Analysis of perception of infectious diseases on the internet in Poland

Andrzej Jarynowski[1,2], Alexander Semenov[3,4], Mikołaj Kamiński[5], Vitaly Belik[1]

[1] *System Modeling Group, Institute for Veterinary Epidemiology and Biostatistics, Freie Universität Berlin, Germany;* [2] *Interdisciplinary Research Institute, Wrocław, Poland;* [3] *University of Jyväskylä, Finland;* [4] *Herbert Wertheim School of Engineering, University of Florida FL, USA.* [5] *Individual Medical Practice, Oborniki, Poland.*

HARMONY
Novel tools for test evaluation and disease prevalence estimation

DFG

Background

We are going to quantify and analyze differences in the perception of infectious diseases on the Internet dependent on host type:

- zoonotic diseases: such as COVID-19 in animal reservoirs (mainly farming as Minks) and Avian Influenza;
- mainly human host diseases as COVID-19;
- animal only diseases as ASF.

Research Problem

Nowadays researchers widely use online social media data to investigate the behavioral and affective dynamics of the public during COVID-19 pandemics. However, non-English European languages are highly underrepresented and other zoonotic diseases are not covered at all.

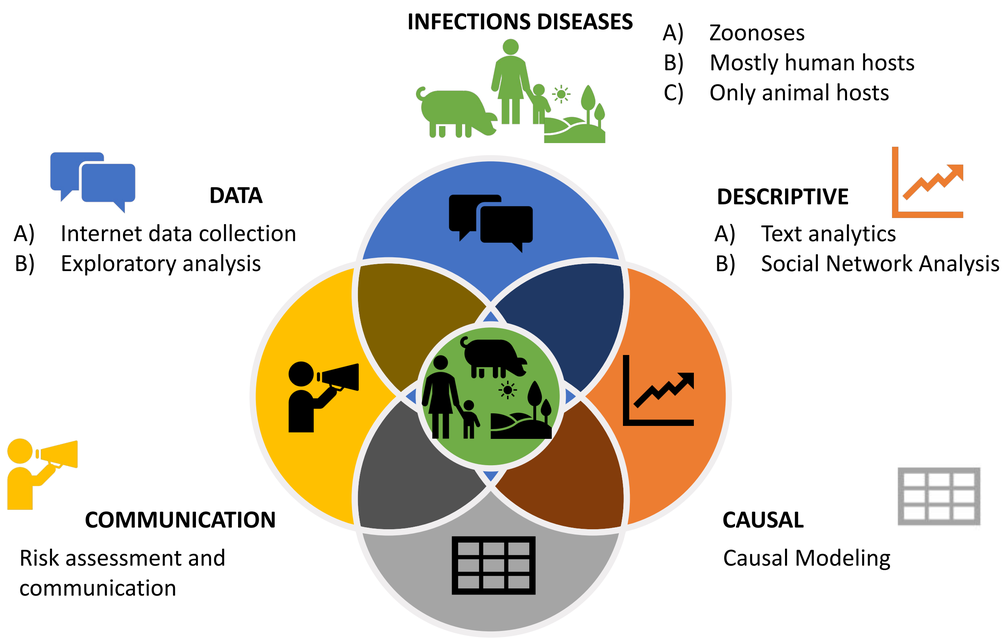


Diagram of project implementation.

Infosurvelliance in prediction/forecasting COVID-19 infection dynamics:

- worked far below expectation in Europe for publicly availabe dataset (i.e. Lampos, V et al. "Tracking COVID-19 using online search"(2021)),
- on the other seems to work with much more precised dataset in China (i.e. Guo, S et al. "Improving Google flu trends for COVID-19 estimates using Weibo posts."(2021)).
- High expectation, little predictive power (low digitalization rates and lack of availability of individual records in Western societies?)

Thus, this project will fill a crucial research gap by investigating and integrating wto major research questions among others:

- **CASE A** What is the hierarchy of social perception of infectious diseases and who are stakeholders interested in animal infections/ in zoonoses/ in human diseases?
- **CASE B** To what extent infoveillance could be used in estimating the burden of disease?

In Case Studies, we show a possibility of using infoveliance/infodemiology as an example of Google Trends, where we have measured weekly interest of a given keyword. Specificity, Sensitivity Estimation (Case A) and understating social (Case B) prevalence.

Outlook Conclusions

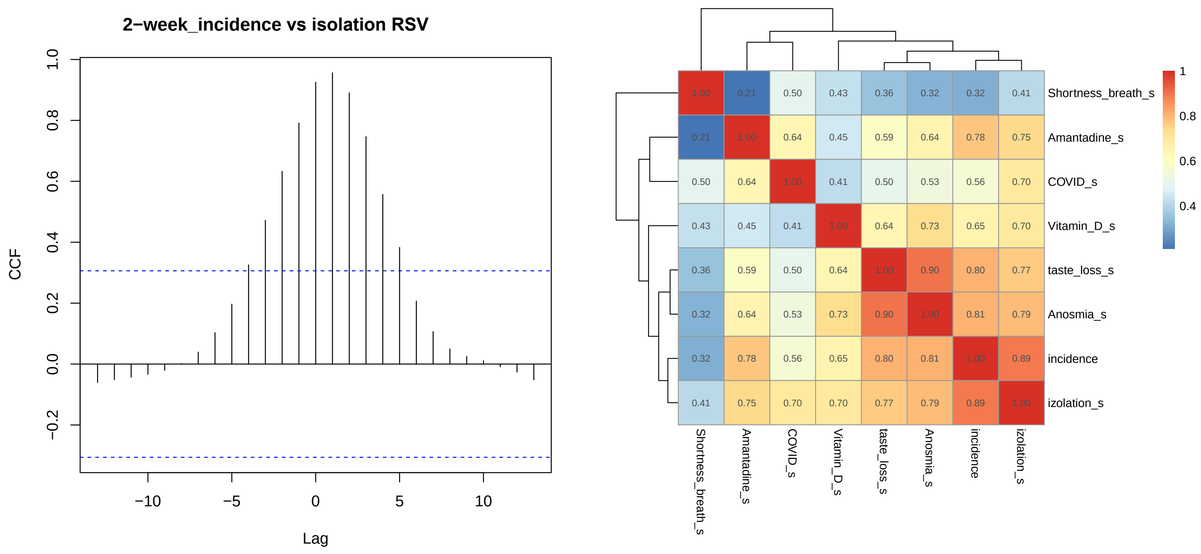
We have just started our project, but we could already state that:

- Infodemiology **is very useful** in understanding social dynamics during the pandemic by quantifying dynamics of interest (demand and supply of content) and discourse patterns. It plays a supplementary role to standard tools such as surveys and allows for the analysis in real time.
- Infosurveillance **could be useful** for public health decision makers in some specific areas such as predicting disease prevalence..

Thus, community-based surveillance in Internet Media formed during COVID-19 pandemic should be investigated further and its potential and experience can be considered in planning active surveillance in future (for animal diseases too).
The topic of infectious disease differs across humans (the highest interest among the general population with i.e. > million tweets monthly), zoonotic (average interest with some peak during local events as SARS-CoV-2 outbreaks among Minks with dozens of thousands Tweets monthly) and non-zoonotic (interest only in engaged group with maximally few thousands Tweets monthly) agents on the Internet. Assuming that the 2-week cumulative incidence is not an ground truth measurement of prevalence, we managed to achieve 96% sensitivity and 91% specificity using the syndromic index for epidemic trend detection. Moreover, keywords (syndromic or infectious disease explicitly) follows their fashion style life-cycles.

Assessment of Disease Se (Sensitivity)/Sp(Specificity)

Case study A)
Mathematical methods assessing disease prevalence based on available information are critically important to both the identification and control of pathogens in humans and animals (including zoonosis). However, prevalence estimates is extremely difficult.
We have analyzed relative search volumes (RSV from 0 to 100 for each keyword) weekly time series of chosen bag of keywords (taste loss, Anosmia, Amantadine, Shortness of breath, Vitamin D, COVID-19, isolation) as well as registered COVID-19 case notifications (since week 10 of the year 2020 till week 5 of 2022).



Left: Cross-correlations of "isolation" and 2-weekly incidence, Right: Spearman correlation matrix of weekly time series of popularity of particular phrases and 2-weekly incidence.

RSV and popularity of given keywords are highly correlated, i.e. popularity of isolation proceed incidence rates.
Thus, we build so called "Syndromic index" (Synd) and compare it with countrywide 14-day cumulative infection rate (as "gold standard" of prevalence estimation) for each week.
For each week our diagnostic measures (Incid: first derivative of cumulative incidence and Synd: first derivative of "Syndromic index") can take:

- + (is growing);
- - (is decreasing).

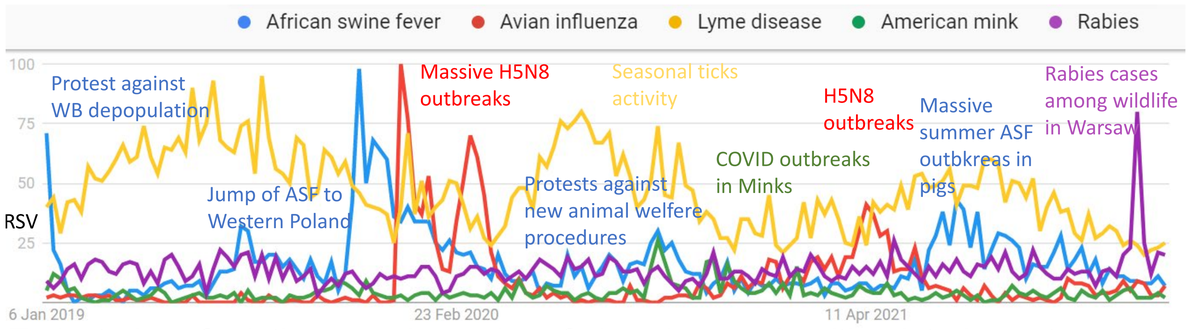
It's mean that if cumulative incidence is greater in a given than in previous week, then Incid is +, - if its smaller.
Se/Sp of "Sign of Syndromic index change" and "Sign of Incidence change" obtained using BLCM (Bayesian Latent Class Model) for given epidemic wave

wave	Incid+/Synd+	Incid+/Synd-	Incid-/Synd+	Incid-/Synd-	Se Incid	Se Synd	Sp Incid
all	44	17	7	25	0.86	0.71	0.99
first&sec	17	7	7	15	0.69	0.69	0.99
third&fourth	27	2	0	41	0.96	0.91	0.99

A given choice of keywords were closer to epidemiological prevalence proxy since second wave of infections. Even ad hoc selection of keywords gives a good agreement with incidence rate, especially when the information needs in first phases of disease has been satisfied.

Assessment of disease prevalence/importance

Case study B
Interest of internet users in selected infectious diseases (RSV) usually peak up during (re-) emergence of diseases in a new region and can be also driven by social-induced events such as street protest. Seasonal patters can be also detected.



Trajectories of relative search queries of given keywords with related events.

Acknowledgement

The authors (AJ, VB) acknowledge financial support by DFG 458528774 as well as presenting is possible due to Harmony - COST Action CA18208. We want to continue wornkin on a project which aims to perform an infodemiological/infoveillance study on Polish society as an example by analysis of the data collected from Internet media (demand - Google search - and supply - Tweets as well as news and Youtube comments) and discourse patterns using state-of-the art machine learning methods (i.e. quantitative media analysis of secondary data).