Opportunities and Challenges Using Unsolicited, Self-Reported, Health Outcomes Data from Open Web-Based Sources

*Mark Wolff, Ph.D.*  
Advisory Industry Consultant  
Chief Health Analytics Strategist  
Health & Life Sciences Global Practice  
SAS Institute  
*mark.wolff@sas.com*
Healthcare Analytics

“While the individual man is an insoluble puzzle, in the aggregate he becomes a mathematical certainty”

Sherlock Holmes
Unstructured Data

Text Analytics

• Tremendous interest in accessing and using unstructured and semi-structured data to **enhance/inform** data driven decision making related to personal and population health.

• Interest not limited to internal/formal data sources
Social Media, Health and Safety
Unsolicited, Self-Reported, Health Outcomes Data

Great interest in collecting and understanding health related information from a variety of Open and Closed Web data sources
FDA’s March 2001 draft guidance for industry entitled Post marketing Safety Reporting for Human Drug and Biological Products Including Vaccines

- Adverse experience information that is submitted via the Internet to an entity with post marketing reporting obligations under 310.305, 314.80, and 600.80 should be reported to FDA if there is knowledge of the four basic elements for submission of an individual case safety report, namely:
  - An identifiable patient
  - An identifiable reporter
  - A suspect drug or biological product
  - An adverse experience or fatal outcome suspected to be due to the suspect drug or biological product

- Draft guidance states that those entities should review any Internet sites sponsored by them for adverse experience information, but are not responsible for reviewing any Internet sites that they do not sponsor; however, if they become aware of an adverse experience on an Internet site that they do not sponsor, they should review the adverse experience and determine if it should be reported to FDA.
Information on Adverse Reactions from the Internet:

- The Marketing Authorization Holder should regularly screen websites under their management or responsibility, for potential reports on adverse reactions.
- The Marketing Authorization Holder is not expected to screen external websites for information on adverse reactions.
- However, if a Marketing Authorization Holder becomes aware of an adverse reaction on any other website the Marketing Authorization Holder should review the case and determine whether it should be reported in expedited manner in accordance with Chapter I.4, Sections 3.1 and 3.5.

The Marketing Authorization Holder should consider utilizing their websites to facilitate adverse reaction collection, e.g. by providing adverse reaction forms for reporting or by providing appropriate contact details for direct communication. In relation to such reported adverse reactions, identifiability of the reporter and Patient refers to the existence of actual people.
Up to 90% of side effects to drugs are not reported, according to some estimates. “Adverse drug reactions (ADRs) are grossly under reported by everyone, including healthcare professionals, but particularly so by patients,”

David Lewis, Novartis

Data from the European Medicines Agency (EMA) indicate that patients are not reporting side effects adequately through official channels and several regulators are exploring the possibility of making use of the abundant information swirling around social networks.

Searching social networks to detect adverse reactions

The Pharmaceutical Journal
JAN 2015
Elizabeth Sukkar
WEB-RADAR

WEB-RADR is investigating the potential for publicly available social media data for identifying drug safety issues

- Establishing the regulatory framework for social media mining for ADRs
- Increasing public access to National Competent Authorities (regulators)
- Engaging patients in the awareness and reporting of safety concerns
- Improving methods of mining social media for pharmacovigilance
- Delivering a platform for monitoring effects of risk minimization
Utilizing Social Media Data for Pharmacovigilance: A Review

Journal of biomedical informatics : February 2015
Sarker, A., Nikfarjam, A., O'Connor, K., Ginn, R., Gonzalez, G., Upadhaya, T., Jayaraman, S., Smith, K.
"Social Media Listening"
Industry Adoption

GSK, Merck use social media to learn how patients use drugs outside the lab

By Jonah Comstock | October 05, 2015

We've written before about Epidemico, a startup that grew out of the Boston Children's Hospital’s work on combing social media for medical and health data. Today at the Health 2.0 conference in Santa Clara, California, Greg Powell, director of pharmacovigilance at GlaxoSmithKline, talked about how the pharma company uses the data it collects from its partnership with Epidemico.

"People really are communicating a lot online on social media," Powell said. "There's actually a wealth of information here that potentially hasn't been tapped into until recently. The question is 'if people are talking about our products, should we be listening to what they're saying?' The answer is 'Of course.'

GlaxoSmithKline collected data about public postings on Twitter and Facebook that mentioned any of a list of 1,000 of the companies drugs. Altogether, they found more than 6 million hits on Twitter and more than 15 million hits on Facebook.
The objective of this requirement is to provide FDA with the resources needed to use social media to inform and evaluate FDA risk communications. Specifically, the objective is to provide FDA with:

- Analyses of social media that provide baselines on consumer sentiment prior to FDA communication and that depict changes in social media buzz following FDA communications
- In-house capability for social media monitoring; and Surveillance through social media listening for early detection of adverse events and food-borne illness.
- The scope of work includes social media buzz reports, a social media dashboard, and quarterly surveillance reports related to specific product classes.

The purpose of this workshop is to discuss ways to incorporate patient preferences on the benefit-risk trade-offs of medical devices into the full spectrum of the Center for Devices and Radiological Heath (CDRH) regulatory decision making. It also aims to advance the science of measuring treatment preferences of patients, caregivers, and health care providers.

Incorporating patient-preference evidence into regulatory decision making

Martin P. Ho, Juan Marcos Gonzalez, Herbert P. Lerner, Carolyn Y. Neuland, Joyce M. Whang, Michelle McMurry-Heath, A. Brett Hauber, Telba Irony
A Patient Reported Outcome is any report of the status of a patient’s health condition that comes directly from the patient, without interpretation of the patient’s response by a clinician or anyone else.

Guidance for Industry
Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims
U.S. Department of Health and Human Services
Food and Drug Administration 2009

Any outcome evaluated directly by the patient himself and based on patient’s perception of a disease and its treatment(s) is called patient-reported outcome (PRO).

Reflection Paper on the Regulatory Guidance for the use of Health Related Quality of Life (HRQL) Measures in the Evaluation of Medicinal Products
European Medicines Agency 2005
Google Trends

...this is not the way
MARCH 4, 2013

Twitter Reaction to Events Often at Odds with Overall Public Opinion

By Amy Mitchell and Paul Hitlin
"The resulting dataset contained a high volume of irrelevant information, but provided a useful starting point."

"We did not seek to verify each individual report as truthful, but rather to identify overall associations between Twitter and official spontaneous report data as a preliminary proof of concept."
Methods and Application for Determining the Integrity and Veracity of Medical Device Safety Related Data in Social Media

Mark Wolff, SAS Institute Inc., Cary, NC, USA
Michael Wallis, SAS Institute Inc., Cary, NC, USA

ABSTRACT

As more individuals, organizations and institutions rely on the internet for information to support decision making, the integrity and veracity of those data have become a critical issue. A key area of interest is the applicability and utility of social media data for device safety monitoring. Such data offer a potentially valuable resource for post-marketing device safety surveillance for the industry and regulators. Adoption of these data as a resource has been hampered by concerns related to the accuracy and reliability of these data and a lack of guidance from regulators. Applying the capabilities of SAS Text Analytics we propose a method for qualifying the veracity of unstructured data collected from internet sources. Further, we describe its application in post marketing medical device safety monitoring and signal detection.
I have been dealing with epilepsy for years now, many of which have included uncontrollable seizures and have been accompanied by symptoms including impaired vision, dizziness, and difficulty breathing. Since I have started using the vagal nerve stimulator, I have found these symptoms to have subsided. Though I have since experienced minor palpitations, such occurrences have been rare. I believe the worst is over. Based upon my experiences with the VNS device, I would recommend it to others as a possible solution.
I have been dealing with epilepsy for years now, many of which have included uncontrollable seizures and have been accompanied by symptoms including impaired vision, dizziness, and difficulty breathing. Since I have started using the vagal nerve stimulator, I have found these symptoms to have subsided. Though I have since experienced minor palpitations, such occurrences have been rare. I believe the worst is over. Based upon my experiences with the VNS device, I would recommend it to others as a possible solution.
I have been dealing with epilepsy for years now, many of which have included uncontrollable seizures and have been accompanied by symptoms including impaired vision, dizziness, and difficulty breathing. Since I have started using the vagal nerve stimulator, I have found these symptoms to have subsided. Though I have since experienced minor palpitations, such occurrences have been rare. I believe the worst is over. Based upon my experiences with the VNS device, I would recommend it to others as a possible solution.
How to build a Lie Detector for the Internet
In Three Challenging Steps

• Semantic Field Normalization/Contextualization for Self-Reported Symptom-Treatment-Outcome Measurement in Web-based Media Sources

• Adaptation of “Semantic Nets” to Establish Veracity of Symptom-Treatment Outcome Reports in Health Related Web Interactions

• Behavioral context as a pathway to crafting semantic field normalization mappings in Clinician/Patient Reported Outcomes Data (C/PROM)
Segmentation
Association Analysis
Multidimensional Temporal Analysis
Visualization and Reporting
Patient preferences considered for the first time in FDA decision to approve first-of-kind obesity device

RTI Health Solutions partnered with the FDA to conduct a study on patients’ preferences which contributed to the Agency’s regulatory decision to approve a first-of-kind device to treat obesity.

This was the first time a patient preference study impacted a new device approval.

Incorporating patient-preference evidence into regulatory decision making

Surgical Endoscopy
January 2015
Martin P. Ho, Juan Marcos Gonzalez, Herbert P. Lerner, Carolyn Y. Neuland, Joyce M. Whang, Michelle McMurry-Heath, A. Brett Hauber, Telba Irony
FDA
Patient Preference

**Obesity**
*Gastric Balloon, Band, Sleeve, Surgery*
- Efficacy
  - Duration
  - Quality of Life
  - Weight Loss
- Safety
  - Adverse Events
  - Device Malfunction
  - Hospitalization
- Usage
  - Daily Life Impact

**Epilepsy**
*RNS (Neuropace), VNS, DBS, AED*
- Efficacy
  - Duration
  - Quality of Life
  - Seizure Reduction
- Safety
  - Adverse Events
  - Hospitalization
- Usage
  - Daily Life Impact
# Harness Patient Preference from Social Media

"Identify and Incorporate the Patient Voice into Our Decision-making on Medical Devices." – FDA Voice

Emily McRae¹, Cheyanne Baird¹, Joe Boland³, Pat Dougherty¹, Martin Ho¹, Telba Irony¹, Mimi Nguyen², Kathryn O’Callaghan², Michael Wallis¹, Mark Wolff¹, Anindita Saha²

¹SAS Institute Inc., Cary, NC, ²Center for Devices and Radiological Health, Food and Drug Administration, Silver Spring, MD

## Objectives
- Explore the feasibility of collecting patient preference information from a variety of social media sources on selected topics.
- Apply sentiment scoring method to reveal context-specific sentiment levels related to medical device treatments.

## Background
- Social media has become a popular medium for individuals to express their opinions.
- After conducting a patient preferences survey on weight loss devices, CDHR explored using sentiment analysis to harness patient preferences from unstructured posts of social media for comparison with the survey results.
- Sentiment analysis is an evolving technology that applies text analytics to analyze a document and infer the author’s sentiment about a topic of interest, such as a medical treatment.
- CDHR and SAS collaborated to capture web-based patient sentiments on the benefits, risks, and other attributes of medical treatment to treat obesity and epilepsy.

## Materials and Methods
- Identified most popular websites on treatments of obesity (surgery, sleeve, band, balloon) and epilepsy (AEDs, DBS, AEDs).
- Veracity Scoring (Signal to noise reduction)
- Segmentation and Data Cleansing
- Sentiment Analysis
- Visualization and Exploration
- Incremental data crawling for real-time sentiment analysis compared to baseline

## Patient Forums

### Epilepsy
- http://www.copingwith-epilepsy.com/
- http://epilepsyfoundation.ring.com/
- http://forum.epilepsysocty.org/

### Obesity
- http://www.bariatriccali.com/
- http://www.obesityhelp.com/
- http://weightlosssurgery.proboards.com/
- http://www.wlsurgery.com/

## Results

### Obesity Domain Dashboard

### Epilepsy Domain Dashboard

## Conclusions

Developed upon advanced text analytics, sentiment analysis is a powerful method to harness timely patient preference information from unstructured yet increasingly big data in the social media to complement data collected from other sources.

---

### An Example of Sentiment Analysis

**Obesity**: One person mentioned that they had surgery on 10/27. My surgery went well, I did take longer in recovery than most but I don’t really remember that to much. I was on my pain drip and I used it every time the light turned green. I didn’t feel pain but they had said to try stay ahead of the pain so that is what I did. I was up and walking around 4am on the day I had surgery. My surgery was at 2pm. My hubby stayed the whole time I was in the hospital and that was a huge help. I just felt more comfortable with him there. I went home Tuesday around suppertime and didn’t need any pain medicine at all once I got home. And by that first weekend I was getting in my protein and my liquids. I was surprised by that. My incisions were all good and are pretty much healed up now. I walked everyday and most days I did get my hour in. It did make me feel better I am sure of that. I have been back to the gym and working infielding my stamina back. I had a good 2 week post op appointment and was happy about that. Everything I introduced my sleeve in the soft food stage has gone well. So far that I am grateful. So I am following my doctors orders and doing two shakes a day and one small meal. I usually have my meal at dinner time with my family. It hasn’t bothered me to cook for my family and then eat my little bit of whatever. There is no way I want to mess this up. I didn’t get this at this not follow the rules and so far following the rules has been working.

Blue = Topic or attribute definitions  ·  Green = Positive  ·  Red = Negative  ·  Black (Bold) = Neutral

---

### Acknowledgements

The authors would like to acknowledge Division of Neurosurgery, Neurosurgery, and Biomedical Informatics and Division of Neurological and Physical Medicine Devices for the input to patient preference attributes and dashboard design.
Reducing Complexity
Increasing Value

• Move away from Linguistic and Semantic Social Media Listening Methods

• Development of Mathematical Approaches to Social media Monitoring

• Reduce Complexity and Effort
• Increase Speed of Analysis
• Facilitate *de Novo* Discovery of Emerging Trends
Reducing Analytic Complexity
...bring in the math!

Singular value decomposition (SVD) is used to transform the high-dimensional, sparse term-vector representation of each document in your document collection into a reduced-dimensional, dense, numeric representation.

Move from a “bag of words” to a matrix of numbers
Analytic Approach

Text Mining

- Extract information from a collection of text documents
- Uncover the themes and concepts that are concealed in them.
- Combine quantitative variables with unstructured text and thereby incorporate text mining with other traditional data mining techniques.
The **Text Cluster** node clusters documents into disjointed sets of documents and reports on the descriptive terms for those clusters.

Two algorithms are available.
- The Expectation Maximization algorithm clusters documents with a flat representation
- Hierarchical clustering algorithm groups clusters into a tree hierarchy.

Both approaches rely on the singular value decomposition (SVD) to transform the original weighted, term-document frequency matrix into a dense but low dimensional representation.
The **Text Topic** node enables you to explore the document collection by automatically associating terms and documents according to both discovered and user-defined topics.

- Topics are collections of terms that describe and characterize a main theme or idea.
- The goal in creating a list of topics is to establish combinations of words that you are interested in analyzing.
- The ability to combine individual terms into topics can improve your text mining analysis.
- Through combining, you can narrow the amount of text that is subject to analysis to specific groupings of words that you are interested in.
Recent Studies
Mathematical Approaches

Comments Related to Pain Medication
Comments related to Medication Adherence / Compliance
Mental Health

The challenge of Improving Suicide Risk Assessment

• Suicide risk assessment in mental health usually relies on data from the patient to identify those at low, moderate or high risk.

• Most people who die by suicide are low risk and most high risk people do not kill themselves.

• It would be valuable to examine the notes made at the time of assessment and subsequent care, yet this is time consuming and is often done only during formal inquiries when a suicide has already occurred.
Suicide Risk Project
University of Ottawa & Brain and Mind Research Institute

Can we identify and change high risk systems of mental health care?

OBJECTIVE: Improving suicide risk assessment through automated and semi-automated knowledge extraction from clinical notes and narrative associated with patient assessment of suicide risk in acute and care management scenarios.

Phase I Text Analytics and Machine Learning applied to clinical notes/narrative
Phase II Develop Linguistic Rules for risk assessment based on mathematical approaches and clinical best practices
Phase III Validate, test and compare approaches against historical clinical data
Text Exploration Framework
Putting Words and Numbers Together

- Interactive application
- Iterative sub-setting with live recalculation
- Fast full-text search
- Data enrichment with SAS Entity Extraction
- Multiple aggregate views
  - Timeline
  - Map (geo-located data)
  - Phrase cloud
  - Authors
- Document tagging
- Ability to set new filters
- Alerts (coming)
- Emerging topic discovery/Unusually frequent
- Enriched data extracts for further analysis
- Viewing social media data
  - Twitter
  - News
  - Blogs
  - Forums
Text Exploration Framework
Putting Words and Numbers Together