WEBRADR WP2B – Analytics – Signal detection Workstream

Contents

1 Team Members

Michael Goodman	AZ
Antoni Wisniewski	AZ
Juergen Dietrich	Bayer
Magnus Lerch	Bayer
Carrie Pierce	Epidemico
Nabarun Dasgupta	Epidemico
John van Stekelenborg	Janssen
Amy Purrington	Janssen
Zeshan Iqbal	Janssen
Marie-Laure Kurzinger	Sanofi
Juhaeri Juhaeri	Sanofi
Stephanie Tcherny-Lessenot	Sanofi
Vroman Benoit	UCB
Nikla Noren	UMC
Ola Caster	UMC
Joanna Hajne	U Liverpool
Danushka Bollegala	U Liverpool
Simon Maskell	U Liverpool
Richard Sloane	U Liverpool

Ad Hoc:

Sabine Brosch	EMA
Telonis Panagiotis	EMA
Jim Slattery	EMA
Paolo Alcini	EMA

2 Logistics

Meeting details:

- 1/week. Thursday 8am EST (1PM GMT)
- Duration: 1 hour

Document repository:

- IMI WEB-RADR Sharepoint - https://sps-ext.nibsc.ac.uk/MHRA/imi → WPB2b → Subteam 3 Signal Detection

3 Recap of Goals

Source: Study Summary

The objective of this sub work package is to develop safety signal detection capabilities appropriate for social media data, including methods to perform local and aggregate trustworthiness assessments The deliverables of the Signal Detection subteam depend critically on the results of the WP2B – ADR Recognition team, which focuses on the accurate extraction of relevant Adverse Event information from single social media posts.

1. First, the WP2B- signal detection team will apply standard algorithms used in spontaneous adverse event databases and electronic health records (e.g. disproportionality type metrics, time trending) to data obtained from social media.
2. A second goal is to develop novel methods by adapting and optimizing these traditional signaling methods to the noisy, low-quality type data that is prevalent in social media networks. In addition, methods to assess the information quality/completeness, both at the individual post/thread level, and in aggregate will be investigated, with the goal to combine these Quality/Completeness scores with signal detection methods.
3. A third goal is to evaluate the abovementioned methods for safety signal detection in social media data.

Scope:

- Adverse Events
- Abuse/misuse
- Lack-of-Effect
- Counterfeit
- Off-label use

The positioning of social media in PV needs to be considered:

- Is social media a stand-alone safety signal detection approach?
- Does social media only play a role in Signal strengthening - ie the social media information is used to stengthen a signal from another source)

The positioning of social media for PV purposes may also depend on the drug lifecycle.

4 Product Lists

The embedded list is current as of June 29, 2015. There are 4 types of products to be used in WP2B:

1. WEBRADR products: those products that were proposed by the EFPIA partners for prospective monitoring. The list comprises 296 products, with 146 unique Active Ingredients

2. Harpaz list: 44 active ingredients to be used as a reference for signal detection methods. Eight of the products on the Harpaz list are also on the WEBRADR list.

3. TIRD (Time Indexed Reference Data): a subset of the WEBRADR products, to be used as a more refined reference data set for signal detection. All the TIRD products are on the WEBRADR list. The TIRD list consists of those products that had at least 1 proto-AE (ie a post containing a likely mention of a product an adverse event) reported in the sampling month of December 2014. As of June 29, the TIRD set is made up of 38 active ingredients. There is some overlap between the Harpaz set and the TIRD set: 7 generics appear in both sets.

4. Gold Standard products: a subset of six of the WEBRADR products, to be used for i) data characterization and ii) annotation, resulting in a human-cuated Gold Standard of posts for ADR detection. All six GS products are in the TIRD set. Four of the six GS generics are also on the Harpaz list.

The embedded document lists the WEBRADR products both by tradename and by active ingredient/generic name.

Note that the 2 generics: ‘ethinylestradiol and gestodene’ and ’ethinylestradiol, gestodene’ apeear as separate lines in the spreadsheet. These are one and the same active ingredient. The original WEBRADR product list contained 2 separate lines for these identical generics, and that redundancy has been maintained in the embedded file.

Master Product Selection list for 2B 2

5 Data Characterization

A first step in understanding the data is to characterize social media data for PV purposes. The goal is to characterize the differences between Social Media and other sources (eg spontaneous reporting data). These differences may inform our approach to signal detection.

The following data descriptors were selected:


Data Characterization FIN

The team will be using the annotation tool developed by GSK, named Insight Explorer. GSK has been conducting research on the use of social media data for pharmacovigilance purposes. As part of this project, they developed Insight Explorer for reviewing and analyzing raw social media data provided by Epidemico. As a recent addition to the WEB-RADR consortium, GSK has offered to share this tool with 2B to assist in the data curation and review processes needed for research. The WP2B ADR Detection team is developing a Gold Standard for the identification of AEs in social media posts, and will be using the same tool on the same posts.

Insight Explorer will be hosted by Epidemico using a server is in EU (Ireland) and the existing AWS cloud. It was confirmed with WEBRADR Project Management (Paul Barrow) that this will meet Data Protection Requirements

One of the attributes of interest is a “Poster Impact Score” - ie “A measure of how influential a poster is”. This descriptor could be a combination of:

- Credibility of source (not easy to determine, since if a blog has many followers, it does not automatically mean that the blog is credible)
- Connectivity
- Relevance (is the author expert in this topic?)
- Frequency (how often new contents will be published?)
- Audience (is there a bigger community who follows the author or the page, maybe special types of audience)
- Awareness-driven aspects („Like“ button, number of followers; however, this does not mean necessarily that a follower is interested in the article, on some pages reading the contents requires being a follower)
- Engagement-driven aspects (number of downloads of active contents e.g. Whitepapers)

- Outcome-driven aspects (feedback is provided, other users comment or write assessments; this parameter is thought to be a good criterion for impact score – the more feedback the higher the impact, people are motivated to comment, since they are seriously interested in a positive or negative way)

Some of these attributes are easy to automate, others virtually impossible. The team decided that a detailed algorithm to calculate this score is beyond the scope of the data characterization exercise.

5.1 Sample Size for Data Characterization

PURPOSE 1: COMPARING SOCIAL MEDIA AND SRS

Assumptions:

1. Testing is at the Product/Source level, eg methylphenidate Facebook posts
2. Each attribute is compared between Social media and an SRS (eg Vigibase)
3. Assume that variables of interest are nominal e,g
    a. Age group: child, adolescent, adult, senior
    b. Medical history present: Y/N
4. Assume that for the SRS, the attributes are usually present ie the comparison proportion is 0.8, ie we assume that 80% of spontaneous cases contain the attribute of interest. Obviosuly the real proportion depends on the attribute of interest.
5. Estimate sample size for various assumed proportions (from 0.1 to 0.7) for Social Media:

```
Test of 2 proportion:
Testing proportion 1 = proportion 2 (versus not =)
Calculating power for proportion 2 = 0.8
Alpha = 0.05

              Sample  Target
Proportion 1    Size   Power  Actual Power
         0.1       7     0.8      0.827961
         0.2      10     0.8      0.817041
         0.3      15     0.8      0.820319
         0.4      23     0.8      0.812223
         0.5      39     0.8      0.805434
         0.6      82     0.8      0.803780
         0.7     294     0.8      0.801139
```

We see that for each Product/Source combination, we would need about 300 posts to be able to distinguish between a proportion of 0.8 in the SRS and 0.7 in social media

PURPOSE 2: ASSESSING SOCIAL MEDIA PROPORTIONS

Assumptions:

1. Testing is at the Product/Source level, eg methylphenidate Facebook posts
2. Each attribute is measured and a percentage is calculated.
3. Required precision: 5%; Confidence level = 95%
4. Assume that variables of interest are nominal eg
    a. Age group: child, adolescent, adult, senior
    b. Medical history present: Y/N
5. Estimate sample size for various assumed proportions (from 0.1 to 0.9) for Social Media:

p	Sample size (assume infinite population)
0.1	138
0.2	246
0.3	323
0.4	369
0.5	384
0.6	369
0.7	323

0.8	246
0.9	138

We see that for each Product/Source combination, we would need a maximum of about 384 posts to be able to calculate a proportion with precision +/- 5%

CONCLUSION – SAMPLE SIZE:

- For each PRODUCT/SOURCE combination, randomly select 400 posts
- Therefore, for each of the 6 Gold Standard substances (which are the substances also used for data characterization), a total of 400 Facebook and 400 Tweets will be analyzed, for a total of 4,800 posts.

6 Reference Data Sets Descriptions

6.1 Gold Standard/Data Characterization dataset

The selection of products is based on Data from 1 month (December 2014). The required sample size (400 per product/source) seems not attainable for all products see 2 right-most columns in the table below. The proto-AE counts were provided by Epidemico for only 1 month (Dec 2014). Data from Datasift is available from Jan-2010 (Twitter) and Jan-2012 for Facebook. The estimates in the table are simply 36*(Proto-AE volume in DEC2014).

Active Ingredients / Generics	Twitter Ingredient Proto-AEs	Twitter Ingredient Mentions	FB ingredient Proto-AEs	FB ingredient Mentions	Total Proto-AEs	Total Mentions	AEs/ Mentions	EFPIA	EXPECTED Twitter proto-AEs	EXPECTED FB AEs	TA
methylphenidate	210	4294	9	1782	219	6076	4%	Novartis	7560	324	Psychiatry
levetiracetam	20	252	17	423	37	675	5%	UCB	720	612	Neurology
sorafenib	0	263	26	31	26	294	9%	Bayer	0	936	Oncology
insulin glargine	21	398	4	347	25	745	3%	Sanofi	756	144	Metabolic
zolpidem	15	344	0	39	15	383	4%	Sanofi	540	0	Psychiatry
terbinafine	2	76	1	1008	3	1084	0%	Novartis	72	36	Anti-infective

6.2 Harpaz reference dataset

Reference: Harpaz, R. et al. A time-indexed reference standard of adverse drug reactions. Sci. Data 1:140043 doi: 10.1038/sdata.2014.43 (2014).

- The reference dataset published by Harpaz et al. is an open access reference. The data set contains 62 positive controls and 75 negative controls spread over 44 different drugs. The positive controls represent US FDA label changes during the year 2013. Negative controls are pairs of random drugs and events among the positive controls that are not labelled as of 2013.

- Drugs in the Harpaz reference data set are defined at the substance level. Events are defined as groups of (or single) MedDRA PTs. Mapped UMLS codes are also provided for the constituent terms.

- The following 8 drugs in the Hapraz set are also on the IMI WEBRADR list of drugs:

    1. anastrozole
    2. clopidogrel
    3. clozapine
    4. levetiracetam
    5. methylphenidate
    6. oxcarbazepine
    7. sorafenib
    8. terbinafine

- Time frame for social media data collection: 1st January 2010 through Time period: data through 2013 ie straddling the label change date.
    - The start date should be set as early as possible; the choice of 1st January 2010 is based on understanding that this is when Epidemico started to gather data in Twitter.
    - End date is based on the positive controls not being labelled until 2013, which makes them 'non-established' or 'emerging' by the end of 2012. This approach should minimise inclusion of well-established associations.
    - Events for background counts: all other drugs monitored by Epidemico (~1700)
    - All extracted proto-AEs for all drugs must be cross-run against the event definitions from Harpaz' reference set. This would make it possible to compute disproportionality measures for all 62 positive and 75 negative controls based on social media data as of 31st December 2012. Hence it would be possible to assess the performance of various disproportionality analysis methods to capture emerging signals in social media data prior to US FDA labelling.

Attachment: Harpaz reference standard:

HARPAZ timeIndexedReferenc

Epidemico provided proto-AE counts for the selection of Harpaz publication products that they are currently monitoring in Medwatcher Social (MWS)- see the counts tab. The "Drugs" tab shows which of these products Epidemico are actively monitoring.

Harpaz ReferenceStandard_I

A possible issue is the low counts for the majority of drugs. The median numberof proto-AEs is 3 (for the cumulative time-period).

Quantiles		
100.0%	maximum	1105469
99.5%		1105469
97.5%		565598
90.0%		18886.8
75.0%	quartile	2887.5
50.0%	median	360
25.0%	quartile	53
10.0%		16.4
2.5%		4.4
0.5%		0
0.0%	minimum	0

Quantiles		
100.0%	maximum	119886
99.5%		119886
97.5%		66324.2
90.0%		975.4
75.0%	quartile	38.5
50.0%	median	3
25.0%	quartile	0
10.0%		0
2.5%		0
0.5%		0
0.0%	minimum	0

6.3 Time-Indexed Reference Data (TIRD)

The disadvantage of Harpaz set is that it solely relies on label changes as the true positives, which is a very high standard, and rarely attained. There are many safety signals that arize in the course of pharmacovigilance that ultimately do not lead to label changes, but have different outcomes (eg inclusion in risk management plansm, enhanced PV, followup studies). Therefore, a more complete dataset extending beyond the reference point of a label change was developed by the team.

CCDS (Company Core Data Sheets) are company's standards which pre-date country label change. This is usually driven by some signal emerging from company's safety DB. An option would be to use the first confirmed occurrence in company DB as an index date. This would be a more stringent standard as these pre-date actual label changes

Possibile options for additional index dates:

- First occurrence of spontaneous case for an event that eventually leads to a CCDS change and label change
- First occurrence of spontaneous case deemed company-related for an event that eventually leads to a CCDS change and label change
- First time drug/event combination exceeds a quantitative statistical threshold

The TIRD reference dataset consists of the following reference datapoints for a given drug and a given event coded by a specific Preferred Term:

1	Date on which this drug/event combination was entered for the first time in the safety database
2	Date on which this drug/event combination was entered for the first time in the safety database with company causality = possible or higher
3	Date on which the event became a Statistic of Disproportionate Reporting (SDR)
4	Date on which the event was identified as a Safety Signal
5	Date on which the event was added to the label – LIMITED to those events that represent a medically significant label changes after first market approval*

(*) Rationale:

- Problem with using ALL labeled events is that many of them were on the label upon first market approval, so that social media discussions are always after that date.
- Some labeled events are not 'real' ADRs, and may have a different provenance (eg suspected class effect, regulatory requirements)

Developing the dataset for a given product may be done following the steps below (if using simple Excel approach)

1. Pull event level data from Spontaneous Reporting System (SRS)
2. Run pivot tables

   Table 1:

   - PT column
   - Event date → minimum

   Table 2:

   - Filter for Company Causality = possible or higher
   - PT column
   - Event date → minimum
3. Match PTs in Table 1 with PTs in Table 2 → capture both event dates
4. Run disproportionality algorithm on historical data → capture all SDRs and their date → select earliest date for each PT
5. Manually go into Signal Tracking System →
   - Download signals
   - Map signals to PTs if necessary
   - Match Signal PTs to SRS PTs and add Signal Detection Date
6. Inspect CCDS for core data label changes

Start date and end date:

1. For single case milestone dates, the start date is the first case received for the product. End date is July 1, 2015
2. For the SDRs: Even though Epidemico will not be able to provide Facebook/Twitter data earlier than 2010, it is recommended to have a start date of ~1/1/2008 for the computed SDRs or earlier for the following reasons:

   - We may be able to retrieve Social Media data from earlier than 2010 (eg Patient Communities)

- We may find a social media signal for an AE that does not report disproportionally in our spontaneous database after 2010, but, for some reason, did exceed a threshold earlier. This is not that common, as SDRs usually stay "high" once they reach a threshold, but not always.

End date: the end date should be as far as possible into "today", especially since there seems to be a dearth of safety signals. Most companies don't have a long history of tracking have reliable signal dates starting in 2012. The TIRD end-date will be July 1, 2015.

Some practical issues may arise in developing the TIRD standard:

1. Not all PTs have a Company Causality >=possible (as expected)
2. Some signals do not have a case with Company Causality >=possible;
3. Some signals do not have a corresponding case (literature, HA inquiries)
4. Some signals cannot be mapped to a specific PT (eg "paediatric issues")
5. Some signals correspond to multiple PTs
6. Major issue - much of the case data is OLD! Most EFPIA partners confirmed that they are facing the same issue.

   As an example for one product, see table with number of "First-received events" by year for one of the products:

Year	Case Count	Percent	CumPercent
2005	1148	35.3%	35.3%
2006	534	16.4%	51.7%
2007	353	10.8%	62.5%
2008	245	7.5%	70.0%
2009	170	5.2%	75.2%
2010	141	4.3%	79.6%
2011	178	5.5%	85.0%
2012	169	5.2%	90.2%
2013	181	5.6%	95.8%

2014	115	3.5%	99.3%
2015	22	0.7%	100.0%
Grand Total	**3256**		

The initial set of drugs comprising the TIRD standard are the 38 drugs on the WEBRADR list that have at least 1 proto-AE in the period December 2014.

The breakdown by company of these 38 products:

Row Labels	Count of Active Ingredients / Generics
Amgen	4
AstraZeneca	7
Bayer	5
Janssen	1
Novartis	14
Sanofi	6
UCB	1
Grand Total	**38**

Confidentiality of the TIRD data: the EFPIA partners who hold the safety databases for their respective products will keep the product-specific TIRD data confidential.

- We recommend that none of the EFPIA companies share any SRS data at all
- We recommend that any analysis take place within each of the companies, with any results reported back in aggregate to the IMI Project team.

- As a very simplified example, an EFPIA company would look at their spontaneous AE data for a given product, and analyze whether any signal detection in social media posts for a given event(s) would have resulted in some type of timing advantage (depending on the type of detection algorithm). They would then report back that for x% of spontaneous AEs would be pre-dated by social media.
- Any social media posts that would support this analysis would be anonymized (product and event) before sharing with the IMI partners.
- For very strong and informative product/event combinations, wider sharing and publication might be considered. We may be able to present results similar to what was done in IMI/PROTECT. This work was published recently in in Drug Safety. Reference: Comparison of Statistical Signal Detection Methods Within and Across Spontaneous Reporting Databases, Drug Safety; v:38 i:6 p:577-587; 6/2015 Springer International Publishing [Cham], Candore, Gianmario; Juhlin, Kristina; Manlik, Katrin; Thakrar, Bharat; Quarcoo, Naashika; Seabroke, Suzie; Wisniewski, Antoni; Slattery, Jim; issn:01145916; eissn:11791942; doi:10.1007/s40264-015-0289-5; coden:DRSAEA; lccn:sn 90031125; itc:79220121; itcp:64506
- Highly conceptual example:
    - Method 1: 10% of PEs have significant advantage; 60% have minor advantage; 30% have no advantage
    - Method 2: 30% of PEs have significant advantage; 30% have minor advantage; 10% have no advantage
- In short, we think that we can still do analyses without sharing company data.

SDR (statistic of Disproportionate Reporting):

1. Each EFPIA partner will run the agreed-upon algorithm(s) (see section on Signal Detection Methods) for their products in their SRS
2. UMC will also run the same algorithms in Vigibase (ie the research instance) for the TIRD products. However, the research instance of Vigibase would be “frozen” and inly include cases up to January 2015

7 Historical usage data Twitter and Facebook

IMI-WEB RADR
WP2B Signal Detectio

8 Social media data requirements

8.1 Drug list

#	Purpose	Drugs	Other drugs needed for expected BG calculations
1	Data characterization	6 drugs 400 posts per source per drug	N/A
2	Method assessment using the reference data set (retrospective)	Harpaz list of 44 drugs TIRD list of 38 Drugs No score cutoff All available Facebook (1 Mar 2012 forward) and Twitter posts ((1 Jan 2010 forward)	As many as possible from Epidemico's list of 1700 Minimum of 1MM posts
3	Prospective monitoring	IMI list of drugs	As many as possible from Epidemico's list of 1700

Note on Facebook: access to text of Facebook posts may be cut off in near future. An initial cutioff date was scheduled for May1, 2015. .However, for Epidemico a ccess was extended to October.

8.2 Data availability

Datasift (Epidemico's data provide) cannot provide earlier Facebook and Twitter data due to restrictions upstream. The earliest Facebook posts are from 1 March 2012, the earliest Tweets are from 1 Jan 2010. We could gain access to earlier Twitter posts via GNIP (allegedly from 2006 onwards), but it seems that earlier Facebook data is not feasible -- at least from our existing data sources. For now, the team will work with the available social media data, and after a first pass will assess whether an effort is required to go back farther in time.

8.3 Data format for signal detection in social media data

In order to apply aggregate signal detection methods, for each product the following information is needed at the drug/event level ie each event identified in the post is one record:

Post ID	Drug	Verbatim Events (referring to drug)[1]	PTs for events (referring to drug)	Date of post	Indicator score	Source	Indications	Serious	HSI	Gender	Geography
123456	Drug A; Tecfidera	temp	pyrexia	03/12/2014	0.8724	Facebook	N/A	N	N	M	(blank)
123456	Tecfidera	cough	cough	03/12/2014	0.8724	Facebook	N/A	N	N	M	(blank)
123456	Drug B; Drug C; Tecfidera	vomiting	vomiting	03/12/2014	0.8724	Facebook	N/A	N	N	M	(blank)
367298	Tecfidera	flush	flushing	03/11/2014	0.90415	Facebook	N/A	Y	N	F	United States
367298	Drug D; Drug C; Tecfidera	fever	pyrexia	03/11/2014	0.90415	Facebook	N/A	Y	N	F	United States

1:Verbatim terms are needed for developing additional signal detection methods that do NOT rely on the coded term. However, providing the verbatims requires additional scripting by Epidemico and is not be feasible in the first iteration of the data pull. For now, ONLY the coded PT will be provided.

Since the data is uncurated, there is no attribution of the event to any one particular drug in the post. Therefore, the drug names will all appear in one column, as shown above. In future iterations of the ADR detection algorithm, specific attribution to one drug may be possible.

Some attributes may not be available for all sources or all posts, specifically:

- Indication
- HSI (Health System Interaction) - only for curated data
- Gender (sometimes prvided by DataSift)
- Geography:
  - For Twitter: geocoordinates (latitude/longitude) or use-input fields
  - Facebook: user-input fields

Conditions for data to be included:

1. **Indicator Score > threshold AND post 'Must contain a symptom'**
2. **Thresholds: {0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9}**

In addition, the same data should be provided for the "background products", where the name of the drug can be masked eg

Post ID	Drug	Verbatim Events (referring to drug)	PTs for events (referring to drug)[2]	Date of post	Indicator score	Source	Indications	Serious	HSI	Gender	Geography	Curated? (Y/N)
345	OTHER	temp	pyrexia	03/12/2014	0.8724	Facebook	N/A	N	N	M	(blank)	Y
345	OTHER	cough	cough	03/12/2014	0.8724	Facebook	N/A	N	N	M	(blank)	N
567	OTHER	vomiting	vomiting	03/12/2014	0.8724	Facebook	N/A	N	N	M	(United States	N

2:Verbatim terms are needed for developing additional signal detection methods that do NOT rely on the coded term. However, providing the verbatims requires additional scripting by Epidemico and is not be feasible in the first iteration of the data pull. For now, ONLY the coded PT will be provided.

Conditions for background data to be included:

1. **Indicator Score > threshold AND post 'Must contain a symptom'**
2. **Thresholds: {0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9}**
3. **Randomly sample 1MM posts at each threshold, under condition that each post contains a symptom, and IS> threshold**

Note: Parent/Child post identification is not possible

9 Signal Detection Methods

9.1 Summary of existing approaches

	Event of interest	All other events	Total
Drug of interest	a	b	a+b = M1
All other drugs	c	d	c+d = M2
Total	a+c = N1	b+d = N2	a+b+c+d = N

Frequentist approaches

- Proportional Reporting Ratio PRR = (a/M1) / (c/M2)
- Reporting Odds ratio ROR = (a/b)/(c/d)= (a*d) / (b*c)

Bayesian methods

- Multi-item Gamma Poisson Shrinker (MGPS) - Empirical Bayes Geometric Mean (EBGM)
- Bayesian Confidence Propagation Neural Network (BCPNN) - Information Component (IC)

Thresholds:

- Common threshold for signals[1, 2, 3:]
    - EB05 ≥2 (EB05: lower bound of the 90% CI of EBGM) – used by FDA
    - PRR ≥2, with a number of reports (N) ≥3, and a Chi-square ≥4 – more used in Europe

1. Evans SJW et al. Pharmacoepidemiol Drug Saf. 2001;10(6):483-6.
2. DuMouchel W. Am Stat. 1999;53(3):177-202.
3. Szarfman A at al. Drug Saf. 2002;25(6):381-92.

4 Results

Table 2 – Signal algorithms & definitions of statistics of disproportionate reporting employed across survey participants' databases

Partner	Algorithm	How is a statistic of disproportionate reporting defined?
UMC	IC	IC025 >0
EMA	PRR	N >2 and lower end of 95%CI on PRR >1
MHRA	EBGM	EBGM ≥2.5, EB05 ≥1.8, N ≥3
AEMPS	IC, ROR	ROR: Lower limit 95% CI >1 and No. of ICSR ≥3 IC: Lower limit 95% CI >0 and No. of ICSR ≥3
BSP	PRR	PRR ≥2, Chi² ≥4, N ≥3
AZ	EBGM	EB05 ≥1.8, and/or +ve trend flag. A trend flag is +ve if either of the following are true: • An EB05 based on current data is >EB95 for the d-e pair 52 weeks ago • An 50% increase in EBGM score when current data are compared with the EBGM score 26 weeks ago
GSK	EBGM	EB05 >2 for non-serious unlisted adverse events; any event whose reporting rate has increased significantly compared to 6 months previously

9.2 Recommended Approach

Methods:
In the first iteration, the following methods will be applied to social media data:

- PRR
- IC025

Thresholds.are not necessary to define ROC curves which should be generated to assess performance.
However, for generating the SDRs for spontanous reporting systems, the following thresholds will be used:

PRR:

- PRR>=2; N>=3
- PRR>= 2; N>=3; Chisq>=4
- Lower 95% CI of PRR >=1

IC025 :

- IC025>0

Time window: we will use a monthly sliding time-window (as SDRs are cumulative, it does not make sense to use a longer period. This is also consistent with IMI PROTECT.

Pooling:

For combination products the question arises whether reports or posts should be combined (“pooled”) for signal detection purposes.

A combination product is an actual drug (package) containing multiple ingredients eg multi-vitamin acetominophen combinations (eg with aspirin, caffeine, dextromethorphan etc). Co-medications are not considered combination products.

Options:

- We can pool together multiple products that share a common active substance, and use that for signal detection
- We can treat each product separately even those that share the same active substance

If pooling is applied, some products may end up in multiple “product pools” eg Excedrin (acetominophen/aspirin/caffeine) would end up in the acetominophen pool AND the aspirin pool.

Whether or not it makes sense to pool, also depends on the product and /or AE, and as such may be an adhoc decision. For example, it will probably be interesting to pool acetominophen products together to monitor liver injury.

Recommendation: **default is to monitor separately ie no pooling**. In certain cases (usually Consumer products), there are many combinations possible. If ***there is a single substance usually responsible for most AEs (eg acetaminophen), then we can pool.***

10 Measuring Performance

Positive Predictive Value (PPV), Sensitivity and Novelty Value (NV) are measures to characterize social media for pharmacovigilance. The most objective measures rely on the i) first identification of a particular event for a drug; and ii) the first occurrence of an SDR for a drug/event combination.

10.1 Drug/Event Occurrence – PPV/Sensitivity

- Assumptions:
    - True Positive = occurrence of reported PT in SRS
    - The set of SRS PTs is complete (ie these are the ONLY TRUE POSITIVES)
- PPV = proportion of occurrences of PTs detected in social media that correspond to a known SRS occurrence. This can be measured varying different tuning parameters in social media ADR detection. This can also be then divided into pre and post SRS occurrence
- Sensitivity = proportion of occurrences of “known” SRS occurrences that also occurred in social media. We can again split in pre and post

Example:

PT	SRS?	SOCIAL MEDIA?	TP/FP/TN/FN	POSITIVE CALL
A	1-JAN-2015	NO	FN	N
B	1-MAR-2014	1-JUN-2012	TP	Y
C	3-MAR-2013	2-JUN-2013	TP	Y
D	NO	1-MAR-2015	FP	Y
E	2-FEB-2012	1-MAR-2011	TP	Y
F	NO	1-JUN-2012	FP	Y

Summary:

- SRS has 4 known occurrences
- Social media has 5 positive calls
- Three of these also occur in social media → overall PPV = 3/5 = 0.6
- Two of these 5 occur in social media before SRS → pre-PPV = 0.4
- One of these 5 occurs in social media after SRS → post-PPV = 0.2
- Sensitivity = ¾ =0.75
- Pre-Sensitivity = 2/4 = 0.5
- Post-sensitivity = ¼ = 0.25
- Specificity = #TrueNegatives/(#TrueNegatives + #FalsePositives)

10.2 Drug/Event Occurrence – NV

- Assumption: True Positive = occurrence of reported PT in SRS **or** in Social Media. An event reported in social media but NOT in SRS is a 'Novel True Positive" (NTP)
- NOVELTY VALUE = ratio of PTs detected in social media that are NOT in SRS vs SRS. This can be measured varying different tuning parameters in social media ADR detection. Example:

PT	SRS?	SOCIAL MEDIA?	TP/NTP/FP/TN/FN	POSITIVE CALL
A	1-JAN-2015	NO	FN	N
B	1-MAR-2014	1-JUN-2012	TP	Y
C	3-MAR-2013	2-JUN-2013	TP	Y
D	NO	1-MAR-2015	NTP	Y
E	2-FEB-2012	1-MAR-2011	TP	Y
F	NO	1-JUN-2012	NTP	Y

Summary:

- Number of PTs detected in SRS = 4
- Number of PTS detected in Social Media = 5, 2 of which do NOT occur in SRS ie 2 are NTPs
- NV = 2/4 =0.5

10.3 SDR – PPV/Sensitivity

- Assumptions: i) True Positive = occurrence of SDR in SRS ii) The set of SRS SDRs is complete (ie these are the ONLY TRUE POSITIVES)

- PPV = proportion of SDRs in social media that correspond to a known SRS SDR. This can be measured for different signal detection methods, and by varying different tuning parameters in each signal detection method. This can then be divided into pre and post SRS occurrence
- Sensitivity = proportion of occurrences of "known" SDRs in SRS that also alerted in social media. We can again split in pre and post
- Example for a particular Method ("method 1"):

PT	SDR in SRS	Alert (Method 1) Social Media	TP/FP/TN/FN	POSITIVE CALL
A	1-JAN-2015	NO	FN	N
B	1-MAR-2014	1-JUN-2012	TP	Y
C	3-MAR-2013	2-JUN-2013	TP	Y
D	NO	1-MAR-2015	FP	Y
E	2-FEB-2012	1-MAR-2011	TP	Y
F	NO	1-JUN-2012	FP	Y

- SRS has 4 known occurrences
- Social media has 5 positive calls
- Three of these also occur in social media → overall PPV = 3/5 = 0.6
- Two of these 5 occur in social media before SRS → pre-PPV = 0.4
- One of these 5 occurs in social media after SRS→ post-PPV = 0.2
- Sensitivity = ¾ =0.75
- Pre-Sensitivity = 2/4 = 0.5
- Post-sensitivity = ¼ = 0.25

10.4 SDR – NV

- Assumptions: True Positive = occurrence of reported PT in SRS or in Social Media. An SDR in social media but NOT in SRS is a 'Novel True Positive" (NTP)
- NOVELTY VALUE = ratio of SDRs detected in social media that are NOT in SRS vs SRS. This can be measured for different signal detection methods, and by varying different tuning parameters in each signal detection method.
- Example for a particular Method ("method 1"):

PT	SDR in SRS	Alert (Method 1) Social Media	TP/NTP/ FP/TN/FN	POSITIVE CALL
A	1-JAN-2015	NO	FN	N
B	1-MAR-2014	1-JUN-2012	TP	Y
C	3-MAR-2013	2-JUN-2013	TP	Y
D	NO	1-MAR-2015	FP	Y
E	2-FEB-2012	1-MAR-2011	TP	Y
F	NO	1-JUN-2012	FP	Y

Summary:

- Number of SDRs in SRS = 4
- Number of SDRs in Social Media = 5, 2 of which do NOT occur in SRS ie 2 are NTPs
- NV = 2/4 =0.5

10.5 Timing Metrics

In addition to the metrics above, we can also characterize methos by calculating simple statistis on their relative time advantages with respect to a spontaneous reporting benchmark.

For the true positives, this would include characterizing the distribution in differences

1. AVERAGE/MEDIAN of difference in detection
2. STANDARD DEVIATIONs
3. PERCENTILES

This may be broken out further by Adverse Event, product, or other relevantstartification factors.

11 Harpaz Results

The embedded file summarizes the findings of the analysis of the Harpaz products. In conclusion, disproportionality analysis on Facebook and Twitter data performed considerably worse than on global spontaneous reporting data, when benchmarked against historical label changes. This is likely due to limited prevalence or retrieval of the labelled events from social media, rather than the disproportionality methods themselves.

UMC_ICPE_2016_OlaCaster_900x1200_f

12 Policy Questions for WP1

1. Would an obligation exist for MAHs and CAs to monitor social media (not under their control) for signal detection purposes?
2. What would be the obligation of an MAH if a 'signal' was found in social media:
    a. Regarding drilldown: would an MAH be obliged to drilldown to the post level?
    b. Would MAH have an obligation to enter corresponding posts in the SRS? (assuming these are de-identified)
3. What would be the obligation of an MAH while reviewing social media posts (regardless whether these are associated with a safety signal) eg if an invalid AE is found, should it be entered in the SRS?
4. Conducting Follow-Up on Adverse Events Found on Social Media
5. What is primary source data and how should it be archived (links, graphics, video, PDF, audio….)
6. WEBRADR: is it permissible to make the Gold Standard Annotation Dataset available for access?
    - Potential issue: Even if we remove PII, the post itself may still be identifiable to the actual poster or his/her friends/family. I am really curious what our PII experts have to say about that. For example, if I were to post something very detailed on Facebook relating my experience with cancer drugs (hypothetical example), the actual content of the post (ie the very specific experiences, locations, dates) may still be identifiable to some individuals. ***Publishing*** these posts by providing free access on behalf of IMI (and its participants) takes things a step further than **analyzing** posts. We would need legal insight, and perhaps an Ethics Committee to oversee?

Appendix A: PRR, Chi-Square and IC algorithms

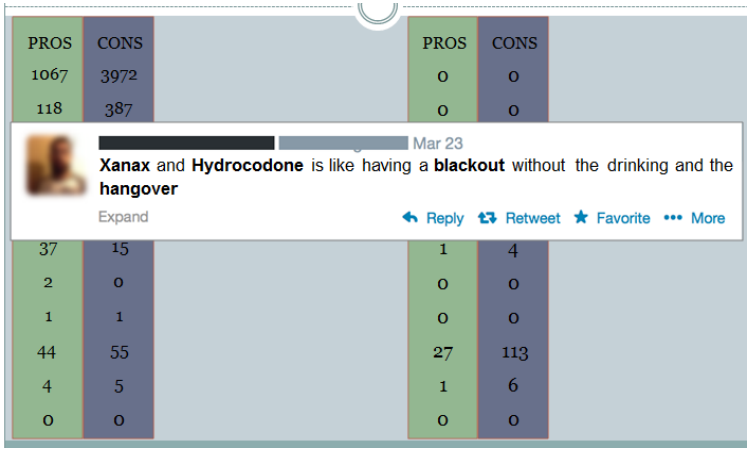
```
*************************************************************************************;
**********                        Prr, Chi-square, UMC IC macro                ************;
*************************************************************************************;
%macro basic_sdr;
*** Calculation of PRR, Chi-square and UMC Information component ***;
a_b = a + b;
a_c = a + c;
b_d = b + d;
c_d = c + d;
a_b_c_d = a + b + c + d;

*** Prr ***;
if c > 0 then
     do;
          prr = (a / a_b) / (c / c_d);
          std_err = sqrt ( (1 / a) + (1 / c) - (1 / a_b) - (1 / c_d) );
     end;
           else
     do;
          c1 = 0.5;
          prr = (a / a_b) / (c1 / (c1 + d));
          std_err = sqrt ( (1 / a) + (1 / c1) - (1 / a_b) - (1 / (c1 + d)) );
     end;
prr_left_95 = prr / exp (1.96 * std_err);
*prr_right_95 = prr * exp (1.96 * std_err);

*** Chi-square ***;
exp_a = (a_b * a_c) / a_b_c_d;
exp_b = (a_b * b_d) / a_b_c_d;
exp_c = (c_d * a_c) / a_b_c_d;
exp_d = (c_d * b_d) / a_b_c_d;
chi_square = (((a - exp_a) ** 2) / exp_a) + (((b - exp_b) ** 2) / exp_b) + (((c - exp_c) ** 2) / exp_c) + (((d - exp_d) ** 2) / exp_d);
pp_chi_square = (1 - cdf('Chisquare', chi_square, 1));

*** Yate Continuity Correction ***;
chi_yate = (((abs(a - exp_a) - 0.5) ** 2) / exp_a) + (((abs(b - exp_b) - 0.5) ** 2) / exp_b) + (((abs(c - exp_c) - 0.5) ** 2) / exp_c) + (((abs(d -
exp_d) - 0.5) ** 2) / exp_d);
pp_yate = (1 - cdf('Chisquare', chi_yate, 1));

*** UMC IC ***;
ratio = (a + 0.5) / (exp_a + 0.5);
ic = log2(ratio);
*ic_025_appr = log2 (ratio) - 3.3 * ((a + 0.5) ** (-0.5)) - 2.0 * ((a + 0.5) ** (-1.5));
```

```
ic_025 = log2 ( quantile ('gamma', 0.025, a + 0.5, (1 / (exp_a + 0.5))) );
```

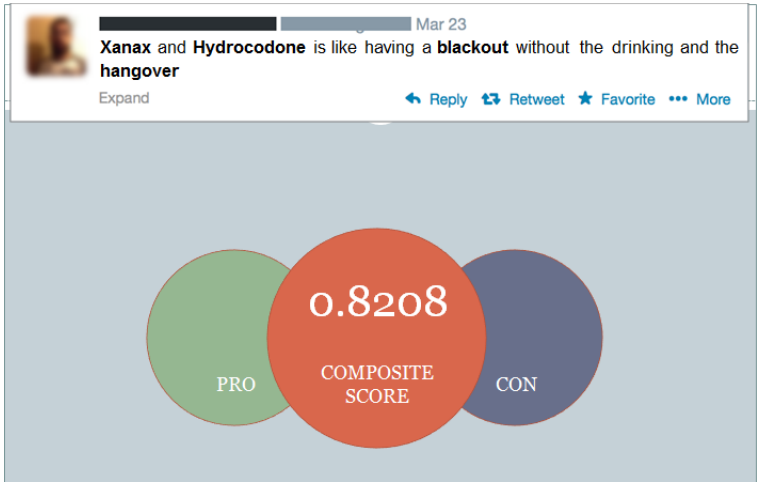
Appendix B: Epidemico Classifier

1. We trained a Bayesian classifier through machine learning by manually coding ~360,000 posts that mentioned a medical product. Posts were analyzed to determine whether or not they described a possible adverse event and then tagged as a Proto-AE (post with resemblance to an adverse event) or not.

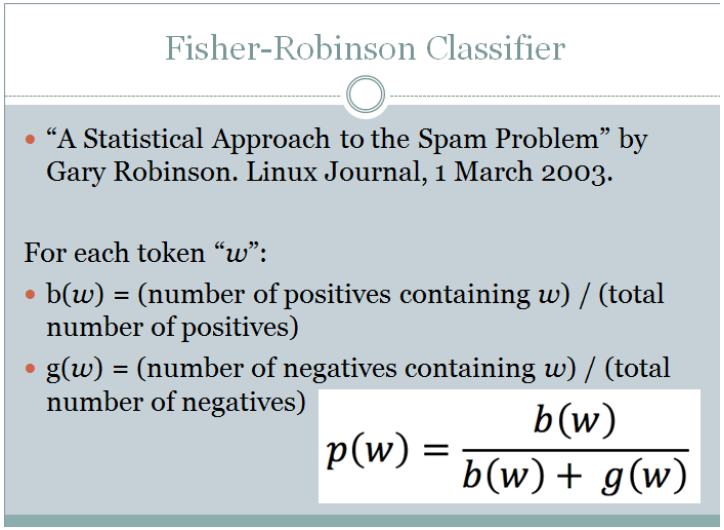
2. Each post that is acquired is broken into one-to-four word phrases. The classifier then compares each phrase to instances where it occurs in the training set. The classifier adds up the nuymber of times each phrase occurs in a manually-coded Proto-AE (“Pros” in this slide) and the number of times each phrase occurs in a manually-coded non-ProtoAE (“Cons” in picture below).



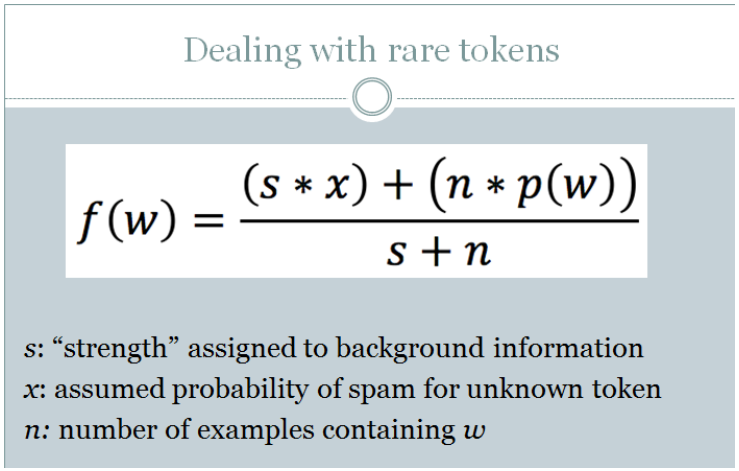
3. The classifier then calculates a composite score from the totals of “Pros” and “Cons” identified in the comparison to the training set data.



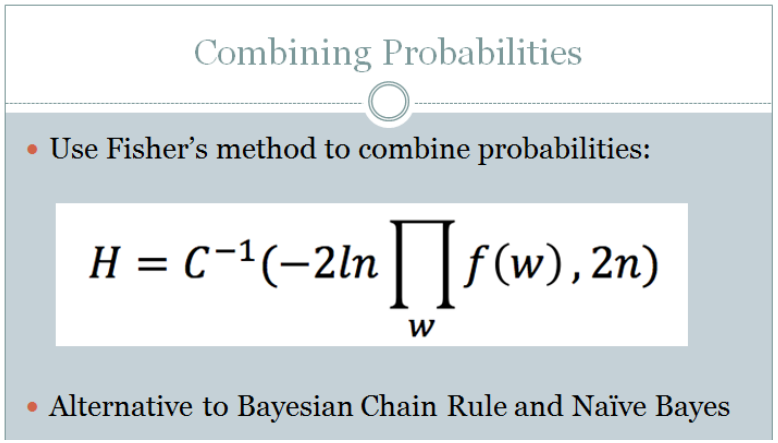
4. p(w) represents the probability of spam for the given word



5. The algorithm provides consideration for rare tokens:



6. Use Fisher's method to combine probabilities (Robinson claims Fisher is empirically better than Bayesian Chain or Naïve Bayes):



7. Compute Indicator Value I:

- Combine inverse probabilities:

$$S = C^{-1}(-2ln\prod_{w} 1 - f(w), 2n)$$

- Average H and S:

$$I = \frac{1 + H - S}{2}$$

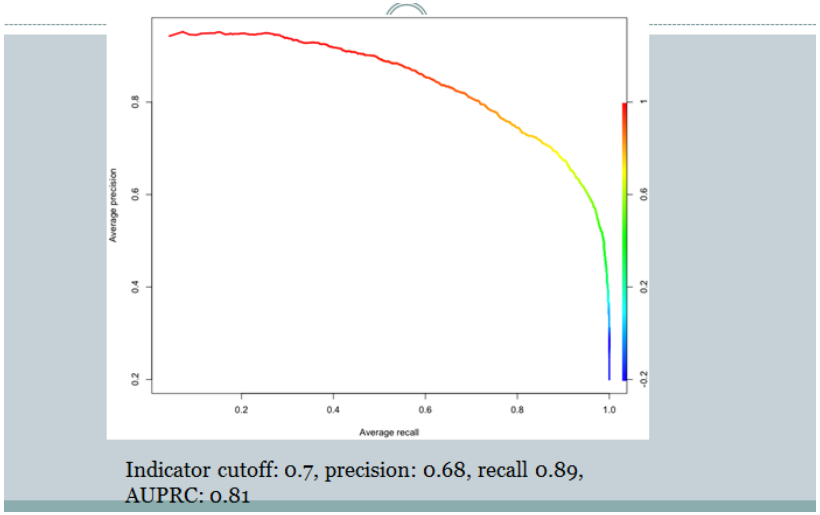
8. Current Automated Classifier:

- Recent test run, attempt to predict AE vs non-AE for 16,834 tweets:

Actual	Predicted	
	AE	Not AE
AE	3,054	1,418
Not AE	394	11,968

- Indicator cutoff 0.7:
  - sensitivity: 0.89; specificity: 0.89
  - precision: 0.68; recall: 0.89; F-score: 0.77

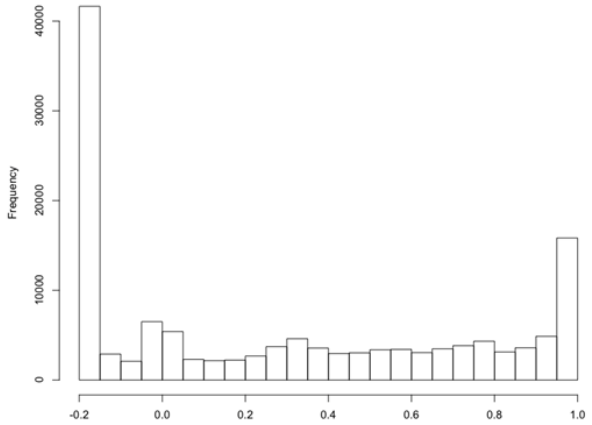
9. Document Classification Performance



Indicator cutoff: 0.7, precision: 0.68, recall 0.89, AUPRC: 0.81

10. Distribution of Indicator scores

Distribution of Indicator, Fisher-Robinson



11. Inverse Chi-squared Function:

```
def C⁻¹(chi, df):
    m = chi / 2.0
    sum = term = math.exp(-m)
    for i in range(1, df / 2):
        term *= m / i
        sum += term
    return sum
```

Appendix C: Facebook and Twitter Tags provided by Datasift/Epidemico

Epidemico data fields.xlsx

Appendix D: Sample Epidemico Data Records

Sample Epidemico data records.docx

Appendix E: SQL script for retrieving cases from ARGUS

The script below shows how to select cases from an ARGUS database corresponding to the earliest observation of an AE for a product, and the earliest observation of an AE with causality >=possible.

```
SELECT a.art_code,
       a.pref_term,
       a.earliest_mentioned,
      b.case_num
  FROM (  SELECT art_code,
                 pref_term,
                 MIN (NVL (CEA.UPDATED, CEA.EFFECTIVE_START_DATE))
                   AS earliest_mentioned
            FROM v$CASE_EVENT ce,
                 v$case_master cm,
                 dlp_owner.dlp_case_event_Assess cea,
                 v$case_product cp,
                 v$lm_product lp
           WHERE     cm.case_id = ce.case_id
                 AND cea.case_id = cm.case_id
                 AND CP.CASE_ID = cm.case_id
                 AND cea.license_id = 0                          -- world
                 AND cea.event_seq_num = ce.seq_num
                 AND cp.seq_num = CEA.PROD_SEQ_NUM
                 AND NVL (cp.product_id, cp.pat_exposure) = LP.PRODUCT_ID -- company drug
                 AND cp.drug_type = 1                         -- suspected
                 AND lp.drug_code = 'BAYxxxx' – Sample drug code
        GROUP BY art_code, pref_term
        ORDER BY 1, 3 ASC) a,
        (  SELECT art_code,
                 pref_term,
                 MIN (NVL (CEA.UPDATED, CEA.EFFECTIVE_START_DATE)) as earliest_mentioned,
                cm.case_num
                    FROM v$CASE_EVENT ce,
                 v$case_master cm,
                 dlp_owner.dlp_case_event_Assess cea,
                 v$case_product cp,
                 v$lm_product lp
           WHERE     cm.case_id = ce.case_id
```

```
AND cea.case_id = cm.case_id
AND CP.CASE_ID = cm.case_id
AND cea.license_id = 0                          -- world
AND cea.event_seq_num = ce.seq_num
AND cp.seq_num = CEA.PROD_SEQ_NUM
AND NVL (cp.product_id, cp.pat_exposure) = LP.PRODUCT_ID -- company drug
AND cp.drug_type = 1                            -- suspected
AND lp.drug_code = 'BAYxxxx'
      GROUP BY art_code, pref_term, case_num) b
where a.earliest_mentioned = b.earliest_mentioned
and a.art_code = b.art_code
order by 2
```

References

To be completed