Early detection of chemotherapeutic skin adverse reactions via social health networks
Azadeh Nikfarjam, PhD1*, Julia D. Ransohoff, MD1*, Erik Jones, PhD3, Brian Loew, AB3, Bernice Y. Kwong, MD2**, Kavita Y. Sarin, MD, PhD2**, Nigam H. Shah, MBBS, PhD1**
*co-first authors
**co-senior authors
1Stanford Center for Biomedical Informatics Research, Stanford University School of Medicine, Stanford, California
2Stanford School of Medicine, Department of Dermatology, Stanford, California
3Inspire, Arlington, Virginia

Corresponding Author:
Julia D. Ransohoff, MD
300 Pasteur Drive, Lane L154
Stanford, California 94305-5133
650-723-6661
ransohoff@stanford.edu

Abstract

Importance: Adverse drug reactions (ADRs) occur in nearly all patients on chemotherapy, causing morbidity and therapy disruptions. Detection of such of ADRs is limited in clinical trials, which are underpowered to detect rare events. Early recognition of ADRs in the post-marketing phase could substantially reduce morbidity and decrease societal costs. Internet community health forums provide a mechanism for individuals to discuss real-time health concerns and can enable computational detection of ADRs.

Objective: To identify cutaneous ADR signals in social health networks and compare the frequency and timing of these ADRs to clinical reports in the literature.

Design: We present a natural language processing (NLP) based ADR signal generation pipeline based on patient posts on internet social health networks. We identify user posts from Inspire health forum related to two chemotherapy classes: erlotinib, an epidermal growth factor receptor inhibitor, and nivolumab and pembrolizumab, immune checkpoint inhibitors. We extract mentions of ADRs from unstructured content of patient posts. We then perform population-level association analyses and time-to-detection analyses.

Results: Our system detected ADRs from patient reports with high precision (0.90) and at frequencies comparable to those documented in the literature, but an average of 7 months ahead of their literature reporting. Known ADRs associated with higher proportional reporting ratios compared to negative controls, demonstrating the robustness of our analyses. Additionally, we discovered the novel ADR of hypohidrosis reported by 23 patients in erlotinib related posts; this ADR was absent from 15 years of
literature on this medication and we recently reported the finding in a clinical oncology journal.

**Conclusions and relevance:** Several hundred million patients report health concerns in social health networks, yet this information is markedly underutilized for pharmacosurveillance. We demonstrate the ability of an NLP-based signal generation pipeline to accurately detect patient reports of ADRs months in advance of literature reporting, and the robustness of statistical analyses to validate system detections. Our findings suggest the important contributions that social health network data can play in contributing to more comprehensive and timely pharmacovigilance.

**Introduction**

Adverse drug reactions (ADRs) are important public health issues, causing considerable patient harm with a high health care cost [1,2]. Serious ADRs occur in over two million patients annually in the US, resulting in 100,000 deaths [2]. ADRs are difficult to be comprehensively characterized during pre-market trials. Many serious ADRs are discovered several years after the drug has been on the market [3]. Trials are limited in ADR detection because of their small sizes of patient groups, short durations, and a lack of enrolled patient diversity [4,5]. ADRs that become apparent in the post-market period result in over 2 million injuries and $75 billion annual health care cost [6,7]. Ongoing surveillance strategies are necessary to monitor the drug safety when drug use is expanded during the post-approval period.

Spontaneous reporting systems (SRS) are drug safety surveillance mechanisms designed by regulatory agencies to monitor drug safety during the post-market period. SRSs can effectively detect rare ADRs by nature of their longitudinal profiling and wide reach of included reports, but suffer due to their reliance on voluntary patient or provider reporting; it is estimated that more than 90% of ADRs are under-reported [8,9]. ADR under-detection has motivated efforts to include complementary, alternative data sources for pharmacovigilance, including electronic health records and administrative claims [10,11,12], biomedical literature [13,14], internet search logs [15], patient posts in social media [16,17] and multimodal systems that jointly analyze multiple sources of information for ADR detection [7].

Over 300 million patients seek and share health-related information from online internet forums [18]. These social health networks provide a platform for patients or caregivers to connect in discussing treatment options, drug side effects, and illness trajectories.

Several studies highlighted the importance of utilizing social media as a resource for pharmacovigilance [19]. User posts often contain informal unstructured text that makes it more challenging for medical information extraction compared to other sources. Therefore, exploring different NLP techniques in ADR concept detection from social media postings is receiving significant attention from the medical informatics community
However, there has been relatively fewer studies focusing on drug-ADR signal generation methods based on social media postings [20].

Here we use Inspire (https://www.inspire.com/), one of the largest online social health networks that contains over 12 million health-related patient posts including discussions of therapy responses, adverse drug reactions, and supplemental treatments [21]. We present an ADR signal generation pipeline based on patient posts in social health networks and compare the timing and the rate of such ADRs with those published in clinical literature. We also demonstrate the capacity for early detection of ADRs using Inspire content.

We focus on two classes of cancer therapy drugs. New cancer therapeutics, the targeted small molecule inhibitors and monoclonal antibodies for anticancer therapies, has greatly impacted oncology treatment but have caused novel ADRs, with cutaneous ADRs occurring in nearly all patients [4,22]. We study the association of cutaneous ADRs with the selected targeted cancer therapy drugs reported in patient postings in Inspire.

Methods

We defined a set of common and rare ADRs to study for their association with two class of drugs: epidermal growth factor receptor (EGFR) inhibitor, erlotinib, and the immune checkpoint programmed cell death–1 (PD-1) inhibitors, nivolumab and pembrolizumab. We focus on eight skin-related ADRs in this work: Rash, Acne, Pruritus (itchy skin), Paronychia (nail changes), Xerosis (dry skin), Hypohidrosis, Bullous eruption (blister), Psoriasis. Figure 1 shows our pipeline for identifying ADR signals from patient posts in Inspire.
Figure 1: Pipeline to identify adverse drug reaction signals associated with EGFR and PD-1 inhibitors in social health networks. For drug-ADR pair extraction, for each drug, we generated a collection of user posts containing at least one mention of the drug. This drug corpus is then processed via DeepHealthMiner to recognize mentions of ADRs. The extracted mentions are then mapped to the corresponding UMLS CUIs. The identified drug-ADR pairs and the related details, for both target and comparison group, are then stored in a relational database. The PRR calculated to quantify the drug-ADR relations. We calibrated the score using the distribution of the negative control set.

Drug Corpus Generation

The Inspire dataset consists of 7,320,546 discussion posts from 2005-2016. For each drug in the study, we generated a corpus defined as a collection of user posts from Inspire, with every post containing at least one mention of keywords corresponding to the drug. To retrieve relevant posts from the Inspire dataset, we used regular expressions, a simple text processing method for string matching. We identified 55,778 posts for erlotinib, and 15,738 for the combined PD-1 inhibitors (nivolumab and pembrolizumab). Because nivolumab and pembrolizumab were more recently introduced and therefore have less discussion content, but nearly identical mechanisms, we combined posts for the two drugs referencing the same ADR.
Detection of Adverse Drug Reactions

To extract ADR mentions from user posts, we use DeepHealthMiner (DHM), a neural network-based named entity recognition (NER) system that is specifically trained to extract drug safety-related entities from user-generated content in social media [23]. DHM is a supervised feed forward neural network that is trained to identify two different entity types: ADRs and drug indications. The original, labeled training data is based on a dataset from the DailyStrength (www.dailystrength.org), an online health community that contains patient-generated reports of treatment experiences very similar to those from Inspire. The labeled training data are the sentences from patient posts manually annotated for the entity span (start and end position offsets) and the type (ADR, Indication). For more information about the data and annotation details please refer to prior publications [24,25]. The unlabeled sentences from user posts are also used for unsupervised training of the word embeddings, which are used as vectors representing the input tokens to the NER system.

We retrained DHM using both labeled and unlabeled posts from Inspire. We retrained 150-dimensional word embeddings by adding the 7 million Inspire posts to the original unlabeled sentences. We manually labeled additional 200 Inspire posts, following the original annotation guideline [24,25], to re-train DHM using Inspire content. The extracted ADR mentions along with the information about the related drugs were stored in a database for further analyses of the associations among drugs and ADRs.

Normalization of extracted mentions

To map the extracted ADR mentions to UMLS CUIs, we generated a lexicon of the eight ADR concepts. We defined a set of seed CUIs for these ADRs as shown in the supplement file (seed_ADR_concepts). Using UMLS hierarchy we expand every seed concept by adding synonyms and alternative names. For “nail changes” we further expanded by including the children of the seed concepts (is-a relation in the hierarchy). Every lexicon entry has a name, UMLS CUI, related seed ADR concept, and an identifier that we defined to group all items corresponding to a specific ADR. As an example, consider the following lexicon entry: ingrown toenail (C0027343) is an expanded concept based on Disorder of nail (C0027339), and “nail changes” is the human readable identifier for the ADR. Additionally, for every ADR, we added a set of colloquial phrases to the lexicon (listed in supplement file seed_ADR_concepts).

Finally, we used Lucene (https://lucene.apache.org/) to index the lexicon. For mapping an extracted ADR mention to the lexicon, we passed the text span extracted by DHM plus a configurable context window of n tokens (n=3) before and after it as a query to the Lucene index. We retrieved a ranked list of the matched, relevant ADR concepts and chose the top ranked concept from this list.
Computing drug and ADR associations

The co-occurrence of a drug and a reported ADR in a post can be considered as a potential association. To quantify the strength of an association between a drug and ADR, we calculated Proportional Reporting Ratio (PRR), a statistic widely used for ADR signal generation from spontaneous reporting databases [20,26] and originally introduced by Evans et al.[27] The PRR quantifies ADR signals by comparing the frequency at which an ADR is reported with a drug of interest compared to the frequency at which the ADR is reported with other drugs (comparison drugs) in the database.

We defined the comparison drugs by sorting drugs mentioned in Inspire by discussion frequency and selecting those most highly-discussed in Inspire discussion posts (frequency > 5000). To reduce noise, we excluded over-the-counter drugs with common indications (e.g. pain and allergy drugs), as well as several anti-inflammatory drugs (e.g. prednisone and Humira) that have indications that match some of our rare ADRs (e.g. bullous eruption and psoriasis), given that patients discussing these drugs were almost exclusively discussing their indications rather than experienced ADRs. The final comparison drug list includes 27 drugs listed in the supplement file (comparison_drugs.txt).

PRR for an adverse reaction R and a drug D is calculated based on the following formula:

\[
PRR(D, R) = \frac{\text{count}(D \cap R)}{\text{count}(D)} \cdot \frac{\text{count}(\neg D \cap R)}{\text{count}(\neg D)}
\]

We define count(D \cap R) as the number of unique users that have reported both D and R in a post, count(D) is the total number of users reporting any ADR for drug D (including all extracted mentions by DHM, e.g. weight gain or fatigue), count(\neg D \cap R) is the total number of unique users that reported R for comparison drugs (excluding D), and count(\neg D) is the total number of unique users that reported at least one ADR for any drug except D.

Calibrating a threshold for the proportional reporting ratio

We performed empirical calibration to find the threshold at which the PRR may represent a true ADR signal [28,29]. We considered the distribution of PRR scores for a set of drug-condition pairs with no known associations. To create this negative control set, we first paired all 27 comparison drugs and the eight target ADRs, totaling 216 drug-condition pairs. We then excluded the drug-condition pairs with known associations (ADR or indication) based on Medi-span ADR database (Wolters Kluwer Health, Indianapolis, IN) and SIDER [30], leaving 86 pairs without any documented associations.
It is important to note that a drug and a condition can be associated for reasons other than ADR or indication and may therefore co-occur frequently within the negative control set. We identified two common reasons that a drug (drug\textsubscript{x}) and a condition (condition\textsubscript{y}) with no documented association co-occur frequently in user posts, therefore representing false positives. Firstly, related to prescribing patterns: if a drug or a set of similar drugs are commonly prescribed with drug\textsubscript{x}, while the reported ADR condition is not attributed to drug\textsubscript{x}, the ADR will appear more frequently with all co-prescribed drugs. For example, mentions of nail changes happen frequently with zoledronic acid, however it is a true ADR associated with capecitabine, which is frequently prescribed with zoledronic acid; zoledronic acid is not responsible for the ADR. Secondly, if condition\textsubscript{y} is one of the syndromic conditions related to the indication for drug\textsubscript{x}, the related condition will co-occur with the drug at a similar rate to the true indication. For example, metformin can be prescribed for polycystic ovary syndrome (PCOS); patients with PCOS often have acne. However, acne is not an ADR for metformin, but will co-occur frequently with mention of metformin by way of acne's association with PCOS.

Because of the potential for observing such drug-condition associations, we manually reviewed the negative control set to exclude them, leaving 81 drug-ADR pairs with no association. The negative control set can be downloaded from here (negative\_control\_pairs.tsv).

**Results**

**Finding drug-ADR signals from forum posts**

We identified 50,574 Inspire posts related to EGFR inhibitors and 16,598 related to checkpoint inhibitors; a total number of 13,600 ADR concepts extracted from the former and 812 concepts from the latter mapped to the 8 target ADRs. To assess system performance in extraction and normalization of ADR concepts, we manually validated the extracted ADRs. Our validation set consisted of 120 posts including 15 randomly selected posts for every ADR (8 ADRs) that the system extracted for one of our target drugs.

We consider an extracted concept as a true positive if the system correctly detected the relevant span of text, classified the entity as an ADR (versus indication), and normalized the ADR to the correct concept. Benchmarking results show that our system achieves a high performance (micro-average precision of 0.90) in recognizing and normalizing the ADRs in the user posts.

The distribution of the frequencies and the calculated PRR of the skin ADRs extracted from the EGFR inhibitors related posts correlates closely with published observed rates in patients (Figure 2), further validating the pipeline for identifying ADRs. The PD-1 inhibitor ADR landscape is largely limited to case reports; forum frequencies and PRR
values correlate with published rates, suggesting that those described to-date accurately represent the emerging toxicity landscape (Figure 2).

![Graph showing Proportional Reporting Ratio (PRR) vs. Reported ADR rate in literature.]

**Figure 2:** Comparison of proportional reporting ratio of skin ADRs reported in social health forums with the ADR rate published in the literature. The ADR rate based on literature are grouped as follows: Not reported, rare = case report, reported = case series or in clinical trial, common = in significant percentage of patients in large trials. The frequency at which ADRs are talked about is shown by the size of the circle or triangle. The most common ADRs reported in the literature are also the most discussed in the patient posts.

To determine the threshold at which PRR represents true signal warranting further investigation, we plotted the PRR distribution for the 81 drug-ADR pairs in the negative control set (Figure 3). More than 95% of the pairs in this set have PRR less than 0.82; therefore, drug-ADR pairs with PRR greater than 1 can be considered as signaling true ADRs.
**Proportional Reporting Ratio Distribution for Negative Examples**

Figure 3: Proportional reporting ratio (PRR) distribution for a set of 28 negative example drugs representing 81 drug-ADR pairs; mean 0.12; median 0.2; maximum 1.4, highlighting a PRR threshold of 1 below which <5% of drug-ADR pairs have true associations.

**ADRs are described in internet forums prior to published reports**

To compare the timing of online ADR reports to initial ADR reports in the literature, we looked at both common and rare events associated with our target drugs. Papulopustular (acneiform) rash and nail-finger changes are well-reported ADRs associated with erlotinib and were first described in published case reports in September 2005 [31] and September 2006 [32], respectively. Inspire posts for these reactions appeared 5 and 3 months in advance of publication, respectively. Psoriasis in the setting of PD-1 inhibitor treatment was first documented in case reports in July 2015 [33] and May 2016 [34]. Inspire forum posts describing this ADR preceded the 2015 case report by 9 months (Figure 4A). Blistering reactions with PD-1 inhibitors were initially published as a case report in June 2015 [35] and 3-case series in May 2016 [36]. Forum
descriptions preceded the first case report by 9 months (Figure 4B). Taken together, our data suggest a significant and consistent reporting lead-time advantage in online posts. Cutaneous ADRs in both EGFR and PD-1 inhibitors have been reported to be associated with cancer response to therapy [37,38,39,40], highlighting the clinical utility of early detection.

Figure 3: Cutaneous ADRs identified in Inspire forums precede initial published clinical reports. We plotted cumulative post count (y-axis) at each date (x-axis) for time-to-detection analysis. (A) Psoriasis was first reported in the literature as individual case reports with PD-1 inhibitors in July 2015 and May 2016, respectively. Inspire users began describing psoriasis flares 9 months prior to the first case report. (B) Bullous reactions with checkpoint inhibitors were first reported in the literature as a case report with pembrolizumab in May 2015 and as a three-case series with nivolumab in May 2016. Inspire cases were reported online 9 months before the initial case report.
**Novel ADR discovery in social health networks**

In early 2017, we documented three patients on erlotinib reporting hypohidrosis. Our system detected 23 unique Inspire users reporting hypohidrosis in causal association with erlotinib (based on manual review of the posts) starting as early as 2006, with a PRR score of 1.90 (Figure). This predicted ADR is absent from over 15 years of EGFR inhibitor literature, implicating a novel ADR detected through our system extractions from online posts. EGFR is expressed in sweat glands [41] and decreases with pharmacologic EGFR inhibition [42]. Interestingly, the hypohidrotic ectodermal dysplasia phenotype is partially-mediated by decreased EGFR signaling [43], which can possibly cause hypohidrosis. These findings demonstrate the potential of ADR extraction from internet forums to not only detect ADRs earlier, but also to contribute to novel ADR discovery.

**Conclusion**

Our study demonstrates that it is possible to detect ADR signals from patient-generated social media posts an average of 7 months earlier than literature reporting, and at frequencies comparable to their eventual literature descriptions. We describe the proof-of-principle construction and validation of a signal generation pipeline for ADR detection from social health networks. We benchmark our system extractions of known drug events against their literature reports to evaluate our pipeline’s accuracy and temporal advantage. Our system is able to extract drug-ADR signals from highly unstructured online patient content with high precision.

In this work, we demonstrate the utility of mining online patient reports to identify signals for both common and rare ADRs with high precision. We envision the potential for social media-based signals to be combined with those derived from alternative modes—electronic health records, insurance claims, FDA reports—to construct the most comprehensive and dynamic catalogue of ADRs.
References

37. R. Perez-Soler, Can rash associated with HER1/EGFR inhibition be used as a marker of treatment outcome? *Oncology (Williston Park, N.Y.)* 17, 23-28 (2003); published online EpubNov (38).