

# On Social Media Mining for HIV Pre-Exposure Prophylaxis

C. Kakalou<sup>1</sup>, J.V. Lazarus<sup>2</sup>, V. Koutkias<sup>1,\*</sup>

\*Contact: ckakalou@certh.gr

<sup>1</sup>Institute of Applied Biosciences, Centre for Research & Technology Hellas, Thessaloniki, Greece

<sup>2</sup>Barcelona Institute for Global Health (ISGlobal), Hospital Clínic, University of Barcelona, Spain



## Background

Pre-exposure prophylaxis (PrEP) is the administration of antiretroviral therapy to prevent HIV infection in seronegative persons. WHO recommends that high-risk individuals should be offered PrEP as an additional prevention choice. Twitter is a popular microblogging platform engaging around 330 million active monthly users.

## Aim

The aim of this study was to investigate PrEP trends in Twitter discussions regarding awareness and expressed sentiment about PrEP, as well as social stigma and side-effects associated with PrEP use

## Main challenges

- Vastness of data, no structure
- Informal, colloquial text but medical-related
- High amount of noise (irrelevant information)
- Ethical implications (public data ≠ available for research)

## Methods

Using open-source technologies, INAB developed a comprehensive data gathering and analytics platform to address the study objectives. Data acquisition from September 2017 to December 2017 relied on the Twitter streaming Application Programming Interface (API). Various text-mining and visualization techniques were incorporated in the platform including sentiment analysis, topic modeling and geographic distributions, and made available via a user-friendly Web-based application. The recently released MedDRA® patient-friendly Low-Level Term (LLT) lexicon was employed to identify mentions of side-effects.

## Conclusions

Our results demonstrate that screening Twitter discussions over time may reveal events impacting the target group's reaction to PrEP. Further research is required to advance the analytics methods (e.g. information filtering and natural language understanding). Our analytics platform can provide insights to public health professionals and policy-makers, helping them to develop strategies for improving HIV prevention.

## Future work

- Create a specialized HIV/PrEP lexicon and an annotated dataset by domain experts
- Exploit further medical terminologies
- Exploit emoticons for further sentiment analysis

## Results

**31,854** tweets were collected using query terms referring explicitly to PrEP and **804** using more abstract terms like "pill prevent HIV". 23 Twitter accounts posted heavily (>9% of the total posts) about PrEP, constituting "PrEP influencers," who contributed to the overall positive sentiment about PrEP. The geographical distribution of tweets was found in accordance with the UNAIDS 2016 map of countries having approved or piloting PrEP. Mentions of social stigma and complaints about potential side-effects were identified, e.g. headache, miscellaneous pain and flu symptoms.

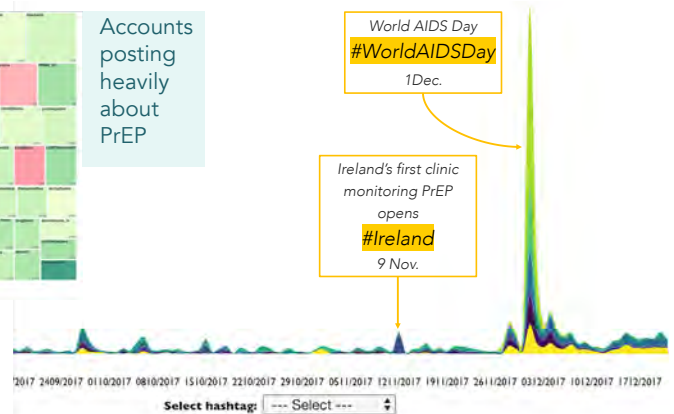
## Interactive Web Application for Trends Monitoring



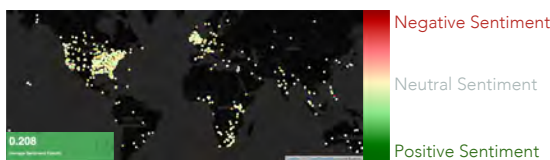
## Influencers



## Popular Hashtags Tracking



## Sentiment Analysis



## The geography of PrEP on Twitter

