WEB-RADR WP2b – STUDY SUMMARY

Analytics - Signal Detection

Type of document:	Study summary
Version:	3.0
Date:	2DEC2014
Authors:	John van Stekelenborg
	Stéphanie Tcherny-Lessenot
	Marie-Laure Kürzinger
	Ola Caster

Table of contents

Introduction

1.1 Purpose of the document

The purpose of this document is to serve as the Work Package 2b – Signal Detection team's operational plan for how to organise and realise the deliverables of this WP as described in the Description of Work (DoW).

1.2 Version history

Version	Date	Comments	Authors
01	27 February 2019		JvS

1.3 Definitions and abbreviations

For definition and abbreviations in this document, please refer to the WEB-RADR Definitions and Abbreviations Document.

2 Rationale & Study Proposal

2.1 Background

Reporting by healthcare professionals (HCPs) and patients by means of mobile devices and the strengthening of safety signals from social media platforms to National Competent Authorities (NCAs) and marketing authorisation holders (MAHs) with subsequent transmission to EudraVigilance is a new and unexplored concept. In addition, the vast amount of information generated through social media requires a well-defined approach as regards monitoring, reporting, analysing and evaluating potential adverse reactions, signals and other medical insights related to medicines.

The automatic identification of pharmacovigilance data such as medication names and adverse events in 'free text', which is the format of tweets, blog posts etc. in social media is a necessary first step in order to apply methods capable of identifying safety signals in such data.

Commercial products (e.g. MedLEE) exist that provide automatic identification of references to medical entities in free text. There are also open source frameworks (e.g. UIMA/cTAKES) that allow further development of methods.. Both existing products and open source frameworks have the capability of mapping identified ADRs to standardized medical terminologies such as MedDRA.

However, the vast majority of studies investigating NLP methods applied them on narratives authored by health care providers in *English* with the aim of providing correct complete information using *standardised vocabularies*. In the current project however, social media will automatically be analysed where information may be *noisy*, *incomplete*, expressed in *non-standard vocabularies*, in *multiple languages*. To address some of these challenges, Freifeld and colleagues demonstrated how a tree-based dictionary-matching algorithm could be used to identify both medicinal products and adverse events in Twitter posts (Freifeld CC, Brownstein JS, Menone CM, Bao W, Filice R, Kass-Hout T and Dasgupta N. Digital Drug Safety Surveillance: Monitoring Pharmaceutical Products in Twitter. Drug Safety, 37:343-350, 2014).

2.2 Purpose

The overall aim of WP2b is to develop and link new and existing analytical tools for the analysis of social media content for pharmacovigilance purposes. WP2b therefore focuses on improving the quality and relevance of data extracted from social media sources and their subsequent use in the detection of safety signals. In order to achieve its objectives WP2b will focus on three areas: 1) Suspected ADR identification; 2) Record linkage and 3) Signal detection.

This study summary applies to the sub work package is “3) Signal Detection”.

3 Study Objectives & Timelines

3.1 Study Objectives

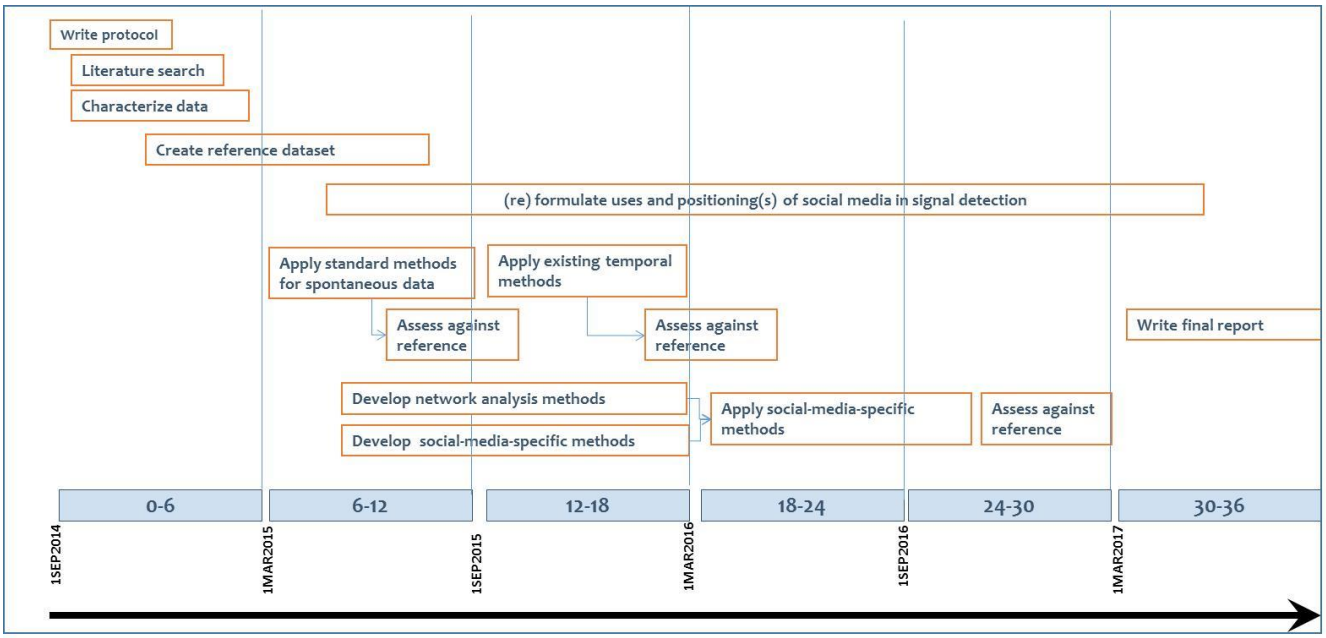
- The overall objective of this sub work package is to develop safety signal detection capabilities appropriate for social media data, including methods to perform local and aggregate trustworthiness assessments (dependency with WP2B – ADR Recognition).
    1. First, the WP2B- signal detection team will 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. (cf task T2B.3.1 and T2b.3.2 in section 4.1))
    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 (cf Task T2b.3.3 in section 4.1). 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. For example, approaches using Bayesian Belief Networks and a process called track‐before‐detect (TkBD) (Rutten 2005) will be investigated.
    3. A third goal is to evaluate the abovementioned methods for safety signal detection in social media data (cf Task T2B.3.4 section 4.1).

3.2 Milestones & Timescales

3.2.1 High-Level milestones

Milestone #	Milestone	Estimated date of completion
M2B.7	Implement standard analysis methods as previously used for signal detection in spontaneous reports, e.g. disproportionality analysis	M12
M2B.8	Temporal analysis implemented , as previously used for signal detection in electronic patient records	M24
M2B.9	Novel methods for safety signal detection in social media implemented	M24
M2B.10	Evaluation of methods for safety signal detection in social media	M28

3.2.2 Detailed project plan and interim deliverables:



3.3 Dependencies

Refer to milestones in WP2a, WP2b (ADR recognition/Record linkage) / WP4 that are either inputs/outputs

Task or Milestone	WP	Milestone	Month	Dependency with WP2b – signal detection
T2A.1	WP2A	Product selection	M2	WP2A → WP2B
T2A.2	WP2A	Create curation platform This will be used as a platform to train the classifier in new languages, develop dictionaries, and refine machine learning algorithms.	M6	WP2A → WP2B
T2A.3	WP2A	Create user interface and database backend	M6	WP2A → WP2B
T2A.7	WP2A	Provide data for WP2b and WP4 Data will be extracted at M6 and M24 for use within the analytics and scientific impact evaluation work packages.	M6 and M24	WP2A → WP2B
T2A.9	WP2A	Create shared collaborative workspace	M25-30	WP2A → WP2B
M2A.7	WP2A	Incorporate feedback from WP2b (ongoing).	M30-35	WP2B → WP2A
T2B.1.2	WP2B-1	Implement existing methods for suspected ADR identification	M6	WP2A-1 → WP2B-3
T2B.1.5	WP2B-1	Implement novel methods for suspected ADR identification	M15	WP2A-1 → WP2B-3
T2B.2.3	WP2B-2	Implement record linkage	M21	WP2A-2 → WP2B-3
T4.7	WP4	Integration of data from different sources and assessment of trustworthiness Reports received from paper reports, website reports, electronic health records, apps, and social media will be integrated together with assessment of trustworthiness.	M35	WP4 ←→WP2B

3.4 Deliverables

Deliverable No.	Deliverable description	Nature (R, P or O)	Expected delivery date
D2B.3	Technical report describing implementation and evaluation of safety signal detection in social media	R	M32

3.5 Participants

The following will be participating in the project:

Name	Organization	Mail
John van Stekelenborg (lead)	Janssen	jvanstek@its.jnj.com
Amy Purrington	Janssen	APurring@its.jnj.com
Geoff Gipson	Janssen	ggipson@its.jnj.com
Zeshan Iqbal	Janssen	ziqbal@ITS.JNJ.com
Michael Goodman	AZ	Michael.Goodman@astrazeneca.com
Niklas Noren	UMC	Niklas.Noren@who-umc.org
Juhaeri Juhaeri	Sanofi	Juhaeri.Juhaeri@sanofi.com
Danushka Bollegala	U Liverpool	danushka.bollegala@liverpool.ac.uk
Simon Maskell	U Liverpool	S.Maskell@liverpool.ac.uk
Stephanie Tcherny-Lessenot	Sanofi	Stephanie.Tcherny-Lessenot@sanofi.com
Ola Caster	UMC	ola.caster@who-umc.org
Marie-Laure Kurzinger	Sanofi	Marie-Laure.Kurzinger@sanofi.com
Juergen Dietrich	BAYER	juergen.dietrich@bayer.com
Antoni Wisniewski	AZ	Antoni.Wisniewski@astrazeneca.com
Carrie Pierce	Epidemico	carrie@epidemico.com
Nabarun Dasgupta	Epidemico	nabarun@epidemico.com
Jim Slattery	EMA	Jim.Slattery@ema.europa.eu
Panagiotis Telonis	EMA	Panagiotis.Telonis@ema.europa.eu
Sabine Brosch	EMA	Sabine.Brosch@ema.europa.eu

4 Methodology

This sub package contains 4 tasks:

Task No.	Task Title / Description
T2B.3.1	Implement cross-sectional analysis methods for safety signal detection in social media, as previously used for signal detection in spontaneous reports, e.g. disproportionality analysis (cf. study objective 1 in section 3.1)
T2B.3.2	Implement temporal analysis methods for safety signal detection in social media, as previously used for signal detection in electronic patient records (cf. study objective 1 in section 3.1)
T2B.3.3	Develop novel methods for safety signal detection in social media Explore novel approaches to safety signal detection in social media, including evaluation of local and aggregate trustworthiness assessments, consideration of Information Quality/Completeness at the aggregate level, accounting for the presence of confounders, and adapting methodology from other areas such as Track-Before-Detect and Bayesian belief networks. We will consider integration of these novel approaches with existing methodology under 2B.3.1 and 2B.3.2. (cf. study objective 2 in section 3.1)
T2B.3.4	Assess relative performance of selected methods Evaluate relative performance against reference set of positive and negative controls from WP4. These may either be historical safety signals in recent years or prospective safety signals during the course of the project. (cf. study objective 3 in section 3.1)

4.1 Implement Current Signal Detection methods

Pre-work:

1. Characterize social media data for PV purposes. The purpose of this exercise is to understand the relevant differences between Social Media and other sources (eg spontaneous reporting databases). These differences may inform our approach to signal detection
    a. Determine relevant descriptors
        - General population descriptors eg gender, age, location,…..
        - Specific descriptors within each of the social media types (Twitter, Facebook, Patient forums, blogs).
    b. Using the agreed-upon descriptors, characterize social media data, spontaneous data, EHR data
2. Formulate various options for positioning of social media in ‘PV surveillance’ eg
    - Signal detection

- Signal strengthening
- Evidence ranking only

The different types of positioning and use of social media monitoring may depend on the drug characteristics, e.g., its current lifecycle.

3. Assessment of the viability of developing a reference data set
4. Scope determination for signal detection:
    - ADRS are in scope, and we will investigate whether these are in scope:
    - Abuse/misuse
    - Lack-of-Effect
    - Counterfeit
    - Off-label use
    - Potential medication errors
    - Overdose (note that the ability to detect overdose is likely strongly dependent on data quality)
5. Determine appropriate level(s) of analysis: verbatim, LLts, PTs
6. Determine options for 'exposure' in social media
7. Investigate applicability of network analysis (source analysis) – stimulated reporting/key-opinion-leaders/dependency of posts; identifying sites/posters that create "noise" in the signal. (potential overlap with 'Record Linkage' team)

Application of existing cross-sectional signal detection methods:

1. Inventory of existing disproportionality methods
2. Assess applicability and select method(s) of preference
3. Implement in Medwatcher Social (dependency on WP2a)
4. Execute monitoring plan on selected products

Contingent on earlier work: develop reference data set for signal detection in social media detection methods:

1. Retrospective time-indexed set of drug and/or drug classes and true positives/true negatives
2. Pull posts retrospectively for drugs/drug-classes of interest corresponding to 'signals' and non-signals
3. Define statistical measures of performance (investigate whether traditional measures eg specificity, sensitivity, PPV) are applicable)

4.2 Implement temporal analysis methods for safety signal detection in social media

1. Inventory of existing trending methods
2. Assess applicability and select method(s) of preference, Assess agains reference dataset.
3. Implement in Medwatcher Social
4. Execute monitoring plan on selected products

4.3 Develop novel methods for safety signal detection in social media

- Use characteristics of social media data (as determined in 4.1) to modify existing disproportionality and trending methods
- Develop methods to assess trustworthiness of social media signals (note: does the concept of 'trustworthiness' lend itself more to the single post level, ie subteam1?)
- Inventory of potential methods from other fields
- Investigate and adapt methods used from other fields eg Track-Before Detect

4.4 Assess relative performance of selected methods

Retrospective assessment:

- Apply traditional and new methodologies to reference data set (retrospectively, with data retrieved for positive controls according to their respective time-index)
- Assess methods using agreed-upon measures (see 4,1)

Prospective assessment:

- Define measures of success
- Apply traditional and new methodologies prospectively