BOSTON UNIVERSITY

COLLEGE OF ENGINEERING

Dissertation

DIGITAL PHARMACOVIGILANCE: THE MEDWATCHER SYSTEM FOR MONITORING ADVERSE EVENTS THROUGH AUTOMATED PROCESSING OF INTERNET SOCIAL MEDIA AND CROWDSOURCING

by

CLARK C. FREIFELD

B.S., Yale University, 2000
M.S., Massachusetts Institute of Technology, 2010

Submitted in partial fulfillment of the

requirements for the degree of

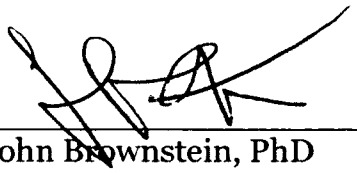
Doctor of Philosophy

2014

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

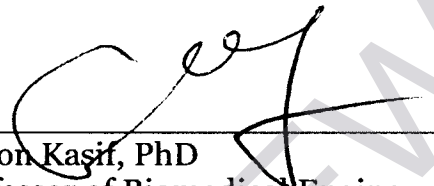
Approved by

First Reader



John Brownstein, PhD
Associate Professor of Pediatrics
Harvard University, School of Medicine

Second Reader

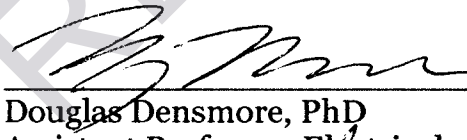


Simon Kasif, PhD
Professor of Biomedical Engineering

Third Reader

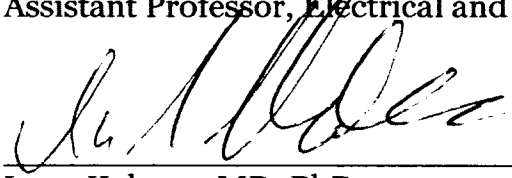
Charles DeLisi, PhD
Metcalf Professor of Science and Engineering

Fourth Reader



Douglas Densmore, PhD
Assistant Professor, Electrical and Computer Engineering

Fifth Reader



Isaac Kohane, MD, PhD
Lawrence J. Henderson Professor of Pediatrics and Health Sciences and Technology
Harvard University, School of Medicine

Acknowledgements

Completing graduate studies is difficult at any phase of life. For me, juggling the obligations of full-time employment and two small children in addition to my academic work has been challenging to say the least. However, throughout the past six years, I have been extremely fortunate to be surrounded by a thoughtful, enthusiastic, patient, and skilled support network. That I stand here today at the end of the journey is credit to all those who have supported me along the way as much as it is to me.

I first thank Professor John Brownstein, my primary research advisor, partner, friend, and mentor. My constant coach, captain, and cheerleader. Dr. Nabarun Dasgupta, the unofficial 6th committee member, has been invaluable as both a friend and pharmacovigilance domain expert. Prof Simon Kasif, my research co-advisor, has continually guided me through the maze of doctoral education, and stepped up to advocate on my behalf on multiple occasions. I thank Prof Charles DeLisi and Prof Douglas Densmore for participating in my committee though the subject matter of the thesis falls outside of the traditional biomedical engineering topic areas; I also thank Prof Zak Kohane, committee member, who has supported me in my career at Boston Children's, and my graduate studies, from the very beginning, some nine years ago. My analyst-curator team has spent countless hours both labeling documents and discussing how best to label, engaging with users and patients, and making MedWatcher better all the time: Chi Bahk, Carrie Pierce, Christopher Menone, Wenjie Bao,

Michael Gilbert—your annotations alone form the lifeblood of the thesis. Thanks. The technical crew has likewise put in many thoughtful hours to help make the software and user experience a reality, on both HealthMap and MedWatcher: Sue Aman, Kate O'Brien, Harold Rodriguez, Rachel Chorney, James Power, Aaron Rietschlin. On the HealthMap side, I also thank Sumiko Mekaru for her epidemiology expertise and strong vision. For all the logistical support, critical but often unsung, thanks go to Robin Heffernan and Leila Amerling, who can make things happen in a pinch. I further thank our collaborators at FDA, visionaries and advocates who helped make the project possible, Ross Filice and Doug Wood. At BU, I extend my gratitude to Irving Bigio for his help securing tuition support, Cathie Klapperich for believing that public health and biomedical engineering go hand-in-hand, Dean Sol Eisenberg for believing in a student outside the mold, and Dafna Shulman-Petrover and Christen Bailey for ever-cheerful navigation of forms and requirements. Thanks to my parents, Charles Freifeld and Marilyn Smith, who have offered moral support, and clutch babysitting, so many times. Thank you Maya and Jonah, for your patience and spirit. Finally, and most of all, thank you to my wife Andee Krasner, who has so thoughtfully and patiently helped me stay on this course, through all the many times I thought about quitting. As much as anyone, this has been Andee's journey as well as my journey and I have been fortunate to have such a companion.

DIGITAL PHARMACOVIGILANCE:

THE MEDWATCHER SYSTEM FOR MONITORING ADVERSE EVENTS THROUGH AUTOMATED PROCESSING OF INTERNET SOCIAL MEDIA AND CROWDSOURCING

CLARK C. FREIFELD

Boston University, College of Engineering, 2014

Major Professor: John S. Brownstein, PhD, Associate Professor of Pediatrics, Harvard University, School of Medicine

ABSTRACT

Half of Americans take a prescription drug, medical devices are in broad use, and population coverage for many vaccines is over 90%. Nearly all medical products carry risk of adverse events (AEs), sometimes severe. However, pre-approval trials use small populations and exclude participants by specific criteria, making them insufficient to determine the risks of a product as used in the population. Existing post-marketing reporting systems are critical, but suffer from underreporting. Meanwhile, recent years have seen an explosion in adoption of Internet services and smartphones. MedWatcher is a new system that harnesses emerging technologies for pharmacovigilance in the general population. MedWatcher consists of two components, a text-processing module, MedWatcher Social, and a crowdsourcing module, MedWatcher Personal. With the natural language processing component, we acquire public data from the Internet, apply classification algorithms, and extract AE signals. With the

crowdsourcing application, we provide software allowing consumers to submit AE reports directly.

Our MedWatcher Social algorithm for identifying symptoms performs with 77% precision and 88% recall on a sample of Twitter posts. Our machine learning algorithm for identifying AE-related posts performs with 68% precision and 89% recall on a labeled Twitter corpus. For zolpidem tartrate, certolizumab pegol, and dimethyl fumarate, we compared AE profiles from Twitter with reports from the FDA spontaneous reporting system. We find some concordance (Spearman's rho = 0.85, 0.77, 0.82, respectively, for symptoms at MedDRA System Organ Class level). Where the sources differ, milder effects are overrepresented in Twitter. We also compared post-marketing profiles with trial results and found little concordance.

MedWatcher Personal saw substantial user adoption, receiving 550 AE reports in a one-year period, including over 400 for one device, Essure. We categorized 400 Essure reports by symptom, compared them to 129 reports from the FDA spontaneous reporting system, and found high concordance (rho = 0.65) using MedDRA Preferred Term granularity. We also compared Essure Twitter posts with MedWatcher and FDA reports, and found rho = 0.25 and 0.31 respectively.

MedWatcher represents a novel pharmacoepidemiology surveillance informatics system; our analysis is the first to compare AEs across social media, direct reporting, FDA spontaneous reports, and pre-approval trials.

Table of Contents

Abstract vi

List of Tables x

List of Figures xi

List of Abbreviations xvi

Introduction 1

Background: Pharmacoepidemiology Foundations 7

HealthMap: Foundations for MedWatcher 10
- **HealthMap Text Processing** 12
- **HealthMap and StreetRx Crowdsourcing** 13
- **HealthMap Surveillance Key Results** 16

MedWatcher Social 19
- **System Architecture** 21
- **Acquisition** 23
- **Supervised Labeling** 25
- **Product and Symptom Taxonomies** 28
- **Information Extraction Algorithm** 30
- **Document Classification Algorithm** 33
  - Feature Extraction 33
  - Fisher-Robinson classifier 36
  - Naïve Bayes Classifer 37
  - Active Learning Approach 38
- **De-Duplication** 39
- **MedWatcher Social: Classifier Evaluation Methods** 40
  - Information Extraction Evaluation 40
  - Document Classification Evaluation by Cross Validation 41
- **MedWatcher Social: Classifier Evaluation Results** 43
  - Classifier performance: information extraction 43
  - Classifier performance: preliminary inter-rater agreement study 45
  - Classifier performance: document classification 45
- **MedWatcher Social: Signal Analysis Methods** 49
  - Product Selection 49
  - Known Product Label Information 51
  - FDA Spontaneous Adverse Event Reporting Data 52
  - Matching Data Sets for Comparison 52
- **MedWatcher Social: Signal Analysis Results** 55
  - Characterizing Twitter and Facebook 55
  - Comparison 1: zolpidem tartrate (Ambien) 57
  - Comparison 2: certolizumab pegol (Cimzia) 63
  - Comparison 3: dimethyl fumarate (Tecfidera) 67
- **MedWatcher Social: Discussion, Limitations** 71
- **Future Work** 72

MedWatcher Personal 74

System Architecture 75
App Frontend Features 75
Report triage, de-identification and submission protocol 77
Promotion: Twitter interaction, advertising, Oprah Magazine 78
MedWatcher Personal: Analysis Methods 80
MedWatcher Personal: Results 81
User Statistics, Report Volume 81
Key Sample Reports 82
Essure Case Study: Background 85
Characterization of Essure MedWatcher Personal and Social Reports 88
Characterization of Essure FDA MAUDE Reports 93
Comparison of Essure Reports Across the product label, MAUDE, MedWatcher Personal, and MedWatcher Social 96
MedWatcher Personal: Discussion, Limitations 99
Future Work 100

Conclusion, Contributions of the Thesis 101

Disclosure 103

Appendices 104
Explanation of Fisher-Robinson and Naïve Bayes classifier methods 104
Symptom cross-mappings for each product 109

Bibliography 113

Author's Vita 125
Education 125
Selected Professional Experience 125
Awards, Honors and Scholarships 126
Articles Cited in the Thesis 127
Articles Not Cited in the Thesis 128

List of Tables

Table 1. Performance summary for the information extraction algorithm 44

Table 2. Full list of new molecular entities (NMEs) as designated by the FDA.... 50

Table 3. Concordance measures between MedWatcher Social Twitter data and FDA FAERS reports for zolpidem tartrate, at varying symptom granularity 57

Table 4. Zolpidem symptoms with greater than 2 percentage point higher prevalence in treatment versus placebo, as reported from clinical trials..... 61

Table 5. Concordance measures between MedWatcher Social Twitter data and FDA FAERS reports for certolizumab pegol, at varying symptom granularity 64

Table 6. Concordance measures between MedWatcher Social Twitter data and FDA FAERS reports for dimethyl fumarate, at varying symptom granularity. 68

Table 7. Summary of reports received through the MedWatcher Personal application..... 82

Table 8. Top symptoms reported for Essure from the 400 MedWatcher direct reports. 89

Table 9. Top symptoms reported in Essure AE tweets 92

Table 10. Top symptoms reported for Essure from the 129 MAUDE reports from 2012 and 2013 93

List of Figures

Figure 1. The FDA "consumer-friendly" voluntary MedWatcher reporting form (3500B) 3

Figure 2. The HealthMap main page (healthmap.org), offering an interactive map visualization of real-time disease outbreak alerts from Internet media..... 11

Figure 3. The HealthMap Outbreaks Near Me mobile application, allowing ready access to location-specific disease outbreak information and direct reporting capability for all users 14

Figure 4. StreetRx (streetrx.com) crowdsourcing community for black market prices for prescription drugs. 15

Figure 5. Price per milligram cross-comparison of StreetRx price reports with law enforcement reports and the Silk Road online marketplace 16

Figure 6. Comparison of formal reporting on avian influenza with reports collected through HealthMap 17

Figure 7. Trends in the time from estimated outbreak onset to (A) outbreak discovery and (B) public communication for selected WHO-verified outbreaks, 1996–2009 [60] 18

Figure 8. Adjusted volume of H1N1 crowdsourced reports from HealthMap Outbreaks Near Me users as compared to weighted influenza-like-illness visit data from the CDC, from September 5, 2009 to January 30, 2010 [48, 49] 19

Figure 9. MedWatcher Social system architecture: the acquisition engine loads posts into the database; we then build tools and data outputs around the database 22

Figure 10. The cycle of data acquisition, classification, training, and evaluation in MedWatcher Social 23

Figure 11. Sample Twitter posts illustrating how the system identifies key terms, and some of the challenges of classification 27

Figure 12. The MedWatcher Social curation dashboard allows analysts to search, review, and label individual posts 27

Figure 13. The HealthMap tree structure for matching dictionary terms against an input text, here applied to symptom matching for MedWatcher Social 31

Figure 14. Semantic highlighting based on the information extraction and alignment process. 32

Figure 15. Summary of computation for Fisher-Robinson document classification algorithm. s and x are hyperparameters of the model; C^{-1} is the inverse X^2 function 37

Figure 16. Computation used in the Naïve Bayes classifier. A represents the "positive" (adverse event post) outcome; D represents the given document; w_i represents the ith token in the document 38

Figure 17. Confusion matrix for inter-annotator agreement study 45

Figure 18. Precision-recall curve for the Fisher-Robinson document classifier on Twitter posts 46

Figure 19. Precision-recall curve for the Fisher-Robinson document classifier, on Facebook posts 47
Figure 20. Precision-recall curve for the Naïve Bayes classifier, on Twitter posts 48
Figure 21. Comparison by magnitude of FAERS spontaneous reports (left) with Twitter AE posts (right) for top reported symptoms (MedDRA Preferred Terms) associated with zolpidem 58
Figure 22. Top reported zolpidem symptoms in FAERS (left) and Twitter (right) at the MedDRA High Level Term level 58
Figure 23. Zolpidem symptom comparison by SOC 59
Figure 24. Sampling of zolpidem AE tweets reporting driving under its influence, including one road traffic accident (highlighted in red) 60
Figure 25. Sample AE tweets and Facebook posts for zolpidem. The tweet in the lower right indicates a case where the abnormal dreams impact the user's experience beyond a minor annoyance 62
Figure 26. Comparison of FAERS spontaneous reports (left) with Twitter AE posts (right) for top reported symptoms (MedDRA Preferred Terms) associated with certolizumab 64
Figure 27. Top reported certolizumab symptoms in FAERS (left) and Twitter (right) at the MedDRA High Level Term level 65
Figure 28. Certolizumab symptom comparison by SOC 65

Figure 29. Selected AE posts from Twitter (top) and Facebook (bottom) for certolizumab 67
Figure 30. Comparison of FAERS spontaneous reports (left) with Twitter AE posts (right) for top reported symptoms (MedDRA Preferred Terms) associated with dimethyl fumarate 68
Figure 31. Top reported dimethyl fumarate symptoms in FAERS (left) and Twitter (right) at the MedDRA High Level Term level 69
Figure 32. Dimethyl fumarate symptom comparison by SOC 69
Figure 33. Comparison of quantitative pre-approval clinical trial data (left) with symptoms identified in Twitter (right) 70
Figure 34. Sample AE posts for dimethyl fumarate from Twitter (top) and Facebook (bottom) 71
Figure 35. MedWatcher Personal application software architecture 75
Figure 36. MedWatcher Personal mobile app screens 76
Figure 37. MedWatcher Personal administrator tools for report review and triage 77
Figure 38. We use MedWatcher Social to identify posts of interest, then solicit reports via the @med_watcher Twitter account 79
Figure 39. MedWatcher Personal user survey 80
Figure 40. The Essure coil and its deployment and functioning 86

Figure 41. Map representing the location information provided in 85 of the 400 Essure reports, showing the national scope of the MedWatcher user community 91
Figure 42. Typical tweets regarding adverse events attributed to Essure 92

List of Abbreviations

AE ..... adverse event

API ..... Application Programming Interface

CDC ..... Centers for Disease Control and Prevention

COD ..... cause of death

DHS ..... Department of Homeland Security

EMA ..... European Medicines Association

FAERS ..... FDA Adverse Event Reporting System

FDA ..... Food and Drug Administration

HHS ..... Health and Human Services

HLGT ..... High-Level Group Term

HLT ..... High-Level Term

HSG ..... hysterosalpingram

IMI ..... Innovative Medicines Initiative

LLT ..... Lower-Level Term

MAUDE ..... Manufacturer and User Facility Device Experience

MedDRA ..... Medical Dictionary for Regulatory Activities

MRI ..... Magnetic Resonance Imaging

MS ..... multiple sclerosis

NIH ..... National Institutes of Health

NME ..... new molecular entity

PDF ..... Portable Document Format

PET ..... polyethylene terphthalate

PT..... Preferred Term

ROC ..... Receiver Operating Characteristic

SAMHSA .....Substance Abuse and Mental Health Services Administration

SNRI ..... serotonin and norepinephrine reuptake inhibitor

SOC..... System Organ Class

SSRI ..... selective serotonin reuptake inhibitor

STD..... sexually transmitted disease

UK..... United Kingdom

URL .....Uniform Resource Locator

US ..... United States

USAID ..... United States Agency for International Development

VAERS ..... Vaccine Adverse Event Reporting System

WHO.....World Health Organization

XML..... Extensible Markup Language

Introduction

Half of Americans take a prescription drug [1]; many others make use of regulated medical devices both at home and in the clinical setting, and vaccination coverage for many vaccines is over 90% [2]. Together, drugs, devices and vaccines have extended our lives and transformed our health. At the same time, nearly all medical products carry a risk of adverse events—harmful impacts caused by the treatment itself. Though in many cases these effects are mild or extremely rare, for some products in some contexts, adverse events can be severe.

The process of clinical trials that precedes government approval is critical for evaluating both safety and efficacy of new medical products. However, clinical trials often exclude participants with multiple health conditions, who are taking other medications, or who don't fit specific clinical criteria [3, 4]. Further, due to inherently smaller sample sizes, clinical trials typically do not detect rare or extremely rare effects [5]. Meanwhile, once a product is approved, it is often prescribed to a much larger population of patients, many of whom don't fit the trial criteria. Further, approximately 20% of prescriptions are for "off label" usage, situations in which clinical trial outcomes have limited applicability [6, 7, 8]. The result is that clinical trials are often insufficient to determine the true risks of a product as it is used in the population. In fact, over half of approved drugs have serious side effects not detected before approval [9]. Moreover, with increased use of opioids and certain new biologics, the incidence of serious adverse events in the US has increased in recent years, even after accounting for

the increase in prescriptions [10].

Agencies such as the Food and Drug Administration (FDA) and Centers for Disease Control and Prevention (CDC) are charged with protecting population health by carefully evaluating and managing the risks and benefits of each product on the market. Because of the clear need to monitor risks beyond clinical trials, the government has established a range of reporting mechanisms for industry, consumers, and clinicians, most notably FDA MedWatch and CDC Vaccine Adverse Event Reporting System (VAERS). For drugs, biologics, and devices FDA MedWatch [11, 12, 13], in particular, the 3500 Form [14], has long been the established channel for clinicians and pharmaceutical companies to report on post-marketing adverse events. For vaccines, VAERS [15, 16] is the preferred reporting mechanism. However, these systems are often cumbersome and oriented to expert clinicians (Figure 1). Even for clinicians, who are under professional obligation report, these forms represent another paperwork burden in an already overburdened workflow. Given these barriers for both patients and clinicians, the result is that adverse events often go unreported [17, 18]. A 2012 study of hospitals by the Department of Health and Human Services Office of Inspector General found that 86% of reportable adverse events went unreported [19]. While not all of these adverse events were related to medical products (some resulted from improper procedure, highly complex cases, or other factors), clearly we face significant underreporting across our health care infrastructure. Of the reports that are collected through existing channels, 80% of drug-related

reports [20], 35% of vaccine reports [21], and 98% of device reports [22] come from manufacturers, who are under legal obligation to report all events they are aware of. While manufacturer reporting is essential, without the direct engagement of patients and clinicians, we are missing a key piece of the picture.

Figure 1. The FDA "consumer-friendly" voluntary MedWatcher reporting form (3500B). This simplified version was introduced in 2013.

Another limitation of existing reporting systems such as FDA MedWatch and VAERS is that they are often slow to publish the data they collect. FDA Adverse Event Reporting System (FAERS), the mechanism by which the FDA publishes MedWatch reports, typically releases data on a quarterly basis, with a one-year delay. At the time of writing in February 2014, the data had not been

updated in over one year [23]. The data that are released come in large relational database files, with duplication and inconsistencies. The FDA's system for disseminating medical device reports, Manufacturer and User Facility Device Experience [24] (MAUDE), is better in that it is typically updated on a monthly basis, and searchable via a Web-based form, but is still challenging for non-expert users. Although public health officials and regulators in government may have ready access to these data sets, the result of the delay is that manufacturers, clinicians, and the general public have limited opportunity to gain insights from these valuable reports.

At the same time as we see both the impact of adverse events and major reporting gaps, we have an explosion in adoption of Internet services and mobile computing devices. In particular, adoption of social media tools and smartphones has surged: 160 million Americans are on Facebook [25], 140 million US-based accounts are on Twitter [26], and 46% of American adults own smartphones [27]. In the health domain specifically, as of fall 2012, 59% of US adults had looked online for health information in the past year, 35% had gone online specifically to determine a medical condition [28], while 52% of smartphone owners reported gathering health information through their phones [29]. Meanwhile, on the clinical side, fully 85% of US physicians own or use smartphones professionally [30].

In response to the challenges of adverse event reporting and the opportunities offered by emerging technologies, we present MedWatcher, a new

system for early detection and tracking of adverse events in the population at large. This new program builds on our earlier work in infectious disease outbreak tracking and in the broader field of digital disease detection. MedWatcher consists of two main components, a data-mining module, **MedWatcher Social**, and a crowdsourcing module, **MedWatcher Personal**. With the data-mining component, we acquire a broad range of data from the Internet, in the form of posts in public forums. We then apply document classification and information extraction algorithms to filter and classify the information and extract both individual and aggregate signals of adverse events. With the crowdsourcing application, we provide a mobile and Web-based application that allows users to both track safety information on their chosen products, and easily submit adverse event reports directly into the system. We then review and filter these reports and similarly look for both individual and aggregate signals of both known and previously unknown side effects.

The job of public health officials, to keep beneficial, sometimes life-changing, and sometimes life-saving therapies available while protecting us from their side effects, remains difficult. Given the noisy data available, primarily in the form of spontaneous MedWatch reports, when an apparent signal arises, there is often no clear answer on whether to change recommendations, labels, or pull a product from the market. With the work of this thesis, we can't hope to answer these thorny questions definitively. However, for the regulators, clinicians, and patients faced with these challenges, we offer a new arrow in the

quiver. We are seeing a situation of vast underreporting of adverse events, even as more and more people are equipped with easy-to-use, high-bandwidth reporting tools. With MedWatcher, we apply the power of new information technologies to provide a new window into understanding the adverse effects of medical products as they occur in the general population.

In the work of this thesis, we seek to answer the following questions: first, can we engineer a system, MedWatcher Social, with sufficient automated classification performance to derive useful data from the noise of Internet media? Second, can we engineer a system, MedWatcher Personal, with sufficient consumer adoption to derive useful data by crowdsourcing direct reports? Third, if so, how do these new sources compare with available pre-marketing and post-marketing surveillance sources? If the new sources match closely, then we have validation of the new signals, and the advantage of our new approach comes from speed and ease of access to the data. If they don't match, then two possibilities arise: one is that our new data sources are simply too noisy to be useful; the other is that our new sources offer important insights that are not available through existing channels.

As shown in our results below, we find that we are indeed able to build effective systems for generating signals, and we find that they match to some degree with existing sources. In general, when we compare social media data with traditional MedWatch reports, we find legitimate adverse event signals in social media, but they tend to be weighted toward the more common, but less severe