DMIN - LATE BREAKING PAPER

Strategies for distributed curation of social media data for safety and pharmacovigilance

Tim A. Casperson[1], Jeffery L. Painter[2], and Juergen Dietrich[3]
[1]tim.a.casperson@gsk.com, GlaxoSmithKline, RTP, NC, USA
[2]jeffery.l.painter@gsk.com, GlaxoSmithKline, RTP, NC, USA
[3]juergen.dietrich@bayer.com, Bayer, Berlin, Germany

Abstract— *We have developed a system for trained medical experts to* curate *patient authored text in social media posts (the process of manually reviewing posts to extract medical insights).*

The system provides annotation capabilities for medication names, events, indications, drug-drug interactions as well as documenting drug use and potential adverse events. This system is currently being used to create a gold standard training set for tuning and comparing the performance of adverse drug reaction detection classification methods.

The software tools facilitate the real-time, distributed annotation and curation of social media data for use by trained medical experts in support of annotating drug names and medical conditions in social media data.

Once complete, this will represent one of the largest collections of annotated social media text available for use in adverse drug reaction detection.

Keywords: social media; adverse event detection; classification

1. Introduction

Web-Recognizing Adverse Drug Reactions (WEB-RADR) is a groundbreaking European Union (EU) Innovative Medicines Innovation funded 3-year initiative to recommend policies, frameworks, tools and methodologies by leveraging these new developments to get new insights in drug safety [1]. Among WEB-RADR's many objectives is the ability to leverage social media for discovering new insights in drug safety.

While there are many on-going efforts to build automatic classification methods for ADR (adverse drug reaction) detection in social media [2][3][4], this project aims to garner one of the largest sets of hand curated data for the purpose of establishing a *gold standard* by which future, automated classification methods may eventually be compared. This is a critical component necessary, and still missing, from the public at large who are interested in cultivating the voice of the patient through social media to expand our capabilities in monitoring the safety of actively marketed drug products.

1.1 Project CRAWL

GlaxoSmithKline, Inc., a pharmaceutical company, initiated their own project investigating the use of social media for pharmacovigilance in early 2012, and that research developed into *Project CRAWL* (Contextualization of Real World Drug Use through Social Listening), working together with an external vendor to collect, aggregate and de-identify social media data. To support all of these efforts, a software tool called *Insight Explorer* was built to help *Project CRAWL* evaluate over 40,000 social media posts, covering a wide range of products and safety related questions [5].

Throughout the *Project CRAWL* initiative, our team worked with our GSK safety scientists to build a prototype application that would enable trained medical experts to manually evaluate social media data for various outcomes related to our own marketed products. The result of this prototype application was then presented to the IMI WEB-RADR team as a potential tool for their efforts.

An example of how the original version of the tool is used is shown in Figure 1 which included the ability for a GSK safety scientist to examine social media post content and meta-data to (1) determine whether it is applicable for a safety assessment or spam, (2) determine whether or not the poster talks about a particular product of interest (*in-scope* for review) or any other products, (3) capture basic demographic information about the poster or patient, and finally (4) to answer an array of questions about the data such as "is this patient seeking information?" or "is this a complaint about a product?".

When we learned of the WEB-RADR project, it was determined that the IMI would also require sophisticated tools to assist their team in meeting the primary goals of manually curating social media data in order to build a "gold standard" of curated *patient authored text*[1], and we launched a new project to make modifications to *Insight Explorer* to help with this effort. This paper reports on our findings to date in building a distributed tool for the curation of social media for a large scale project around safety and pharmacovigilance.

[1]We acknowledge that it is not necessarily the case that when a person posts on social media that the person is talking about their own experiences. Instead of belaboring the nuances of whether or not a social media poster refers to him or herself directly or to another (such as a friend or family member), we will simply refer to all of these social media posts as *patient authored text*.

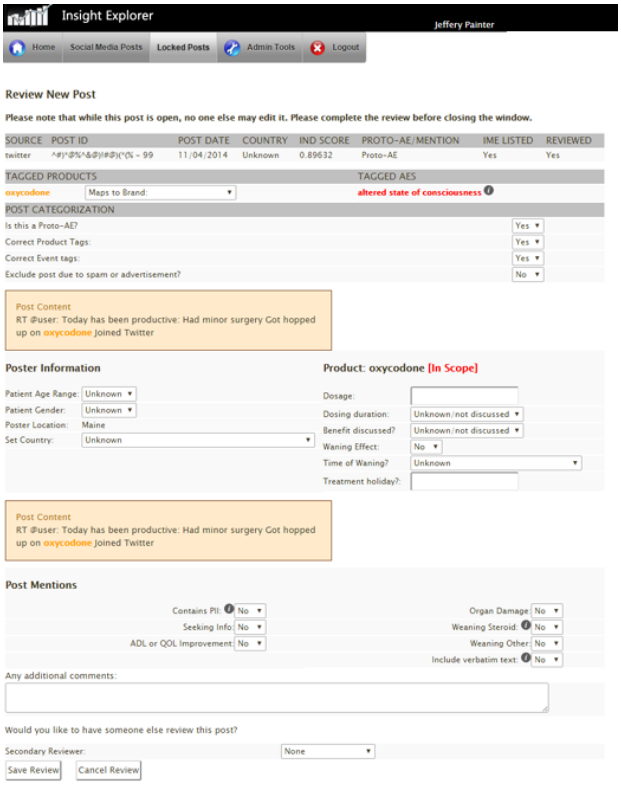

Fig. 1: Insight Explorer: Review Post

1.2 WEB-RADR Initiative

From the project charter, Johan Ellenius, UMC, explains it best:

“We are focusing on developing methods to automatically annotate medication names, Proto-AEs and unrelated events. One task in our work stream is to compare the classification performance of various methods for named entity recognition of these entities. In order to do the comparison, we must have a gold standard classification of the entities that we attempt to classify. So for example, in order to calculate the recall of our method for medication name annotation, we must be able to relate the correct identification of medication names with all occurrences of medication names in the tweets and posts. This is why we need a gold standard for medication names."

“The predictive models that perform the named entity recognition may draw on a lot of different information extracted from the posts. That is indeed part of the challenge, to identify purposeful features to use in the models. However, they don’t need to be specifically gold standard classified, because the aim of our research is not to assess e.g. how well we are able to determine location of tweets or the sex of the author or something else that might serve as useful predictors in our models."

2. Background

Safety surveillance for adverse event (AE) detection encompasses a history which may well serve as a blueprint for the evolution of data mining in general. Since the discipline of pharmacovigilance first emerged in the 1960s, the strengths and weaknesses of this science have become quite evident. The early days of AE detection relied on the use of self-reporting from patients who experienced an unwanted side effect as well as their health care professionals through systems such as the FDA AERS[6] (adverse event reporting system) as well as homegrown solutions by each of the pharmaceutical companies [7].

The primary issue surrounding these self-reporting systems is that there was never evidence of the total population exposed to medications, and therefore, a much needed denominator to determine the likelihood of the event was missing [5].

Using observational data such as electronic health records and insurance claims data (where it is possible to compute a denominator) can supplement findings from spontaneous adverse event reports, but this type of data has additional problems with bias (e.g. indications are reported for insurance claims purposes rather than to indicate the exact patient experience).

In addition, access to these types of data typically require signing expensive license arrangements with individual data providers, and again, the population under study may be subject to underlying biases in the market these payers and providers serve (e.g. those who are insured are typically younger and healthier than the overall population, or a Medicaid or Medicare population which may be sicker than normal) [8].

By using social media sources, it is hoped that for the first time, it may be possible to capture the true voice of the patient in their own words. However, people do not typically speak the same language as health care professionals, and therefore, one must learn how to discern what patients mean when they express their experiences and interactions with health care systems through social media.

There is a need for a highly specialized social media curation tool for medical within the IMI (Innovative Medical Initiative). The IMI has kicked off a 3 year project which is partly for the study of publicly available patient authored text on social media. This project is called WEB-RADR. (WEB-Recognizing Adverse Drug Reactions)

The WEB-RADR project aims to supplement our current knowledge of patient authored text by manually curating thousands of social media posts gathered from Twitter and Facebook. The amount of data that the expert curators must process however required construction of new software tools and an infrastructure to help them distribute the workload evenly and minimize the risk of losing the valuable information they are collecting.

Early social media post curation was initially accomplished by using spreadsheets as the main tool. This process was tedious and not well suited for collaborative curation across a geographically distributed team.

Our team has custom designed our social media tool, *Insight Explorer*, for WEB-RADR's social media curation needs. This is a web based tool allowing remote access to a distributed curation team and is highly scalable.

3. Distributed Curation

The problem to be solved with distributed social media curation has many facets. Curation teams are scattered geographically into different time zones and need a customized system that enables them to collaborate and work remotely. In addition, if a team annotates social media posts in a traditional tool such as a spreadsheet, the data is very wide, is difficult to read, and is cumbersome to divide up posts between curators and collaborate across the team.

Social media posts need the information within it tagged and categorized into groups such as medication reported, brand name, generic name, medical events and indications reported such as flu or fever, as well as poster information if given, and any further comments from the curator. Furthermore, curators in our case are usually highly qualified medical experts. Any small improvements that can be made to their curation process that can save time will affect cost drastically.

Before the implementation of *Insight Explorer* as the WEB-RADR curation tool, curators would divide up posts equally among team members and place them in a spreadsheet for review. If the teams were to grow or shrink size once the curation process had begun, it would cause additional complications partly because it would be monotonous to reallocate the posts based on the new team size while ensuring that previously curated data was preserved.

Furthermore, the analysis of the data after curation would be more difficult because the data would need to go through the additional step of being read into an analytics platform. In *Insight Explorer*, the curation data is automatically saved in a relational database.

Our approach to this multi-faceted curation problem was to create a new version of *Insight Explorer*, specially customized to meet the needs of a diverse geographically distributed team of reviewers.

We worked with the WEB-RADR project leads to gather system requirements which have drastically increased the speed with which the teams can curate the data, as well as collecting all the results in a centralized data repository for analysis.

3.1 Incorporating Structured Terminologies

One of the requirements was to add a feature to enable the automatic look up and association between generic and brand name drugs. To accomplish this, we utilized the

Fig. 2: Brand Name and Generic Name are automatically looked up based on the reported term

Unified Medical Language System (UMLS) to develop our extracts of the *RxNorm* and MedDRA dictionaries using the UMLS MetaThesaurus [9] .

Much of our group's prior history was rooted in the development of tools and methods to exploit the use of standardized structured terminologies in support of safety activities [10], [11], [12]. As such, we have taken much of that history and knowledge and applied it to the development of tools such as this one to ease the curation process for our trained medical experts.

To assist our users from having to perform the unnecessary task of leaving the review screen in order to lookup codes from a standard terminology, we sought a method to incorporate them directly into the user experience to (1) improve the consistency of the annotations created by the expert reviews and (2) to reduce the time and effort of these valuable human resources from having to consult external tools to determine appropriate coding of the social media posts.

Therefore, we were able to add both a medical condition terminology (MedDRA) and a standardized drug catalog (RxNorm) to support these goals. The most recent version of the UMLS (2015AA) at the time the system was being developed was used to generate extracts of both the MedDRA terminology as well as RxNorm.

MedDRA was established in the late 1990's by the ICH (international Council for Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use) to promote the sharing of medical information about medical conditions internationally. It serves as a standard terminology for classifying medical conditions into "Preferred Terms", also known as PT's (e.g. influenza) and Lowest Level Terms (LLTs such as flu) [17].

While MedDRA itself is a hierarchical terminology, we were only interested in the support of mapping user reported terms to lower level terms within the MedDRA ontology [13]. Since MedDRA then maps every low level term (LLT) to a preferred term (PT), those mappings are automatically displayed to the user once a match has been identified. If the expert reviewer knows the preferred term directly, they can also enter that and the corresponding LLT is automatically populated for them instead.

There were instances of confusion among curators which leave room for improvement in this area. In MedDRA version 17.1, not every LLT necessarily maps to a PT. One of our user's did experience this in the case of the term "locally advanced breast cancer", which is a valid LLT in

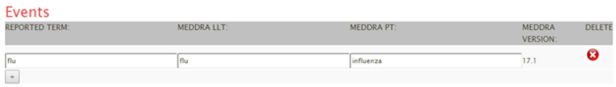

Fig. 3: Patient events are recorded and the patient reported text is mapped to a MedDRA term

MedDRA. The user expected that this term should map to the PT for “breast cancer", but it does not. Within the the UMLS concept hierarchy, the LLT was denoted within a single concept (CUI=C3495949), which did not have any corresponding PT codes. However, the UMLS does express a relationship between “locally advanced breast cancer" as being *classified_as* the PT for “breast cancer". At the moment, the current version of *Insight Explorer* is not taking advantage of these broader relationships, and is limited to more restrictive synonymy relation only.

Each of the relations between the LLT and PTs are stored within our MySQL database and loaded upon the review stage of the process to make lookup and retrieval appear in real time with minimal response gap to the end user.

This is similar to our process for mapping drug names automatically as well. In this case, we were able to extract from RxNorm each of the recognized brand names and map them directly to their generic counterparts. As an expert reviewer notes the patient authored representation of a drug concept, the reviewer copies that patient authored text directly into the interface, and then has the opportunity to attempt to describe the product either through the generic name or the brand name directly. If the generic is entered first, then a corresponding list of appropriate brand names is presented to the reviewer for selection as shown in Figure 2.

After identifying the potential medications that a poster may have mentioned, the curator is tasked with also trying to identify potential medical events related to the patient’s experience with the medication. These can be added dynamically through the “event editor" section of the review screen shown in Figure 3.

3.2 Poster and Patient Demographics

In addition to annotating the post itself, separate information about the poster and the patient (if the posters are referring to someone other than themselves) is annotated as well as show in Figure 4.

Here, the application allows reviewers to incorporate information given about the person including age, gender and location if that is determinable from the content of the post. While some location information may be provided as meta-data from some social media sources, this is not always reliable, and in some cases, may be at too detailed a level to comply with our de-identification process (e.g. some posts contain latitude and longitude coordinates which can pinpoint a post to a street address). For the work done prior

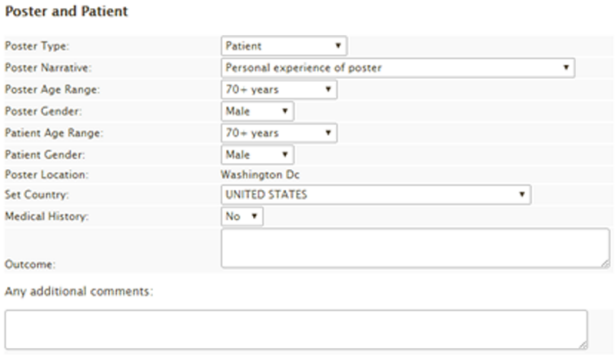

Fig. 4: Annotated poster and patient data

to the WEB-RADR initiative, these details were masked from our reviewers to prevent identifying a particular poster. We have employed methods to generalize coordinates to a zip code or state level where possible.

Gender identity may be inferred from the pronoun usage of a poster, or if they are speaking about a particular patient, from the language used to describe the person of interest (e.g. “my son" or “my daughter"). Additionally, a database of names has been utilized to help identify gender based on the username of the poster [14]. Again, the unique usernames have been de-identified prior to our loading of the posts into *Insight Explorer*, but the attributes are preserved for the reviewer indicating if the poster was male, female or unknown. However, our early studies have shown demographic data is generally difficult to infer. Gender is usually the easiest to identify, but age and other information is not as prevalent from our prior findings [15].

When a curator captures the medication and condition event information, the automatic term lookup pulls concepts for indications from MedDRA[16] version 17.1 (Medical Dictionary for Regulatory Activities). Our local version of MedDRA was extracted using the UMLS[9] MetaThesaurus.

When curators evaluate a social media post, they first review the verbatim text as entered by the poster, and then attempt to map that into a MedDRA term at the LLT and PT levels. *Insight Explorer* helps to ease this task by allowing the users to enter either a LLT or PT, and will then attempt to automatically map to the appropriate MedDRA entry, displaying the PT or LLT mappings automatically on the screen. This helps to insure that the curators are systematically entering data which conforms with the standardized terminology, and can later help in our knowledge discovery process to automate the mappings between patient authored text and a controlled vocabulary.

4. Deployed System

Insight Explorer in support of the WEB-RADR project for social media curation is deployed through Amazon Web Services running on Ubuntu Linux. From there, *Insight*

Explorer is accessed using a web browser after the user authenticates to the system. The application is compatible with most modern web browsers including Internet Explorer, Mozilla Firefox, and Google Chrome.

Adding additional curators are managed through the application by a system administrator. This is accomplished first by adding the users IP address to the Amazon hosted system's white list and then granting corresponding login credentials.

Administrators to the system have an additional security step of connecting through an encrypted SSH key exchange with the server. Passwords can only be incorrectly entered up to three times before accounts are locked which then requires that the account be unlocked by an administrator to enhance our security protocols and prevent tampering with the system.

4.1 System Architecture

The architecture of *Insight Explorer* follows a standard model-view-controller (MVC) framework. The presentation layer uses Apache Velocity templating language to render the page views in HTML and Javascript. The server tier is a Java™web application utilizing a MySQL database instance for persisting the social media reviews by each curator.

4.2 Team Collaboration

The four expert curators were initially divided into two teams of two. When both teams agree on a social media post, the post will be considered as "gold" and marked final in the database for classification as such.

When two teams disagree on their annotation of a particular post, then a super user team determines the final resolution of the disagreement. The decision of the super user team will determine the outcome to be marked finally as "gold." The super user team can be much smaller because the number of posts to be reviewed is estimated to be only 15% of the total. Based on prior research, we expect the teams to agree 85% of the time [18].

A flow diagram demonstrates how the review process occurs and the steps taken to reconcile differences between the review teams by a super user shown in Figure 5. The diagram outlines that we have multiple instances of *Insight Explorer* running, each possessing its own unique database instance for recording the expert's annotations of the social media posts. Each team reviews all of the posts selected for this ADR detection experiment.

To preserve the process, each team will review all of the posts in exactly the same order. Once the user begins the review process, a social media post is selected from the database and presented to the reviewer, at which point he or she will go through the following steps: (1) read the post, (2) identify relevant text, and (3) map those terms for drugs and conditions.

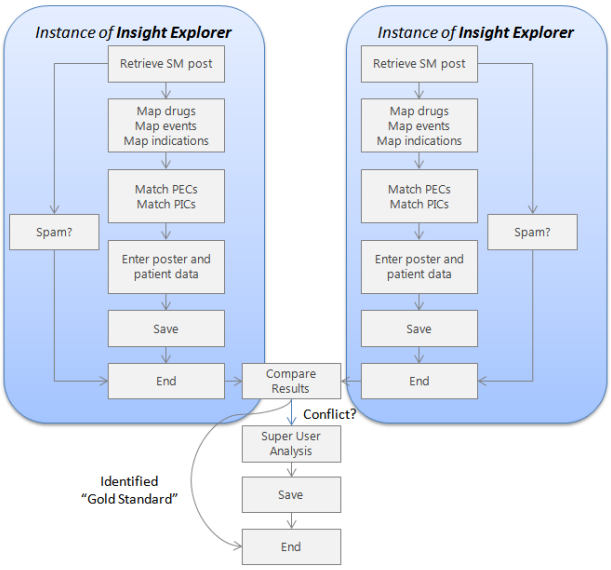

Fig. 5: Process flow supporting parallel teams using multiple instances of our tool

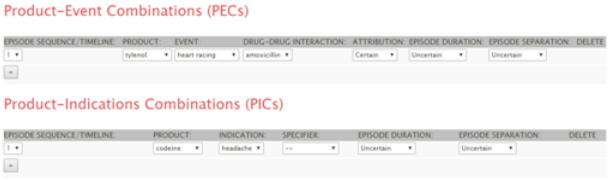

Fig. 6: Event modeling for product indications and events in social media

4.3 Modeling Complex Relations

Once a reviewer has identified the drugs and conditions in a post, then the reviewer next evaluates the post to determine if there are any complex events, such as product-event combinations (PECs) or product-indication combinations (PICs) present. The reviewer can then build those relations based on the drugs and conditions they identified above as shown in Figure 6.

A product event combination occurs when it is discernible from the post that a patient experienced an adverse event directly related to taking a product. A product-indication combination (PIC) is when a post definitively provides context that the patient and/or poster is taking the product to treat a specific medical condition (i.e. the indication) for that drug.

When curators tagged an associated PEC for a post, social media posters claimed about 35% of the time that a medication listed in their post caused an adverse event (such as loss of appetite or insomnia). The other 65% of the time it was uncertain whether the poster claimed a medication referenced in the post caused the adverse event or not. Posters almost never mentioned how long the duration that

the adverse event occurred.

When curators tagged an associated PIC for a post, the posters mentioned about 25% of the time the indication (e.g “psoriasis" or “epilepsy") that they were taking the medication for.

Again, the PEC and PIC annotations will ultimately go through the review process and the outcomes of these complex relations will be resolved by the super user if the curation teams are in disagreement on the models.

5. Discovery

Post curation has been underway for several months, and at the time of this writing, the system contains approximately 60,000 social media posts with the potential for more to be added. The team is currently reviewing social media data for six products of interest to build the gold standard. At this point of the curation process, more than 90% of the content has been marked as spam. Initially, a small team was curating posts only part time, and they averaged just a few thousand posts per month.

In the interest of time, each team was increased to approximately 10 members, with some of them being recruited full time to help complete the project sooner. The system is flexible enough to allow teams to grow and shrink as needed with minimal impact on the curation process while maintaining consistency in the reporting results.

Less than 500 of the 10,000 posts reviewed were not marked as spam. Of those 500 posts, 23 unique events have been recorded (associated with a MedDRA PT) and 132 unique indications have been mapped to MedDRA PT codes from the patient authored text.

The Web-RADR team did not initially want to filter spam programmatically since they wanted to be certain they were reviewing all posts that might have mentioned one of the products of interest in the study. However, due to time constraints, and the large volume of spam found in the data, we have adapted the system to automatically tag spam to speed the curation process. A spam filter was designed in an earlier version of *Insight Explorer* for internal use by the GSK team which was subsequently applied to the WEB-RADR project.

The majority of posters did not specify if they were the patient. However, when they did specify, the poster usually was the patient. To date, the reviewers have found only a few posts that were designated as either friend of the patient, patient health care provider, family, or simply not the patient. Otherwise, the reviewers mark them as an “unknown" poster type.

6. Future Direction

Insight Explorer has proven itself as a solid platform for the annotation and curation of social media data for safety and pharmacovigilance. We hope to expand on the software’s capability to incorporate some of our group’s other efforts around automated trend and topic analysis. Combining these facets of text mining to the annotation process would lead us toward developing a system by which we can eventually understand the strengths and weaknesses inherent to social media data as applied to ADR detection and other safety related activities – to understand truly the voice of the patient in a way that was previously inaccessible to the pharmaceutical industry.

To further this aim, we are exploring the tool’s capability to deal with longitudinal data through extractions of threaded discussions in online patient forums, as well as making the tool more flexible for trained medical experts to define their own protocol of questions that they may wish to ask of the data without having to spend as much time customizing the tool for each particular task.

7. Conclusions

Although WEB-RADR is still early in the process of curating thousands of social media posts, once completed, this will represent one of the largest collections of annotated patient authored texts available for furthering our knowledge and understanding of how the patient’s experiences with medicines are communicated through social media networks.

This will add a valuable resource to measure the performance of machine learning algorithms in their ability to automatically tag and predict adverse drug reactions in social media.

Additionally, the research initiated by the GlaxoSmithKline Safety Listening Laboratory, aims to develop specific classification algorithms around topics of interest beyond ADR detection, and this data set will help to further those goals as well.

8. Acknowledgments

The authors wish to thank the IMI WEB-RADR project leads and Greg Powell (GSK) for arranging the connection between the developers of *Insight Exporer* and the WEB-RADR team. We would also like to acknowledge the work performed by Epidemico for data collection and anonymization which was gathered based on the key word selections of products identified by the WEB-RADR team. We extend thanks as well to Magnus Lerch (Bayer) and Antoni Wisniewski (AstraZeneca) for working to develop the user requirements and giving feedback on test versions of the software.

Also, thanks to Grant Thomson of GSK for providing graphics support. Final thanks to GlaxoSmithKline for generously agreeing to donate the Insight Explorer software to the WEB-RADR project. The GSK Safety Listening Lab is actively pursuing continued research in the field of social media use for pharmacovigilance efforts and is open to collaboration opportunities with other research institutions in the field.

References

[1] R. Ghosh and D. Lewis, "Aims and approaches of web-radr: a consortium ensuring reliable adr reporting via mobile devices and new insights from social media," *Expert opinion on drug safety*, vol. 14, no. 12, pp. 1845–1853, 2015.

[2] C. C. Yang, H. Yang, L. Jiang, and M. Zhang, "Social media mining for drug safety signal detection," in *Proceedings of the 2012 international workshop on smart health and wellbeing*. ACM, 2012, pp. 33–40.

[3] A. Patki, A. Sarker, P. Pimpalkhute, A. Nikfarjam, R. Ginn, K. O'Connor, K. Smith, and G. Gonzalez, "Mining adverse drug reaction signals from social media: going beyond extraction," *Proceedings of BioLinkSig*, vol. 2014, pp. 1–8, 2014.

[4] R. Feldman, O. Netzer, A. Peretz, and B. Rosenfeld, "Utilizing text mining on online medical forums to predict label change due to adverse drug reactions," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2015, pp. 1779–1788.

[5] Powell, Seifert, Reblin, Burstein, Blowers, Menius, and Painter, "Social media listening for routine post-marketing safety surveillance," *Drug Safety*, pp. 1–12, 2016.

[6] FDA, *FDA Adverse Event Reporting System*, 2016 (accessed Feb 8, 2016), http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Surveillance/AdverseDrugEffects/.

[7] S. J. Reisinger, P. B. Ryan, D. J. O'Hara, G. E. Powell, J. L. Painter, E. N. Pattishall, and J. A. Morris, "Development and evaluation of a common data model enabling active drug safety surveillance using disparate healthcare databases," *Journal of the American Medical Informatics Association*, vol. 17, no. 6, pp. 652–662, 2010. [Online]. Available: http://jamia.oxfordjournals.org/content/17/6/652

[8] W. Hersh, M. Weiner, and P. Embi, "Caveats for the use of operational electronic health record data in comparative effectiveness research," *Medical care*, vol. 51, no. 8, pp. S30–S37, 2015.

[9] Aronson and Lang, "An overview of metamap: historical perspective and recent advances," *Journal of the American Medical Informatics Association*, vol. 17, no. 3, pp. 229–236, 2010.

[10] J. L. Painter and N. L. Flowers, "Codeslinger: An interactive biomedical ontology browser," in *Artificial Intelligence in Medicine*. Springer, 2009, pp. 260–264.

[11] G. H. Merrill, P. B. Ryan, and J. L. Painter, "Construction and annotation of a umls/snomed-based drug ontology for observational pharmacovigilance," *Methods*, 2008.

[12] J. L. Painter, K. M. Kleiner, and G. H. Merrill, "Inter-translation of biomedical coding schemes using umls," in *AAAI Fall Symposium*, 2006.

[13] G. H. Merrill, "The meddra paradox," in *AMIA annual symposium proceedings*, vol. 2008. American Medical Informatics Association, 2008, p. 470.

[14] M. Kantrowitz, "Name corpus: List of male, female, and pet names," http://www.cs.cmu.edu/afs/cs/project/ai-repository/ai/areas/nlp/corpora/names/0.html, 1994 (accessed 25-March-2015).

[15] R. L. DiSantostefano, J. L. Painter, M. Thomas, and G. Powell, *Safety Assessment and Selection Bias: Who uses social media to communicate about medications?*, August 2015, iCPE, Boston, MA (Poster session).

[16] I. F. of Pharmaceutical Manufacturers, "Meddra (medical dictionary for regulatory activities)," http://www.meddra.org, 2015 (accessed June 1, 2015).

[17] Brown, E. G., L. Wood, and S. Wood, "The medical dictionary for regulatory activities (meddra)," *Drug Safety*, vol. 20, no. 2, pp. 109–117, 1999.

[18] D. L. MacLean and J. Heer, "Identifying medical terms in patient-authored text: a crowdsourcing-based approach," *Journal of the American Medical Informatics Association*, vol. 20, no. 6, pp. 1120–1127, 2013.