On Social Media Mining for HIV Pre-Exposure Prophylaxis

C. Kakalou¹, J.V. Lazarus², V. Koutkias¹

¹Institute of Applied Biosciences, Centre for Research & Technology Hellas, Thermi, Thessaloniki, Greece
²Barcelona Institute for Global Health (ISGlobal), Hospital Clínic, University of Barcelona, Spain

Contact: cakalou@certh.gr
Social Media Mining for Public Health
Social Media (SoMe) Mining

«Social Media Mining is the process of representing, analyzing, and extracting actionable patterns from social media data.»

How is it done?

- Natural Language Processing
- Text mining
  - Sentiment Analysis
  - Event Detection
- Network modelling
Why is it important?

- Collection of user-generated data: spontaneous, unsolicited and up-to-date, with a high degree of variance

- Obtain insights from underrepresented groups of patients or from people that don’t have easy access to the healthcare services

  (“Listening to the patient’s voice”)

- Information that patients are reluctant or even afraid to provide to healthcare providers
What to watch out for

- Issues with credibility, recency, uniqueness, frequency and saliency
- Novel information sources
  - Different analysis approaches
  - Difficulties in evaluating results
- “Big Data Hybris”: Should complement, not replace traditional techniques
- Ethical Implications (Public data ≠ available for research?)
Application: Mining PrEP Trends on Twitter
Research Question

Can we identify PrEP-related trends discussed on Twitter?
Research Question

- What is the sentiment regarding PrEP?
- Investigate public awareness, views and emotions
- Detect main discussion topics
- Are there any side-effects mentioned?
- What are the technical difficulties on the systematic collection & analysis of Twitter data regarding Public Health?
3 Methodology and Implementation
Our data gathering & analytics Web platform

- Developed and maintained by the Institute of Applied Biosciences (INAB)
- Using open-source technologies
- Unstructured data gathering from multiple sources
- Integrating various text-mining and visual analytics methods
- Aims to support public health surveillance scenarios
Keywords

2 datastreams:

- truvada OR (truvada AND PrEP) OR (hiv AND PreP)
  OR (preexposure AND prophylaxis)

- (pill AND prevent AND HIV) OR (pill AND protect AND HIV) OR (pill AND protect AND AIDS)
31,854 tweets
“PrEP” stream

804 tweets
“Pill” stream

~4 months
Continuous data collection
Data Analysis

- Only tweets in English collected
- Identification of hashtags (#), mentions (@), links (t.co) and location (📍)
- Hashtags are kept in the text corpus, since they give meaning
- Text Cleaning/Preprocessing
- Hashtags tracking
- Geographical mapping of tweets
- Sentiment Analysis ➔ 2 adapted sentiment lexicons (AFINN & NRC)
#PrEP #Truvada

Sample tweets
RT @shyronn_jones: Things that protect us from HIV: Undetectable Partner, PrEP, PEP, Condoms. Things that don't: HIV Criminalization. #HIV
This truvada is TEARING my stomach tf up tonight
Yes. Placebo PREP trials (more ok in women and LMICs!?)—low adherence—null results. Let's treat cis & trans women?

https://t.co/WDNnUbFTDv
4 Results – Exploratory Data Analysis
Overview – “PrEP” datastream

Analysis of Twitter data regarding the use of Pre-Exposure Prophylaxis (PrEP)

Data collected from 1 September 2017 to 21 December 2017

Keywords used:

truvada, truvada & PrEP, PrEP & HIV, pre-exposure prophylaxis

31854 Tweets collected
19162 Retweets
111 Days collecting data

Original vs Retweets:
Overview – “Pill” datastream

Analysis of Twitter data regarding the use of Pre-Exposure Prophylaxis (PrEP)

Data collected from 1 September 2017 to 21 December 2017

Keywords used:

pill prevent HIV, pill protect HIV, pill protect AIDS

- 804 Tweets collected
- 502 Retweets
- 111 Days collecting data

Original vs Retweets:
Analysis of Twitter data regarding the use of Pre-Exposure Prophylaxis (PrEP)

Data collected from 1 September 2017 to 21 December 2017

Sentiment Polarity per Day

Average Sentiment Polarity: 0.208

PrEP dataset is selected
Influencers

Accounts posting heavily about PrEP, representing more than 9% of total posts collected.
Hashtags Timeline

World AIDS Day
#WorldAIDSDay
1 Dec.

Ireland’s first clinic monitoring PrEP opens
#Ireland
9 Nov.
Geography of Tweets

Regulatory Status of Truvada for PrEP

*Approved via import license from South Africa.
1. India, SA and Kenya have generic PrEP approvals.
2. The European Commission granted Gilead marketing authorization for Truvada as PrEP in Q3 2016. This should encourage countries within the EU to make PrEP available within their national health systems, based on cost factors and individual country regulatory requirements.

Updated October 2017
Geography of Tweets

Analysis of Twitter data regarding the use of Pre-Exposure Prophylaxis (PrEP)

Data collected from 1 September 2017 to 21 December 2017

The geography of PrEP on Twitter
Sentiment Geography of Tweets

Analysis of Twitter data regarding the use of Pre-Exposure Prophylaxis (PrEP)

Data collected from 1 September 2017 to 21 December 2017

The geography of PrEP sentiment on Twitter

Negative Sentiment
Neutral Sentiment
Positive Sentiment
5 Discussion
Conclusion

- Large volume of data with low amounts of noise (irrelevant information)
- PrEP discussions can be successfully identified on Twitter
- PrEP discourse seems to be continuously increasing
- Overall sentiment is positive
- Main concerns: stigma, medication cost, insurance coverage and vulnerable population groups
- Detection of critical events is possible
Challenges / Points of Interest

- NLP tasks harder for short & medical-related twitter text
- Sarcasm, humour, subculture references hard to detect
- Machine learning needed to classify personal to informational tweets
- «Prolific» Tweeters: Analyse timeline and social sphere/network
- Monitoring of the propagation of information through the retweeted/shared links
Future Work

◉ Create a specialized HIV/PrEP lexicon
◉ Create an annotated dataset by domain experts (sentiment, NLP Tasks, personal experience detection)
◉ Improve text mining and NLP analysis of data
◉ Further use of medical terminologies (MeSH, MedDRA)
◉ Exploit emoticons for further sentiment analysis
◉ Make the platform available to researchers, healthcare professionals and policy makers
Thank you! Questions?

On **Social Media Mining** for HIV Pre-Exposure Prophylaxis

C. Kakalou\textsuperscript{1}, J.V. Lazarus\textsuperscript{2}, V. Koutkias\textsuperscript{1}

\textsuperscript{1}Institute of Applied Biosciences, Centre for Research & Technology Hellas, Thermi, Thessaloniki, Greece
\textsuperscript{2}Barcelona Institute for Global Health (ISGlobal), Hospital Clínic, University of Barcelona, Spain

Contact: [ckakalou@certh.gr](mailto:ckakalou@certh.gr)