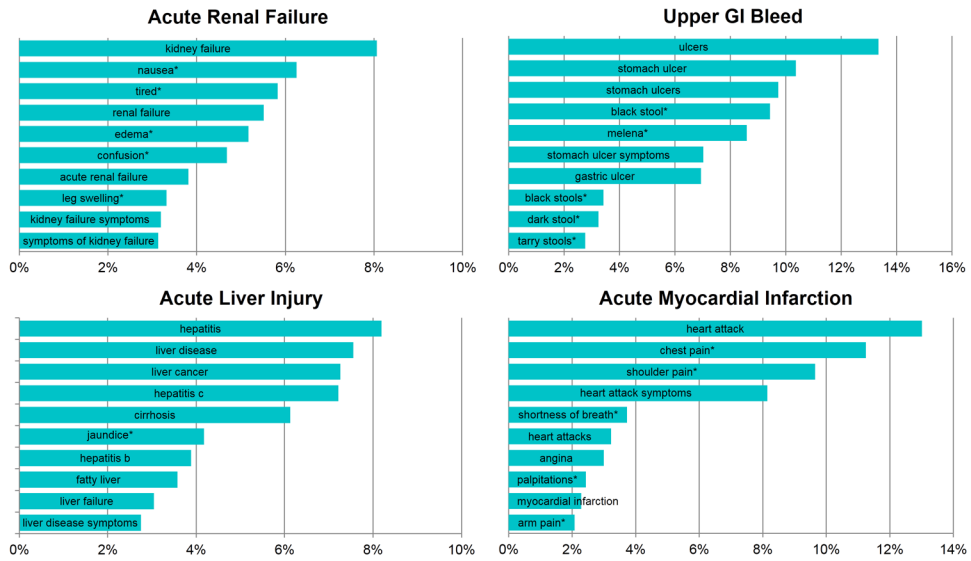
Figure 2.
Top 10 search terms for each of the OMOP outcomes. Percentage denotes the fraction of queries with the search term for the outcome. A search term corresponding to a symptom of a condition is suffixed with (*).
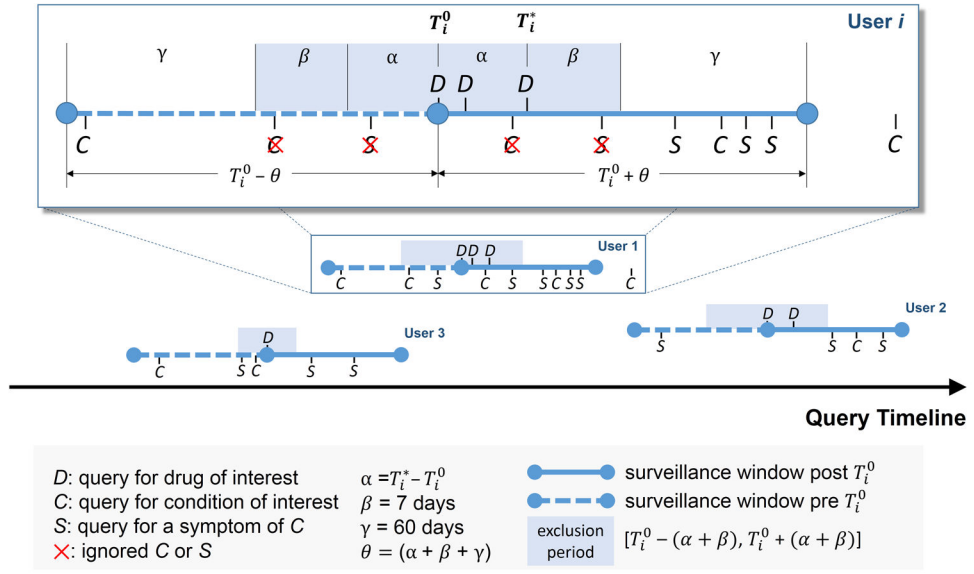
Figure 3.
Illustration of the components used to compute associations between drugs and medical conditions for signal detection using search logs. Each user is associated with a surveillance period (blue line) centered on the time of the users' first query (T_i^0) for a drug of interest (D). Associations are estimated by calculating the query rate ratio (QRR)—the ratio between the number of queries for a condition (C) or symptom (S) of interest outside the exclusion period (shaded region) after and before T_i^0. In this example, QRR=(4+3+2)/(1+1+1)=3.

Table 1

Distribution of OMOP test cases used in the evaluation.

	Test Cases		
	Positive	**Negative**	**Total**
Acute Renal Failure	20	55	75
Upper GI Bleed	19	47	66
Acute Liver Injury	65	34	99
Acute Myocardial Infarction	32	53	85
Total	136	189	325

Table 2

Comparison of signal detection accuracy for AERS and search logs.

	Full AUC			Partial AUC at 0.3 FPR		
	AERS (EB05)	Search Logs (QRR05)	AUC difference	AERS (EB05)	Search Logs (QRR05)	AUC difference
Acute Renal Failure	0.88	0.88	-4%	0.19	0.19	-2%
Upper GI Bleed	0.89	0.92	29%	0.21	0.22	17%
Acute Liver grievance	0.79	0.81	12%	0.14	0.16	10%
Acute Myocardial Infarction	0.70	0.73	9%	0.10	0.14	19%
Average	0.81	0.83	11%	0.16	0.18	12%

EB05, QRR05: association statistics used to quantify signals generated from AERS and search logs respectively. AUC: area under receiver operating characteristic curve (AUC),AUC difference is the proportion of error reduction gained by using one data source over the other

Table 3

Signal detection accuracy for a strategy that combines signal generated from AERS and search logs.

	Full AUC		Partial AUC at 0.3 FPR	
	AERS+Search Logs (IVW05)	AUC difference	AERS+Search Logs (IVW05)	AUC difference
Acute Renal Failure	0.93	45%	0.23	40%
Upper GI Bleed	0.92	-3%	0.23	14%
Acute Liver Injury	0.86	24%	0.19	22%
Acute Myocardial Infarction	0.75	8%	0.14	2%
Average	0.86	19%	0.20	19%

IVW05: statistic used to quantify signals generated by combining AERS and search logs through inverse variance weighting of AERS and search logs association statistics. AUC: area under receiver operating characteristic curve (AUC).AUC difference is defined as the proportion of error reduction gained by using the combined signals over the better performing individual data source.