Post-market drug surveillance sans trial costs: Discovery of adverse drug reactions via large-scale analysis of Web search queries

Abstract

Background: Postmarket drug safety surveillance largely depends on spontaneous reports by patients and healthcare providers, hence less common adverse drug reactions—especially those caused by long-term exposure, multidrug treatments, or specific to special populations—often elude discovery.

Objective: Here we propose an ultra-low-cost fully automated method for continuous monitoring of adverse drug reactions in single drugs and in combinations thereof, and demonstrate the discovery of heretofore unknown ones.

Materials and Methods: We use aggregated search data of large populations of Internet users to extract information related to drugs and adverse reactions to them, and correlate these data over time. We further extend our method to identify adverse reactions to combinations of drugs.

Results: We validate our method by showing high correlation of our findings with known adverse drug reactions (ADRs). However, while acute, early-onset drug reactions are more likely to be reported to regulatory agencies, we show that less acute, later-onset ones are better captured in Web search queries.

Conclusions: Our method is advantageous in identifying previously unknown adverse drug reactions. These ADRs should be considered as candidates for further scrutiny by medical regulatory authorities, e.g., through Phase IV trials.

Keywords: machine learning, side effects, infoveillance, infodemiology

Introduction

Existing mechanisms for post-market drug surveillance work well in many cases, but failures resulting in harm to patients and even fatalities are widely documented [1], including the withdrawal of Thalidomide in the 1960's [2], and more recently of Cerivastatin [3], Troglitazone [4], and Rofecoxib [5]. Two main kinds of post-market drug surveillance mechanisms exist today. One kind is run by regulatory agencies, such as MedWatch and the Vaccine Adverse Event Reporting System (VAERS) by the FDA and CDC in the United States, the Yellow Card Scheme by the MHRA in the United Kingdom, and the International Drug Monitoring Programme by WHO. These are supplemented by public (or public-private cooperation) initiatives such as Research on Adverse Drug events And Reports (RADAR) and Web sites such as eHealthMe.com, which collect patient-reported information on drug outcomes. The most serious limitation of these data collection initiatives is that they rely on the patients and their health care providers to make the association between the adverse drug reaction (ADR) and the drug. This can be especially difficult when the adverse reaction appears only after the drug is taken for a lengthy period of time, or when the patient takes several medications concurrently. To alleviate this problem, projects such as the FDA's Sentinel Initiative [6], the EU-ADR initiative [7], and the Observational Medical Outcomes Partnership (OMOP) [8, 9, 10, 11] are beginning to use observational data, including administrative claims and electronic health records, to identify adverse drug reactions.

Our proposed approach uses a novel kind of observational data, namely, Web search query logs. Search queries contain a cornucopia of world knowledge, and prior studies used query logs to track tropical storms [12], certain life events [13], and the spread of disease [14]. As such, this approach is an example of infodemiology [15], and is enabled by the fact that as many as 80% of U.S. Internet users seek health information online [16]. Consequently, our methodology allows analyzing the data from literally

hundreds of millions of people, and in some cases, a significant percentage of the patients using a given drug. Performing such analysis continually allows for long-term monitoring, while grouping search requests by geographical location facilitates demographic segmentation of the population [17].

Existing drug surveillance mechanisms often depend on the need for medical providers or patients to realize the connection between the treatment and its side effects (adverse or otherwise). This inherent limitation poses a challenge to testing new methods for ADR discovery, since existing data is not comprehensive enough to be considered a gold standard, considering that patients and medical providers might not realize the connection between treatments and some ADRs. We therefore adopt a two-pronged approach for validating our method. First, we show that it can reliably identify currently known ADRs. While the findings of our method are positively correlated with existing data, this correlation is not perfect as we discover new, previously unknown ADRs. Second, we characterize the differences between the known and the newly discovered ADRs, and identify the Most Discordant ADRs (MDADRs) between these two sources. We show that the ADRs found by our method are usually less acute reactions (i.e., not requiring immediate medical attention) with much later onset, which is exactly why they elude detection by conventional mechanisms. Over all the drugs examined, we found that the ADRs “apnea" and “cramps" are consistently overlooked in the FDA data (as reported in the Adverse Event Reporting System, AERS, see below), while “tiredness" and “weight loss" are frequent ADRs of vaccines that are overlooked in VAERS reports. We propose that the ADRs newly discovered by our method be further investigated in carefully-designed clinical trials, which should be lengthy enough to allow detection of late onset reactions.

Materials and Methods

Our method, called Query Log Reaction Score (QLRS), quantifies the prevalence of ADRs for a given drug, as explained below. We used QLRS to identify ADRs of top selling drugs and vaccines based on the queries submitted to the Yahoo U.S. Web search engine during 6 months in 2010. A total of 176 million unique users, as identified by a unique signature of the users' browser, were included in this study. The search logs were anonymized according to the Yahoo privacy policy by scrambling actual user identifiers. This was achieved by using a one-way cryptographic hash function, which makes it impossible to map the resultant hash values back to the original user identifiers, while keeping the probability of collisions very low. As explained below, only the search counts were considered, which were aggregated across the users. Furthermore, the research described herein was carried out according to the Yahoo guidelines on human subject research.

We investigated the 20 drugs listed in Table 2 (additional results for the top 100 drugs are provided in Appendix), which are the top selling drugs in the United States by revenue [18]. We analyzed these drugs for two reasons. First, these findings would likely impact the largest number of people. Second, data is more abundant for these drugs, and thus results are likely to be more significant for these. We note that all of these drugs are usually taken for long periods of time (however, we have also demonstrated the applicability of our method to vaccines, as detailed in the Appendix, which are usually administered a limited number of times to each patient). We limited our work to non-generic versions of these drugs, to reduce the chance of additional confounding influences, and because brand names are mentioned 88% more often than generic names in the query log (not statistically significant). However, we also discuss the differences in ADRs of similar drugs in the Results section.

A total of 195 ICD10 symptoms [19] were studied as manifestations of possible ADRs. We filtered the symptoms according to http://en.wikipedia.org/wiki/List_of_medical_symptoms to facilitate replicability of our method in other languages. This list of symptoms was further expanded with synonyms (see below), since patients frequently use nonmedical terminology to describe their symptoms. Basing our work on terms from Wikipedia (a highly popular information source), and identifying synonyms using behavioral data makes our approach suitable for identifying ADRs as described by non-professionals.

We limited the symptoms under consideration for each drug to the 50 most frequently queried symptoms for that drug. We identified possible ways in which nonprofessionals describe their health symptoms using two query expansion methods. First, we selected the most frequent search terms that led users to click on the Wikipedia page that described each symptom [20,21,22]. Second, we extracted frequently occurring lexical affinities [23], namely, word pairs appearing in close proximity, in the 50 highest ranked Web search results returned when the symptom name is used as a query. The two top terms from each of the two methods were used as alternative names for each symptom. For each symptom discussed in the paper, the various possible search terms expressing it have been mapped to the same medical term. For example, the ADR "diplopia" could have been searched for using a colloquial term "double vision".

Three medical professionals (two medical doctors, one nurse practitioner) independently labeled the expansion terms with respect to their relevance as an expansion term to each specific medical term. The inter-annotator agreement estimated using the Fleiss Kappa statistic [24] was 0.44 (statistically significant, $p<10^{-5}$). This is a medium level agreement. However, for 88% of the terms the majority of annotators (i.e., two or more) agreed that the term was an appropriate expansion of the medical term. Thus, our expansion method indeed constructs a high precision dictionary of terms. In order to maintain

the automated nature of our method, the results reported here are based on all the expansion terms, not just those marked as relevant by the annotators.

For each drug, we first identified all the users who had searched for the drug name. We define Day Zero for each user as the day when that user has first searched for the drug. Day Zero for all other users, who did not search for the drug, was defined as the midpoint of their observed query history. We then counted the number of times each symptom was queried for before and after Day Zero by each user. The purpose of using the data from people who did not search for the drug was to normalize against environmental effects, e.g., seasonal allergies. This is in contrast with most prior infoveillance research [15], which is concerned with whole-population prevalence, rather than the comparison of specific sub-populations, e.g., people using or not using the drug.

For each drug-symptom pair we constructed a two-way contingency table counting the number of times a symptom was searched for before and after Day Zero, for users who did and did not search for the drug (see Table 1). For each symptom, we scored its prevalence as a reaction to the drug using the Pearson's Goodness of Fit test (chi^2 test statistic [25]). We refer to this score as Query Log Reaction Score (QLRS). Additional results for the top 100 drugs are provided in Appendix.

We used two reference datasets to assess the validity of our findings. Adverse Event Reporting System (AERS) is the database of the FDA's postmarket safety surveillance program for approved pharmaceutical drugs. The Side Effect Resource (SIDER) lists known ADRs for marketed drugs, extracted from public documents and package inserts [26].

AERS data was downloaded from the FDA AERS website, and included reports submitted between January 2004 and June 2010. Reports were mapped to the same list of symptoms as QLRS, using the same synonym list. In total, 47% of the cases in AERS were matched to at least one of the 195 symptoms or their synonyms, indicating good coverage by the symptoms list used in our study. Similar analysis was

performed for SIDER. To assess the overall quality of ADR discovery by our method, we computed the Spearman's rank correlation coefficient between two lists of ADRs for each drug, one ordered by QLRS and one by the number of AERS reports.

Observe that AERS data is complicated by the fact that multiple reports can be submitted to the FDA for the same case, and that reports can pertain to side effects of the drug, the underlying disease, or other concurrently taken drugs [27]. Therefore, ADR prevalence according to AERS should be considered a noisy reference. We employed several approaches to computing the correlation. First, we used the raw report counts in AERS, with the corresponding correlation denoted by $\rho 1 = \text{Corr}(\text{QLRS}, \text{AERS count})$. We also used the AERS data to compute two regularized measures of disproportionality that are commonly employed for analyzing adverse side effect reports. Specifically, we used the empirical Bayes geometric mean (EBGM) [28,29], with the correlation denoted by $\rho 2 = \text{Corr}(\text{QLRS}, \text{AERS EBGM})$, and the Information Component (IC) [30], with the correlation denoted by $\rho 3 = \text{Corr}(\text{QLRS}, \text{AERS IC})$.

We hypothesized that some ADRs are more likely to be reported to the FDA, while others tend to be self-addressed by patients through online research. Consequently, if our method is to discover previously unknown ADRs, the correlation can never be perfect. Therefore, we first analyze the commonalities between the ADRs we discovered and those already known. Then, we analyze the properties of the newly discovered ADRs.

To focus on the ADRs identified by both our method and the AERS data, we removed the five symptoms that most reduced the value of one of the metrics, $\rho 1$, using a greedy selection process. We call the removed symptoms Most Discordant ADRs, or MDADRs. Specifically, we iteratively identified and removed the ADR that most reduced the Spearman's rank correlation between the AERS counts and the QLRS ranking of ADRs. An alternative method of removing discordant ADRs would focus on reaching

statistically significant values of ρ1. However, we chose to use a fixed number in order to facilitate the analysis of MDADRs, as performed below.

Identification of Adverse Drug Reactions for Multiple Drugs

Some individuals are prescribed multiple drugs to be taken simultaneously. The interaction between these drugs may give rise to specific ADRs that are not present (or are present at different severity) if each drug is taken individually. Thus, in the following we show how our method can be used to identify ADRs that are associated with taking pairs of drugs. Our method attempts to remove the ADRs attributed to individual drugs so as to identify those ADRs that arise from the combination thereof.

For each pair of drugs that were analyzed, we identified their characteristic ADRs due to the interaction by discounting the probability of the ADRs arising from each of the individual drugs. This was done by subtracting the contribution of ADRs of the individual drugs as predicted by a linear regression model. We hypothesized that the ADRs observed in patients who only take one of the drugs, will appear at a similar ratio for the patients who take both drugs. However, new ADRs that are due to the interaction of the two drugs will not be reliably predicted by modeling each drug separately, and will therefore appear at a substantially different ratio than the prediction.

For each pair of drugs, we identified three disjoint groups of users: the first two groups are those who searched for only one of the two drugs, and the third group searched for both. For the first two groups, we counted the number of times each ADR was searched for before (after) Day Zero. For the third group, i.e., users who searched for both drugs, we defined Day Zero as *max*(date-first-search($drug_i$), date-first-search($drug_j$)), where date-first-search() is the earliest date on which the user searched for a

given drug. We denote these numbers (before/after Day Zero) for the *i*-th ADR in population p by $n^{p}_{i,b}$ ($n^{p}_{i,a}$). Next, we defined the ratio of change in the ADR prevalence (due to the commencement of treatment with the second drug) as $n^{p}_{i,a}/(n^{p}_{i,b} + n^{p}_{i,a})$. Finally, we built a regression model to predict the probability of change in the third population (patients taking both drugs) given the corresponding values in the first two populations. This regression model effectively discounts the effect of the ADRs caused by each drug separately. We also identified MDADRs for pairs of drugs in a similar way as for individual drugs.

Results

We counted the number of times each drug appeared in AERS, and found it to be highly correlated with the number of online searches for that drug. For the drugs listed in Table 2, Spearman's correlation is $\rho = 0.66$ ($P = 0.002$; $n = 20$). The correlation becomes even more pronounced for pairs of drugs, $\rho = 0.73$ ($P = 10^{-68}$; $n = 380$). The correlation between the sales figures (as represented by the number of prescriptions sold) and the number of web searches is $R^2=0.26$ ($p<0.01$). A linear model that uses both the number of AERS reports and the sales figures to predict web search volume yields R^2 greater by 2.4% than the one using only AERS reports. We believe these findings mean that the search volume is more indicative of the prevalence of ADRs rather than actual sales. Thus, the popularity of a drug in web queries is highly representative of its appearance in AERS, suggesting that web queries are strongly reflective of real-world phenomena.

As noted in the Methods section, we assessed the overall quality of ADR discovery by our method by computing the Spearman's rank correlation coefficient between two lists of ADRs for each drug, one ordered by QLRS and one by the number of AERS reports. First, we used the raw report counts in AERS, with the corresponding correlation denoted by $\rho 1 = Corr(QLRS, AERS\ count)$. We also used the AERS data to compute two regularized measures of disproportionality that are commonly employed for analyzing adverse side effect reports. Specifically, we used the empirical Bayes geometric mean (EBGM) [28,29], with the correlation denoted by $\rho 2 = Corr(QLRS, AERS\ EBGM)$, and the Information Component (IC) [30], with the correlation denoted by $\rho 3 = Corr(QLRS, AERS\ IC)$.

Table 2 reports the values of $\rho 1$ after removing 5 MDADRs for each drug. QLRS predictions are relatively highly correlated with the AERS counts, and the correlation is statistically significant ($P < 0.05$) for 12 of the 20 drugs (Olkin-Pratt (DSL) fixed-effect meta-analytical approach [31], $P<0.001$; $n = 20$). Positive

correlation was not found in only one of the drugs (Singulair). Interestingly, removing 15 MDADRs for this drug (instead of 5) resulted in a statistically significant correlation of $\rho 1 = 0.48$ ($P = 0.02$), thus suggesting a particularly high discrepancy between the prevalence of ADRs as predicted by QLRS and as registered in AERS for this drug.

We also note that although the majority of observed correlation values are significant, they are far from indicating perfect correlation. This is to be expected, as the correlation would only have been perfect if our method were exactly re-discovering the known ADRs. However, as we discover previously unknown ADRs, we obviously achieve an imperfect match to the list of known ones in AERS. In the following section, we analyze the differences between the known ADRs and those identified by our method. Additionally, there is a small negative correlation (Spearman $\rho=-0.22$, $p=0.02$) between the number of users who queried for a drug and $\rho 1$. This demonstrates that higher correlations are obtained when more data is available, and is an additional cause for the imperfect correlations.

Statistically significant correlations with EBGM and IC were also found (see Appendix), and the meta-analysis is highly statistically significant at $P< 10^{-3}$; $n = 100$. Note, however, that EBGM and IC are measures designed to enhance the detection of ADRs that are especially prevalent in a given drug under study, compared to all other drugs. At the same time, raw AERS counts (used for the computation of $\rho 1$) are more likely to be associated with the appearance of an ADR regardless of any other drug. This explains the higher correlation we observed of QLRS with the raw AERS counts ($\rho 1$) than with EBGM and IC ($\rho 2$ and $\rho 3$, respectively).

SIDER [26] contains information on ADRs extracted from public documents and package inserts. Due to regulatory and legal requirements, it is overly inclusive in its listings, which makes it a noisy reference as well. The SIDER data is essentially binary, without relative frequency or absolute counts, which makes the above correlation analysis inapplicable. We used SIDER to assess the accuracy of QLRS by computing

the Area Under the ROC Curve (AUC [32]) and the F-measure [33], taking as positive examples all the ADRs listed in SIDER for the drug. AUC measures the method's ability to correctly identify known ADRs, while the F-measure simultaneously considers precision and recall. Only 8 of the 20 drugs we analyzed appeared in SIDER, and the corresponding accuracies are reported in Table 3 (again, after removing MDADRs). The results suggest that our method is able to reconstruct known ADRs, as measured both in terms of AUC and the F-measure.

Most Discordant Adverse Drug Reactions (MDADRs)

Analyzing the most discordant ADRs reveals characteristic differences between the known ADRs (registered in AERS and SIDER) and those identified by our method. We show that ADRs identified as most discordant are not random; instead, they belong to one of the following two classes. The first class includes ADRs that are readily recognized by patients and medical professionals due to their acuteness and fast onset. The other class includes later-onset, less acute ADRs, which are more difficult to identify using self-reporting methods.

As noted above, upon removing as few as 5 MDADRs, the correlation between QLRS and AERS counts ($\rho 1$) frequently becomes statistically significant. As opposed to this, removing a random subset of 5 symptoms only results in a negligible, statistically insignificant change in the correlation.

Interestingly, although the MDADRs were identified separately for each drug, they were highly consistent across drugs. Of the 32 MDADRs we identified overall, 22 were chosen for more than one drug (mean 3.1, s.d. 2.3). Significantly, these ADRs were always overemphasized either in the query log

or in AERS, but never in both (for different drugs). The likelihood of such behavior at random is smaller than $1 : 10^5$. A typical example is the ADR "nausea", which appeared at a far higher rank (i.e., more prevalent) in the AERS dataset than in the QLRS ranking for 7 out of the 20 drugs, and was never found at a rank below that of QLRS for the other drugs.

Most importantly, MDADRs that are prominent in queries and in AERS have notable differences in their temporal behavior. As an illustrative example, we used the query log to compute the cumulative density functions (CDFs) over time for two MDADRs for the drug Effexor, one overemphasized in AERS ("nausea") and one overemphasized by QLRS ("sleepiness"). Figure 1 shows the difference between the CDFs of the two MDADRs, starting from the time the drug is first searched for (day 0). As Figure 1 demonstrates, each of these two MDADRs is more likely to occur in a different time range. Observe that the symptom prominent in AERS ("nausea") is usually searched for shortly after the first query about the drug (i.e., several days after day 0), when it is much more likely that the other symptom. In contrast, the symptom ranked highly by QLRS ("sleepiness") appears much more prominently 45 to 75 days after the commencement of treatment, where the likelihood of nausea drops significantly.

We measured the difference in the time of onset (defined as the number of days between the first search for the drug and the first search for the ADR in the query log) for the MDADRs that were overemphasized by QLRS and in the AERS data. Averaged over all the drugs, the difference was 7.3 days (statistically significant, Wilcoxon's signed-rank test [34], two sided, $P = 0.014$; $n = 15$). Based on these findings, we conclude that ADRs are more likely to be reported to the regulatory authorities if they appear shortly after commencing the treatment (as it might be easier for patients and caregivers to link the ADRs to the treatment), and that are serious enough to warrant reporting. Conversely, ADRs identified by our method usually appear much later after the beginning of treatment, hence their possible association to the drug is often overlooked.

Thus, the MDADRs overemphasized by QLRS represent an interesting class of reactions that are harder to discover using traditional methods.

Correlation between QLRS scores of similar drugs

Several of the drugs we investigated are different brands of essentially the same drug. All other things being equal, we expected that patients taking two different brand versions of the same drug would experience similar ADRs. To evaluate this conjecture, we conducted two evaluations. First, we measured the Spearman's correlation between the QLRS scores of ADRs for the multiple brand versions of the same drug. Second, we evaluated the correlation between QLRS scores and AERS counts after aggregating the chi^2 contingency tables for drugs that have the same generic names.

The intra-drug correlation (different brands of the same drug) was, on average, 0.42, compared to 0.23 for all the other pairs of drugs ($P = 0.03$, one-sided ranksum test). Thus, while the ADRs are somewhat different among the near-identical drugs, the correlation is statistically significantly higher than that observed for random pairings of drugs. The imperfect correlation can be explained by several factors.

First, different manufacturers may produce drugs with slight variations in inactive ingredients, coloring agents, and fillers. The change in fillers in the drug Eltroxin in Israel has been associated with a large number of patients experiencing major side effects, including changes in heart rate, dizziness, and difficulties in breathing (See, for example, http://www.haaretz.com/news/national/israel-drug-company-summoned-to-hearing-over-thyroid-drug-eltroxin-1.409067). Furthermore, there could be demographic differences between the populations taking those drugs. For example, different brand versions may be prescribed in different markets or different geographical regions. Finally, while many health-care providers report that they do not employ special monitoring after switching from brand

names to generic drugs, some have encountered specific ADRs caused by such switching [35]. Thus, there are known differences in ADRs caused by different versions of similar drugs.

Nonetheless, the relatively high correlation between the ADRs of similar drugs provides additional supporting evidence that the ADRs discovered by QLRS are a genuine reflection of actual patient experiences.

Using RxNORM (http://www.nlm.nih.gov/research/umls/rxnorm/overview.html, accessed January 2013.) we identified 30 drugs that are different brand versions of 14 generic drugs, out of the 100 drugs that we analyzed (see Appendix). For example, Procrit and Eprex are two brand versions of the generic drug Erythropoietin. This allowed us to focus the analysis on the generic component of the drug rather than the specific brand names, by computing an aggregated contingency table for all brand name versions of the same generic drug prior to computing the QLRS score (the AERS counts were aggregated similarly).

The resulting correlations are reported in Table 5. As can be seen from the table, the values of $\rho 1$ aggregated over all the brand names of the same generic drug are significantly higher than those for individual brand names (on average, aggregated $\rho 1$ equals to 0.62, compared to 0.35 for individual brand names). In all 14 cases, statistically significant correlations between QLRS scores and AERS counts were found. MDADRs found in the aggregated data were the same as those identified in the brand name versions of the drug in 79% of the cases.

We attribute these findings to several reasons. First, by aggregating different brand name versions of a drug we focus on the active ingredient of the drug, which is more likely to be reported in AERS by medical personnel. Second, averaging these additional observations over multiple drugs reduces the amount of noise in the data, and thus increases the correlation with AERS. Finally, analyzing the active

ingredient reduces the effect of individual manufacturing procedures and components, and thus focuses the analysis on a simpler set of chemical components, which may have a smaller set of ADRs.

Adverse Reactions to Multiple Drugs

Observe that some ADRs occur only when two drugs are taken concurrently or in close temporal proximity. These ADRs can be especially difficult to detect because they occur infrequently, and only in a population that takes both drugs. To this end, we extended our method to identify ADRs of pairs of drugs.

The correlation between QLRS rankings and AERS (raw counts, EBGM, and IC), for the 10 most common pairs of drugs, is shown in Table 6. These correlations are lower than those for individual drugs, but are nonetheless statistically significant (Olkin-Pratt (DSL) fixed-effect meta-analytical approach [31], $p<0.001$; $n = 10$). We believe this result is noteworthy because it may be more difficult to include in a clinical trial those patients who take multiple drugs concurrently.

Based on these findings, we believe our method can also be applicable to combination products (i.e., drugs that contain two or more active substances), if each of the active substances is marketed also as a separate drug, in similar doses. We plan to extend our method to combination products in our future work, and intend to investigate whether the correlation can be increased, for example, by using non-linear correlation measures.

In their recent work, White et al. [36] used search logs to study the side effects of one specific drug pair, Paroxetine and Privastatine, whose interaction was reported to cause hyperglycemia. Their finding confirms the utility of search logs in identifying drug interactions, which were later validated by the FDA.

However, there are several key differences between their study and ours. First, the method proposed by White et al. performs a direct count of symptoms, thus not taking into account seasonal and other effects handled by the QLRS method. Second, in order to identify queries that are indicative of hyperglycemia, White et al. constructed a list of hyperglycemia-related terms manually, by reviewing the relevant medical literature. In contrast, we map user queries to medical terminology in an automated way, building on query expansion methods developed in the field of information retrieval. Finally, whereas White et al. only analyzed one particular condition (hyperglycemia) and one specific pair of drugs (Paroxetine and Pravastatin), our study is conducted at a substantially bigger scale. Specifically, we automatically mine side effects of 100 top selling drugs and their combinations, as well as side effects of vaccines.

Discussion

Clinical trials of pharmaceutical drugs are limited in their extent owing to their prohibitively high cost and insufficient diversity among participants. On the other hand, voluntary reporting of ADRs by patients and healthcare professionals is limited because of the extra effort required, and because of the difficulty of linking the ADRs to the drug that caused them (especially when these ADRs have a late onset or are due to multi-drug treatments). We proposed a novel, low-cost method for discovering adverse drug reactions from aggregated Web search data of large populations of Internet users. We demonstrated that our method allows analyzing the ADRs of drugs and vaccines in dramatically larger populations than typical clinical trials, and can assist in identifying ADRs that so far eluded discovery by the existing mechanisms.

We believe our method constitutes a new, complementary approach to pharmacovigilance, due to its computational efficiency as well as access to vastly larger and more diverse populations. There are multiple avenues for future work. The effectiveness of our method can be validated by analyzing medical records (e.g., OMOP), or by assessing its ability to predict changes in safety labels by regulatory authorities. It would also be interesting to compare QLRS scores to those derived from the analysis of social media. Finally, a validation of MDADRs through clinical trials would be of significant value to validate our method. Specifically, we propose to test the MDADRs found by our method, which are underemphasized in current ADR databases (e.g., AERS), in a clinical setting or through phase 4 trials. Such trials should be prioritized by the severity, volume of searches, and uniqueness of the ADRs discovered. Once verified, these MDADRs will become an important addition to the list of known ADRs of which patients are informed. Finally, quantifying the strength of the protopathic bias (if any) in our data would serve to strengthen the validity of ADRs discovered by our method.

Our work falls within the domain of infodemiology, that is, the study of Internet media to inform public health and policy [15]. Much previous work in this area has centered on detection and characterization of transient events (i.e., disease outbreaks [15] and special events [36]) and the analysis of the kinds of information available to users [38]. Our paper is novel in that it makes use of search engine queries to identify transient events at the individual level, and more importantly, to discover associations between events [15] that eluded detection by the patients themselves or their health practitioners.

The main drawback of relying on Web search data is that it is inherently noisy. It is often impossible to ascertain whether a person searching for drugs and ADRs is doing so out of curiosity, or conducting research for himself, a relative, or even for a patient. Admittedly, Internet users comprise a biased sample of the population, and so the ADRs discovered may not be fully representative of the entire population. Nonetheless, our results suggest that the sheer size of the data alleviates these concerns, and the proposed method is able to identify adverse effects of drugs that are not captured by existing surveillance mechanisms.

Another limitation of this study is using a restricted set of symptoms expanded through the use of synonyms. While using a larger dictionary would have allowed identification of additional (and possibly rarer) ADRs, our focus on more common symptoms is likely to lead to better identification of the more common concerns to patients. Future work will focus on professionally-used term dictionaries in order to focus on more knowledgeable patients and health providers. Another way to strengthen our results is the use of non-English search data, which will increase the volume of data (and the size of the observed population) thus enabling the analysis of less frequent drugs and ADRs.

Finally, although this work is based on data from a large Internet search engine, it does not cover the entire population. However, privacy concerns preclude conducting our analysis across search engines**, as the latter never share information about their users**. Nevertheless, given the sheer number of users

whose data was analyzed in the study (176 million, which is especially notable compared to most other pharmacovigilance studies), we believe our findings are still highly significant. It should also be emphasized that QLRS discovers adverse drug reactions via aggregating queries across multiple users and query sessions. Consequently, the output of our method comes in the form of a list of newly discovered ADRs for each drug, and does not include any private, personal, or user-specific data whatsoever.

Our approach is mostly language-independent except for the initial list of symptoms (http://en.wikipedia.org/wiki/List_of_medical_symptoms), and obtaining the latter from non-English Wikipedias will allow one to apply the method to additional languages, markets, and populations. Extending the coverage is particularly important for studying rare events, such as ADRs in patients that take a plurality of drugs. Computational pharmacovigilance, which uses observational data such as Web search query logs, is complementary to the existing data collection mechanisms, and the ADRs it identifies should be considered as candidates for further investigation.

ACKNOWLEDGMENTS: We thank Eugene Agichtein, Elmer Bernstam, David Carmel, Tania Fuchs, Aleksandr Kagan, Bo Pang, and Michael Schwarz for comments.

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

COMPETING INTERESTS: There are no competing interests.

References

1. Ray WA, Stein CM. Reform of Drug Regulation: Beyond an independent drug safety board. N Engl J Med 2006;354:194–201.

2. Lary J, Daniel K, Erickson J, Roberts H, Moore C. The return of Thalidomide: Can birth defects be prevented? Drug Saf 1999;21:161–169.

3. Furberg C, Pitt B. Withdrawal of cerivastatin from the world market. Curr Control Trials Cardiovasc Med 2001;2:205–207.

4. Gale E. Troglitazone: the lesson that nobody learned? Diabetologia 2005;49:1–6.

5. Topol E. Failing the public health -- Rofecoxib, Merck, and the FDA. N Engl J Med 2004;351:1707–1709.

6. Platt R, Wilson M, Chan KA, Benner JS, Marchibroda J, McClellan M. The new sentinel network — Improving the evidence of medical product safety. N Engl J Med 2009;361:645–647.

7. Trifiro G, Fourrier Reglat A, Sturkenboom MCJM, Acedo CD, van der Lei J. The EUADR project: preliminary results and perspective. St Heal T 2009;148:43–49.

8. Overhage JM, Ryan PB, Reich CG, Hartzema AG, Stang PE. Validation of a common data model for active safety surveillance research. J Am Med Inform Assn 2012;19:54–60.

9. Madigan D, Ryan P. What can we really learn from observational studies? The need for empirical assessment of methodology for active drug safety surveillance and comparative effectiveness research. Epidemiology 2011;22:629–631.

10. Stang PE, Ryan PB, Racoosin JA, Overhage JM, Hartzema AG, Reich C, Welebob E, Scarnecchia T, Woodcock J. Advancing the science for active surveillance: rationale and design for the Observational Medical Outcomes Partnership. Ann Intern Med 2010;153:600–606.

11. Ryan PB, Welebob E, Hartzema AG, Stang PE, Overhage JM. Surveying US observational data sources and characteristics for drug safety needs. J Pharmaceut Med 2010;24:231–238.

12. Backstrom L, Kleinberg J, Kumar R, Novak J. Spatial variation in search engine queries. WWW 2008; 357–366.

13. Richardson M. Learning about the world through long-term query logs. ACM Trans Web 2008;2:1–27.

14. Ginsberg J, Mohebbi M, Patel R, Brammer L, Smolinski M, Brilliant L. Detecting influenza epidemics using search engine query data. Nature 2009;457:1012–1014.

15. Eysenbach G. Infodemiology and Infoveillance: Framework for an Emerging Set of Public Health Informatics Methods to Analyze Search, Communication and Publication Behavior on the Internet. J Med Internet Res 2009;11(1):e11

16. Fox S. Health Topics 2011. Pew Internet & American Life Project 2011.

17. Weber I, Jaimes A. Who uses web search for what? And how? WSDM 2011;15–24.

18. Humphreys A. MedAdNews 200 World's BestSelling Medicines. MedAdNews 2007.

19. Janca A, Ustun T, van Drimmelen J, Dittmann V, Isaac M. ICD10 Symptom Checklist for Mental Disorders. WHO 1994.

20. Cui H, Wen JR, Nie JY, Ma WY. Query expansion by mining user logs. IEEE T KnowL Data En 2003;15(4):829-839.

21. Baeza-Yates R, Tiberi A. Extracting semantic relations from query logs. Proc SIGKDD 2007:76-85.

22. Van Der Heijden M, Hinne M, Kraaij W, Verberne S, van der Weide T. Using query logs and click data to create improved document descriptions. Proc workshop on Web Search Click Data 2009:64-67.

23. Carmel D, Farchi E, Petruschka Y, Soffer A. Automatic query refinement using lexical affinities with maximal information gain. SIGIR 2002;283–290.

24. Fleiss JL. Measuring nominal scale agreement among many raters. Psychol bull, 1971;76(5):378.

25. Upton G, Cook I. Oxford dictionary of statistics. Oxford University Press, 2002.

26. Kuhn M, Campillos M, Letunic I, Jensen L, Bork P. A side effect resource to capture phenotypic efects of drugs. Mol Syst Biol 2010;6.

27. DuMouchel W. Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. American statistician 1999;53:177–189.

28. Zorych I, Madigan D, Ryan P, Bate A. Disproportionality methods for pharmacovigilance in longitudinal observational databases. Stat Methods Med Res 2011 [Epub ahead of print].

29. Szarfman A, Machado SG, O'Neill RT. Use of screening algorithms and computer systems to efficiently signal higher than expected combinations of drugs and events in the US FDA's spontaneous reports database. Drug Safety 2004;25:381–392.

30. Wilson AM, Thabane L, Holbrook A. Application of data mining techniques in pharmacovigilance. Brit J Clin Pharmacology 2003;57:127–134.

31. Schulze R. Meta-analysis: a comparison of approaches. Hogrefe and Huber, 2004.

32. Duda R, Hart P, Stork D. Pattern classification (2nd ed.). Wiley, 2001.

33. van Rijsbergen C. Information Retrieval (2nd ed.). Butterworth, 1979.

34. Demsar J. Statistical comparisons of classifiers over multiple data sets. JMLR 2006;7:1–30.

35. Rouleau B, Lavoie L, Leblanc J, Moretti S, Collin C. Reporting of adverse drug reactions by community pharmacists: a qualitative study in Quebec. Drug Inf J 2011;45(5):627-639.

36. White RW, Tatonetti NP, Shah NH, Altman RB, Horvitz E. Web-scale pharmacovigilance: listening to signals from the crowd. J Am Med Inform Assoc 2013; 0:1-5.

37. Ayers JW, Althouse BM, Allem JP, Ford DE, Ribisl KM, Cohen JE. A Novel Evaluation of World No Tobacco Day in Latin America. J Med Internet Res 2012;14(3):e77

38. Hill S, Mao J, Ungar L, Hennessy S, Leonard CE, Holmes J. Natural Supplements for H1N1 Influenza: Retrospective Observational Infodemiology Study of Information and Search Activity on the Internet. J Med Internet Res 2011;13(2):e36

Tables

Table 1. The chi^2 contingency table for a given drug and ADR, used for computing the QLRS score

		User queried for the drug?	
		No	Yes
Number of times that the user queried for the ADR	Before Day 0	N_{11}	N_{12}
	After Day 0	N_{21}	N_{22}

Table 2. Spearman's rank correlation between QLRS and the number of ADR reports in AERS (with MDADRs removed). For statistically significant correlations, p-values are provided in parentheses (n=45). MDADRs shown in bold are those emphasized in QLRS, and those in regular type are prominent in AERS.

Drug	**ρ1**	**MDADRs**
Advair	0.28	Anxiety, **Apnea**, Chest pain, Cough, **Weight gain**
Aranesp	0.30	Asthenia, Back ache, Back pain, **Edema**
Diovan	0.34	Chest pain, **Cramp**, **Sleepy**, Wound
Effexor	0.54 ($8*10^{-4}$)	Nausea, **Phobia**, **Sleepy**, **Weight gain**
Enbrel	0.39 (0.02)	Back pain, Cough, Diarrhea, Fever, **Weight gain**
Lipitor	0.54 ($9*10^{-4}$)	Asthenia, Constipation, Diarrhea, Dizziness, Nausea
Mabthera	0.38 (0.01)	Chest pain, Fever, Headache, Malaise, **Wound**
Nexium	0.45 ($8*10^{-3}$)	Abdominal pain, **Tired**, **Weak**, **Weight gain**
Norvasc	0.34	**Apnea**, Constipation, **Cramp, Tired**, **Weight loss**
Pantoloc	0.49 (10^{-3})	Chest pain, Fever, Headache, Malaise, Nausea
Pantozol	0.51 ($5.7*10^{-4}$)	Chest pain, Fever, Headache, Malaise, Nausea
Plavix	0.25	Back pain, Chest pain, Cough, Paresthesia
Protonix	0.25	Abdominal pain, Diarrhea, Nausea, **Vomit**
Remicade	0.37 (0.04)	Chest pain, Fever, **Infertility**, Paresthesia, Rash
Risperdal	0.40 (0.02)	Diarrhea, Headache, Insomnia, **Weight gain**
Rituxan	0.23	Abdominal pain, Diarrhea, Paresthesia, **Weak**
Seretide	0.41 ($4*10^{-3}$)	Chest pain, **Dyspnea**, Headache, Malaise, Nausea
Seroquel	0.48 ($3*10^{-3}$)	**Apnea**, Dizziness, Headache, **Weight gain**
Singulair	-0.06	**Apnea**, Dizziness, Insomnia, **Tired**
Zyprexa	0.61 ($2*10^{-4}$)	Constipation, Diarrhea, Nausea, Paresthesia, Somnolence

Table 3. Accuracy of ADR identification using QLRS, tested against the SIDER dataset (MDADRs removed).

Drug	F-measure	AUC
Advair	0.77	0.67
Diovan	0.43	0.71
Effexor	0.94	0.67
Lipitor	0.76	0.7
Pantoloc	0.44	0.57
Pantozol	0.44	0.64
Plavix	0.55	0.59
Singulair	0.52	0.64

Table 4. MDADRs we identified (the number of drugs in which each MDADR appeared, out of the 20 drugs analyzed, is shown in parentheses).

MDADRs overrepresented in AERS	**MDADRs overrepresented in the query logs**
Abdominal pain (3)	Apnea (4)
Asthenia (2)	Cramps (2)
Back pain (3)	Sleepy (2)
Chest pain (8)	Tired (3)
Constipation (3)	Weak (2)
Cough (3)	Weight gain (6)
Diarrhea (8)	Wound (2)
Dizziness (3)	
Fever (5)	
Headache (6)	
Insomnia (2)	
Malaise (4)	
Nausea (7)	
Paresthesia (4)	

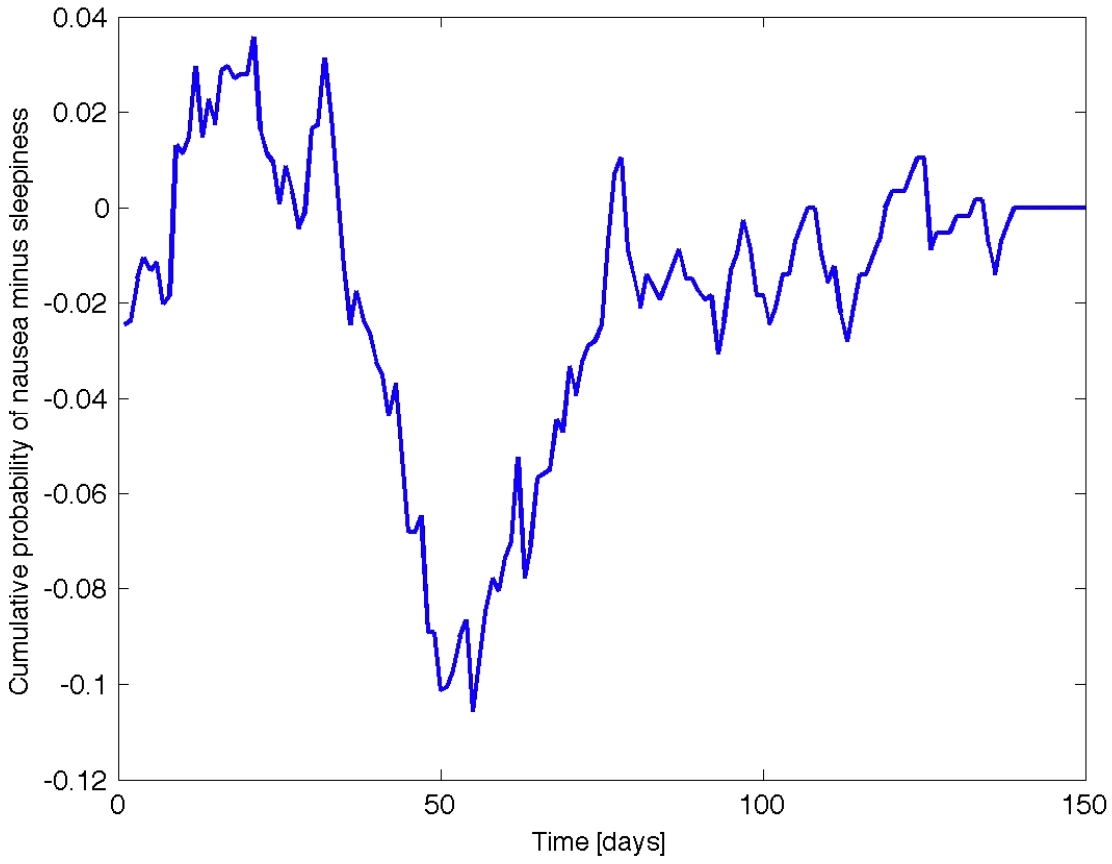
Table 5. Spearman's rank correlation between the ADRs of generic drugs as identified by QLRS and by the number of reports in AERS (MDADRs removed according to the raw report counts in AERS). QLRS scores and AERS counts for the generic drugs were computed by aggregating over multiple brand names of the same generic drug. For statistically significant correlations, p-values are provided in parentheses (n=45).

Brand names	**Generic name**	**Averaged ρ1 of individual brand names**	**ρ1 aggregated over all the brand names of the same generic drug**
Procrit, Eprex	Erythropoietin	0.35	0.72 ($1*10^{-7}$)
Neulasta, Neupogen	Filgrastim	0.43	0.54 (0.003)
Lantus, Humalog	Insulin analog	0.43	0.60 ($2*10^{-5}$)
Avonex, Rebif	Interferon beta-1a	0.60	0.83 ($2*10^{-8}$)
AcipHex, Pariet	Rabeprazole	0.21	0.53 ($2*10^{-4}$)
Protonix, Pantozol, Pantoloc	Pantoprazole	0.42	0.67 ($4*10^{-7}$)
TriCor, Lipanthyl	Fenofibrate	0.40	0.70 ($1*10^{-7}$)
Rituxan, MabThera	Rituximab	0.31	0.52 ($3*10^{-4}$)
Advair , Flovent	Fluticasone	0.31	0.64 ($2*10^{-6}$)
Cozaar , Hyzaar	Losartan	0.42	0.56 ($7*10^{-5}$)
Losec , Prilosec	Omeprazole	0.27	0.65 ($2*10^{-6}$)
Paxil , Seroxat	Paroxetine	0.13	0.44 ($2*10^{-3}$)
Avandamet , Avandaryl, Avandia	Rosiglitazone	0.38	0.67 ($4*10^{-7}$)
Imigran , Imitrex	Sumatriptan	0.26	0.60 ($1*10^{-5}$)

Table 6. Spearman's rank correlation between the ADRs of pairs of drugs identified by QLRS and by the number of reports in AERS (MDADRs removed according to the raw report counts in AERS).

Drug 1	**Drug 2**	**ρ1**	**ρ2**	**ρ3**
Risperdal	Seroquel	0.27	-0.08	-0.16
Effexor	Advair	0.28	-0.11	0.04
Zyprexa	Seroquel	0.1	0.15	-0.05
Advair	Lipitor	-0.21	0.05	0.02
Plavix	Lipitor	0.3	0.29	0.24
Lipitor	Effexor	0.14	-0.18	-0.02
Advair	Plavix	0.14	0.15	0.05
Nexium	Plavix	0.19	0.11	0.33
Seroquel	Effexor	0.32	0.15	0.11
Lipitor	Nexium	0.12	0.07	0.14

Figure 1: Temporal behavior of ADRs. The difference between the cumulative probabilities of the ADRs “nausea" and “sleepiness" for the drug Effexor. The ADR highly ranked by QLRS (“sleepiness”) has a much later onset (45-75 days).



Appendix

Identification of Adverse Drug Reactions in Vaccines

Another class of commonly prescribed drugs is vaccines, which are administered mostly to children, but also to adults. Though serious ADRs of vaccines are extremely rare, less severe reactions are known to occur. The FDA keeps track of such adverse reactions using a database similar to AERS, called the Vaccine Adverse Event Reporting System (VAERS). We used VAERS as the reference dataset for identification of ADRs in vaccines. VAERS data was downloaded from the FDA VAERS website, and included reports submitted between January 2004 and December 2009.

Using our methodology, we compared the ADRs identified by QLRS with those reported to VAERS, for the 11 vaccines most commonly mentioned in the queries (see Table S3). In general, correlations between ADRs identified by our method and those registered in VAERS are high. Indeed, for 10 of the 11 vaccines the Spearman correlation was statistically significant (after removal of MDADRs). The MDADRs emphasized in VAERS were “diarrhea", “fever", “nausea", and “headache". On the other hand, the most prominent MDADRs according to QLRS were “tiredness" and “weight loss".

These findings demonstrate that the proposed approach for identifying ADRs of pharmaceutical drugs can also be applied to vaccines, which are usually administered only a limited number of times to each patient.

Table S1. Spearman's rank correlation between QLRS and the number of ADR reports in AERS, EBGM, and IC (with MDADRs removed according to the number of reports in AERS) for the top 100 drugs. For statistically significant correlations, p-values are provided in parentheses (n=45).

Drug	**ρ1**	**ρ2**	**ρ3**
Abilify	0.05	0.04	0.04
Aciphex	-0.02	-0.19	-0.19
Actonel	0.05	0.18	0.18
Actos	0.16	0.3	0.3
Advair	0.28	0.14	0.14
Aranesp	0.32	0.34 (0.04)	0.34 (0.04)
Atacand	0.39 (0.02)	0.09	0.09
Augmentin	0.27	0.17	0.17
Avandamet	0.39 (0.01)	0.4 (0.009)	0.4 (0.009)
Avandaryl	0.4 (0.009)	0.33 (0.03)	0.33 (0.03)
Avandia	0.35 (0.04)	0.32	0.32
Avastin	-0.01	-0.18	-0.18
Avonex	0.48 (0.002)	0.47 (0.003)	0.47 (0.003)
Betaseron	0.51 (0.0007)	0.47 (0.002)	0.47 (0.002)
Celebrex	0.16	0.08	0.08
Cellcept	0.61 ($<10^{-5}$)	0.68 ($<10^{-5}$)	0.68 ($<10^{-5}$)
Cipralex	0.74 ($<10^{-5}$)	0.66 ($<10^{-5}$)	0.66 ($<10^{-5}$)
Clexane	0.72 ($<10^{-5}$)	0.69 ($<10^{-5}$)	0.69 ($<10^{-5}$)
Copaxone	0.26	0.13	0.13

Coreg	0.22	0.09	0.09
Cozaar	0.48 (0.003)	0.21	0.21
Crestor	0.26	-0.07	-0.07
Cymbalta	0.42 (0.01)	0.14	0.14
Delix	0.35 (0.02)	0.34 (0.03)	0.34 (0.03)
Depakote	0.42 (0.01)	0.11	0.11
Detrol	0.71 ($<10^{-5}$)	0.33 (0.03)	0.33 (0.03)
Diovan	0.34	0.1	0.1
Duragesic	0.46 (0.002)	0.28	0.28
Effexor	0.54 (0.0008)	0.23	0.23
Eloxatin	0.49 (0.0008)	0.44 (0.004)	0.44 (0.004)
Enbrel	0.39 (0.02)	0.32	0.32
Eprex	0.33 (0.03)	0.28	0.28
Erbitux	0.31 (0.04)	0.33 (0.03)	0.33 (0.03)
Evista	0.39 (0.02)	0.34 (0.04)	0.34 (0.04)
Flomax	-0.08	-0.33	-0.33
Flovent	0.34 (0.05)	0.21	0.21
Floxin	0.08	0.1	0.1
Fosamax	-0.02	0.08	0.08
Gemzar	0.5 (0.0008)	0.49 (0.001)	0.49 (0.001)
Gleevec	0.18	0.08	0.08
Harnal	0.38 (0.01)	0.29	0.29
Herceptin	0.77 ($<10^{-5}$)	0.65 ($<10^{-5}$)	0.65 ($<10^{-5}$)

Humalog	0.48 (0.002)	0.48 (0.003)	0.48 (0.003)
Humira	0.58 (0.0001)	0.51 (0.001)	0.51 (0.001)
Hyzaar	0.36 (0.03)	0.15	0.15
Imigran	0.33 (0.03)	0.2	0.2
Imitrex	0.18	0.04	0.04
Kaletra	0.58 ($<10^{-5}$)	0.45 (0.003)	0.45 (0.003)
Lamictal	0.27	-0.18	-0.18
Lantus	0.38 (0.02)	0.26	0.26
Levaquin	0.61 (0.0001)	0.29	0.29
Lexapro	0.12	0.13	0.13
Lipanthyl	0.34 (0.03)	0.34 (0.03)	0.34 (0.03)
Lipitor	0.54 (0.0009)	0.2	0.2
Losec	0.3	0.07	0.07
Lotrel	0.42 (0.01)	0.19	0.19
Lovenox	0.35 (0.04)	-0.03	-0.03
Lupron	0.08	-0.2	-0.2
Lyrica	-0.12	-0.24	-0.24
Mabthera	0.38 (0.01)	0.3	0.3
Micardis	0.24	0.3	0.3
Neorecormon	0.34 (0.03)	0.23	0.23
Neulasta	0.53 (0.0003)	0.42 (0.006)	0.42 (0.006)
Neupogen	0.33 (0.04)	0.33 (0.04)	0.33 (0.04)
Nexium	0.45 (0.008)	0.12	0.12

Norvasc	0.34	0.14	0.14
Olmetec	0.51 (0.0007)	0.4 (0.01)	0.4 (0.01)
Pantoloc	0.49 (0.001)	0.46 (0.002)	0.46 (0.002)
Pantozol	0.51 (0.0006)	0.45 (0.003)	0.45 (0.003)
Pariet	0.43 (0.004)	0.38 (0.01)	0.38 (0.01)
Paxil	-0.05	-0.1	-0.1
Pegasys	0.49 (0.001)	0.39 (0.01)	0.39 (0.01)
Plavix	0.25	-0.04	-0.04
Pravachol	0.21	-0.1	-0.1
Premarin	0.34	0.01	0.01
Premphase	0.4 (0.01)	0.31 (0.04)	0.31 (0.04)
Prempro	0.62 ($<10^{-5}$)	0.5 (0.001)	0.47 (0.003)
Prevacid	0.29	-0.05	-0.05
Prevnar	0.6 ($<10^{-5}$)	0.6 ($<10^{-5}$)	0.6 ($<10^{-5}$)
Prilosec	0.23	0.06	0.06
Procrit	0.37 (0.03)	0.43 (0.01)	0.43 (0.01)
Prograf	0.61 ($<10^{-5}$)	0.62 ($<10^{-5}$)	0.62 ($<10^{-5}$)
Protonix	0.25	0.04	0.04
Pulmicort	0.53 (0.0006)	0.49 (0.002)	0.49 (0.002)
Rebif	0.71 ($<10^{-5}$)	0.61 ($<10^{-5}$)	0.61 ($<10^{-5}$)
Remicade	0.36 (0.04)	0.41 (0.02)	0.41 (0.02)
Risperdal	0.4 (0.02)	0.01	0.01
Rituxan	0.23	0.3	0.3

Seretide	0.58 (0.00006)	0.25	0.23
Seroquel	0.48 (0.003)	-0.11	-0.11
Seroxat	0.3	0.24	0.24
Singulair	-0.06	-0.13	-0.13
Spiriva	0.46 (0.006)	0.33	0.33
Symbicort	0.27	0.21	0.21
Synagis	0.63 ($<10^{-5}$)	0.65 (0.000004)	0.65 (0.000004)
Tamiflu	0.69 ($<10^{-5}$)	0.69 (0.000001)	0.69 (0.000001)
Taxotere	0.52 (0.0004)	0.38 (0.02)	0.38 (0.02)
Topamax	0.36 (0.04)	0.26	0.26
Toprol	0.51 (0.002)	0.14	0.14
Tricor	0.45 (0.006)	0.37 (0.03)	0.37 (0.03)
Tritace	-0.08	-0.08	-0.08
Truvada	0.45 (0.003)	0.59 (0.00004)	0.59 (0.00004)
Valtrex	0.12	0.02	0.02
Viagra	0.28	-0.07	-0.07
Vytorin	0.39 (0.02)	0.41 (0.01)	0.41 (0.01)
Wellbutrin	0.08	0.11	0.11
Xalatan	0.3	0.3	0.3
Zetia	0.26	0.06	0.06
Zocor	0.09	-0.04	-0.04

Zofran	0.1	0.15	0.15
Zoladex	0.46 (0.002)	0.33 (0.03)	0.33 (0.03)
Zoloft	0.12	0.19	0.19
Zometa	0.33 (0.04)	0.27	0.27
Zyprexa	0.61 (0.0002)	0.11	0.11

Table S2. Accuracy of ADR identification using QLRS for the top 100 drugs, tested against the SIDER dataset (MDADRs removed).

Drug	F-measure	AUC
Abilify	0.79	0.86
Aciphex	0.46	0.86
Actonel	0.64	0.81
Advair	0.78	0.7
Atacand	0.87	0.76
Avandia	0.83	0.55
Blopress	0.50	0.39
Celebrex	0.61	0.79
Coreg	0.79	0.82
Cozaar	0.74	0.8
Crestor	0.69	0.75
Cymbalta	0.72	0.81
Delix	0.57	0.33
Diovan	0.71	0.78
Duragesic	0.67	0.5
Effexor	0.67	0.9
Eloxatin	0.71	0.6
Eloxatine	0.50	0.44
Evista	0.88	0.6
Flomax	0.62	0.64

Flovent	0.67	0.6
Gemzar	0.69	0.5
Gleevec	0.68	0.61
Harnal	0.60	0.29
Hyzaar	0.69	0.75
Imigran	0.55	0.42
Imitrex	0.47	0.86
Lamictal	0.61	0.9
Lipanthyl	0.56	0.33
Lipitor	0.70	0.75
Losec	0.72	0.62
Lotrel	0.63	0.75
Lyrica	0.51	0.92
Micardis	0.55	0.71
Pantoloc	0.57	0.44
Pantozol	0.64	0.44
Pariet	0.65	0.46
Paxil	0.73	0.83
Plavix	0.59	0.55
Pravachol	0.66	0.65
Prilosec	0.58	0.9
Prograf	0.79	0.83
Protonix	0.73	0.9

Pulmicort	0.88	0.8
Risperdal	0.88	0.94
Seloken	0.50	0.44
Seroquel	0.66	0.89
Seroxat	0.54	0.44
Singulair	0.64	0.52
Spiriva	0.90	0.7
Symbicort	0.61	0.6
Takepron	0.50	0.58
Taxotere	0.74	0.67
Toprol	0.76	0.85
Tricor	0.72	0.78
Tritace	0.57	0.33
Viagra	0.79	0.75
Wellbutrin	0.77	0.91
Zetia	0.94	0.8
Zocor	0.55	0.57
Zofran	0.48	0.47
Zoloft	0.57	0.81

Table S3. Spearman's rank correlation between QLRS and the number of ADR reports in VAERS (with MDADRs removed according to the raw report counts in VAERS). For statistically significant correlations, p-values are provided in parentheses. Also shown are MDADRs for each drug and the main indications for the drug. MDADRs shown in bold are those emphasized in QLRS, and those in regular type are prominent in VAERS.

Vaccine	**Indications**	**ρ1**	**ρ2**	**ρ3**	**MDADRs**
DPT	Diphtheria, pertussis, tetanus	0.46 (0.01)	0.44 (0.01)	0.36 (0.04)	**Alopecia**, **Apnea**, **Dry mouth**, **Depression**, Edema
Hepatitis	Hepatitis	0.41 (0.02)	0.07	0.21	**Apnea**, **Depression**, Nausea, **Weight loss**
HPV	Human papiloma virus	0.33 (0.05)	0.26	0.15	Headache, **Infertility**, **Miscarriage**, **Tired**
IPV	Polio	0.51 (0.003)	0.48 (0.005)	0.41 (0.02)	Bleeding, Malaise, **Weight loss**, Urticaria
MMR	Measles, mumps, rubella	0.45 (0.01)	0.40 (0.02)	0.50 (0.002)	Edema, **Sleepy**, **Tired**, **Weight loss**
OPV	Polio	0.69 ($<10^{-5}$)	0.58 (10^{-4})	0.72 ($<10^{-5}$)	**Bloody show**, **Chest pain**, **Hirsutism**, **Tired**
Polio	Polio	0.51 (0.002)	0.40 (0.02)	0.41 (0.01)	**Depression** , **Itch**, **Phobia**, **Wound**

Smallpox	Smallpox	0.44 (0.01)	0.42 (0.02)	0.31	Diarrhea, Fever, Headache, Paresthesia, **Tired**
Typhoid	Typhoid fever	0.33 (0.05)	0.41 (0.01)	0.33	Diarrhea, Fever, **Weight loss**, **Xerostomia**
Varicella	Chickenpox	0.18	0.10	0.10	Fever, **Miscarriage**, **Tired**, Urticaria
Yellow fever	Yellow fever	0.46 (0.005)	0.37 (0.02)	0.55 ($5*10^{-4}$)	**Drug overdose**, Nausea, **Tired**, **Weight loss**