Connecting Information to Improve Detection in Event-Based Surveillance

David Buckeridge, MD PhD
Professor and Canada Research Chair
McGill University, Montreal, Canada

Context

EPI-AI project

- **Goal:** incorporate modern NLP into EBS and develop aberration detection methods that exploit modern NLP
- Using BioCaster as model system

Study of global variation detection in EBS

- **Goal:** Assess variation in frequency-based aberration detection using seasonal influenza as a proxy
- Data from HealthMap and EIOS

GPHIN workflow analysis

- **Goal:** Develop user experience for incorporating clustering of reports into EBS
- User stories, interface mock-ups, and algorithm development with GPHIN analysts

1. What *influences detection* in event-based surveillance?

2. How can we *exploit these factors* to improve detection?

1. What *influences detection* in event-based surveillance?

2. How can we *exploit these factors* to improve detection?

Evidence About Determinants of Detection

Reports

- Source
- Frequency
- Extraction

Disease

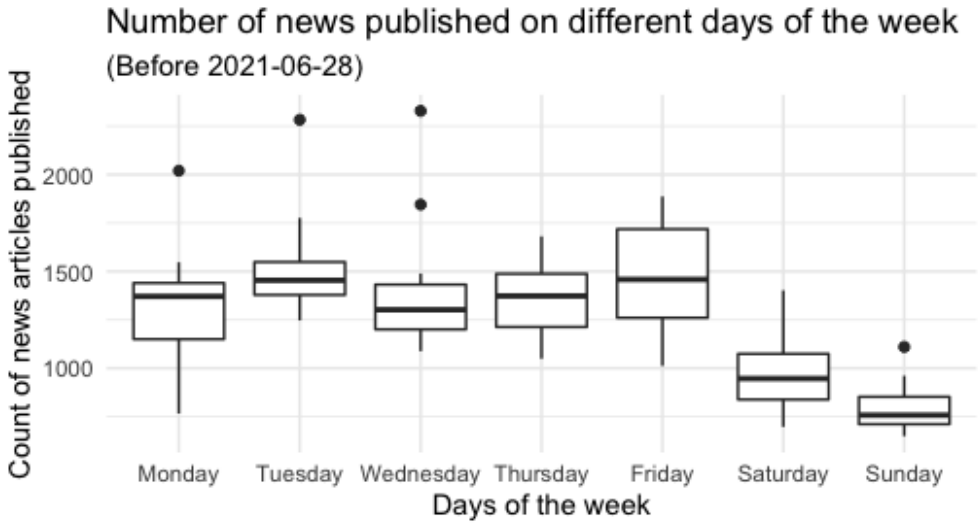
- Novelty
- Presentation

Context

- Health system
- Media landscape
- Resources

Reports

- Source
  - Type of media
  - Formal vs informal
  - Language
- Frequency
  - Day-of-week
  - Over time within an event
  - Crowding by other disease
- Extraction of information

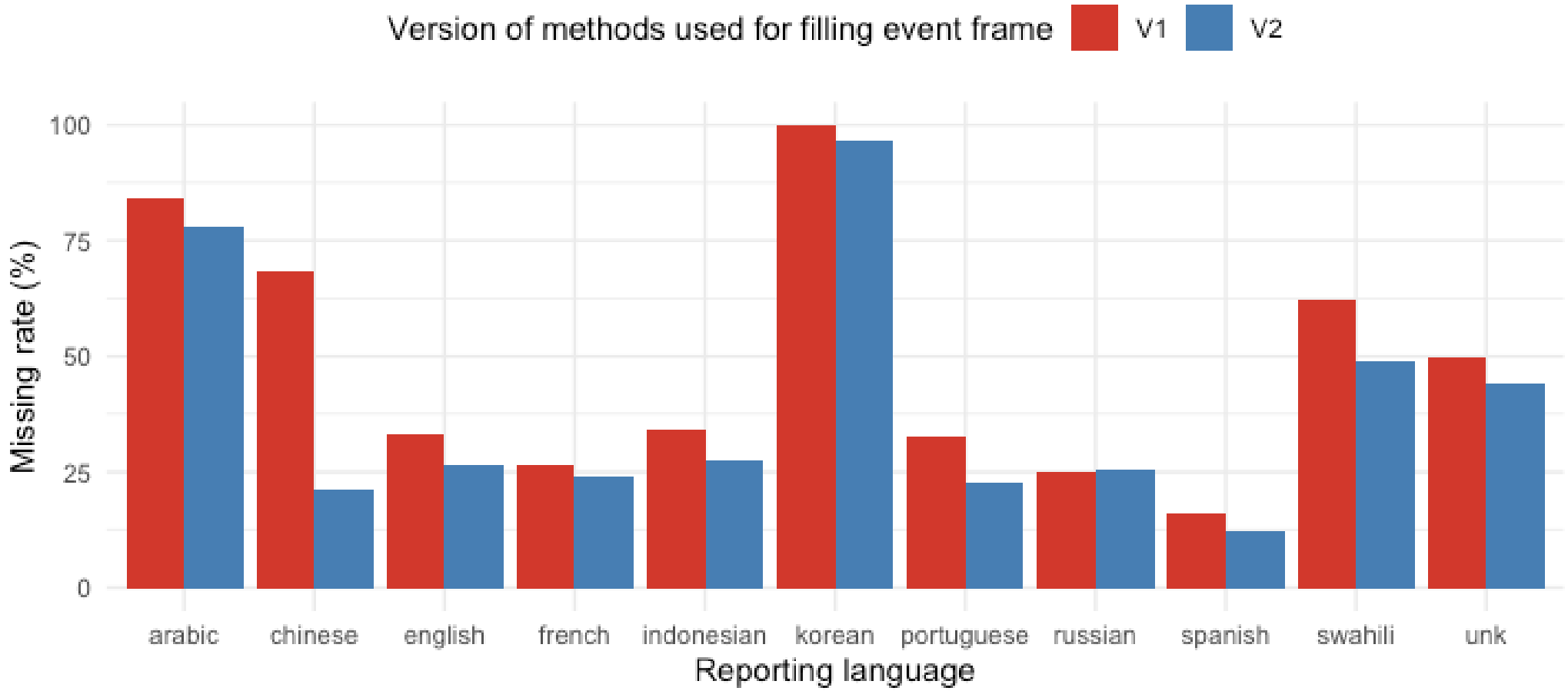


Scales D, Zelenev A, Brownstein JS. Quantifying the effect of media limitations on outbreak data in a global online web-crawling epidemic intelligence system, 2008-2011. Emerg Health Threats J. 2013 Nov 8;6:21621.

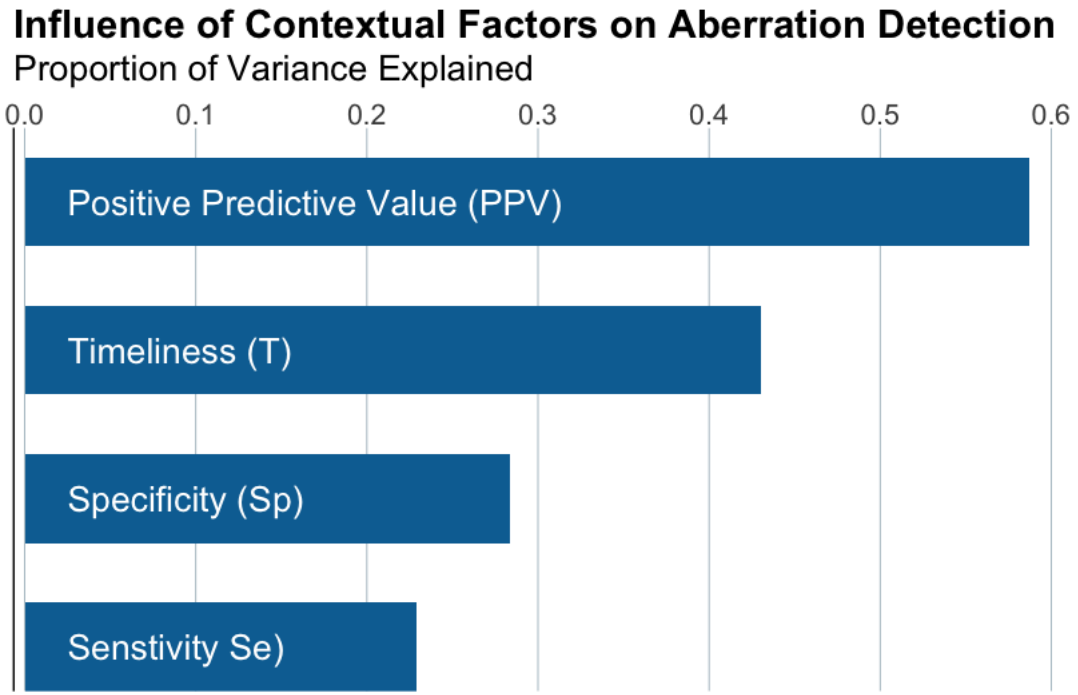
Smith KC et al. Understanding newsworthiness of an emerging pandemic: international newspaper coverage of the H1N1 outbreak. Influenza Other Respir Viruses. 2013 Sep;7(5):847-53.

Drivers of Emerging Infectious Disease Events as a Framework for Digital Detection. Olson SH et al. Emerg Infect Dis. 2015 Aug;21(8):1285-92.

Variation in Extraction by Report Language

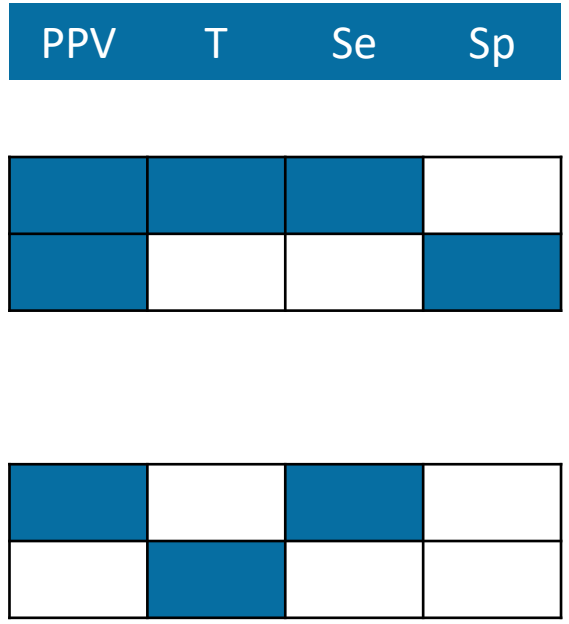


Context



Influence of Individual Factors

Significance of Factor for Predicting Metric



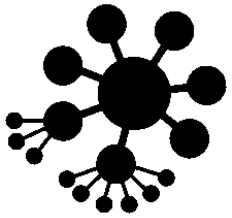
Ganser I, Thiébaut R, Buckeridge DL. Global Variations in Event-Based Surveillance for Disease Outbreak Detection: Time Series Analysis. JMIR Public Health Surveillance 2022 Oct 31;8(10):e36211.

1. What *influences detection* in event-based surveillance?

2. How can we *exploit these factors* to improve detection?

Strategies to Improve Detection

Account for background variation in frequency of reports



Connect information about same event across reports

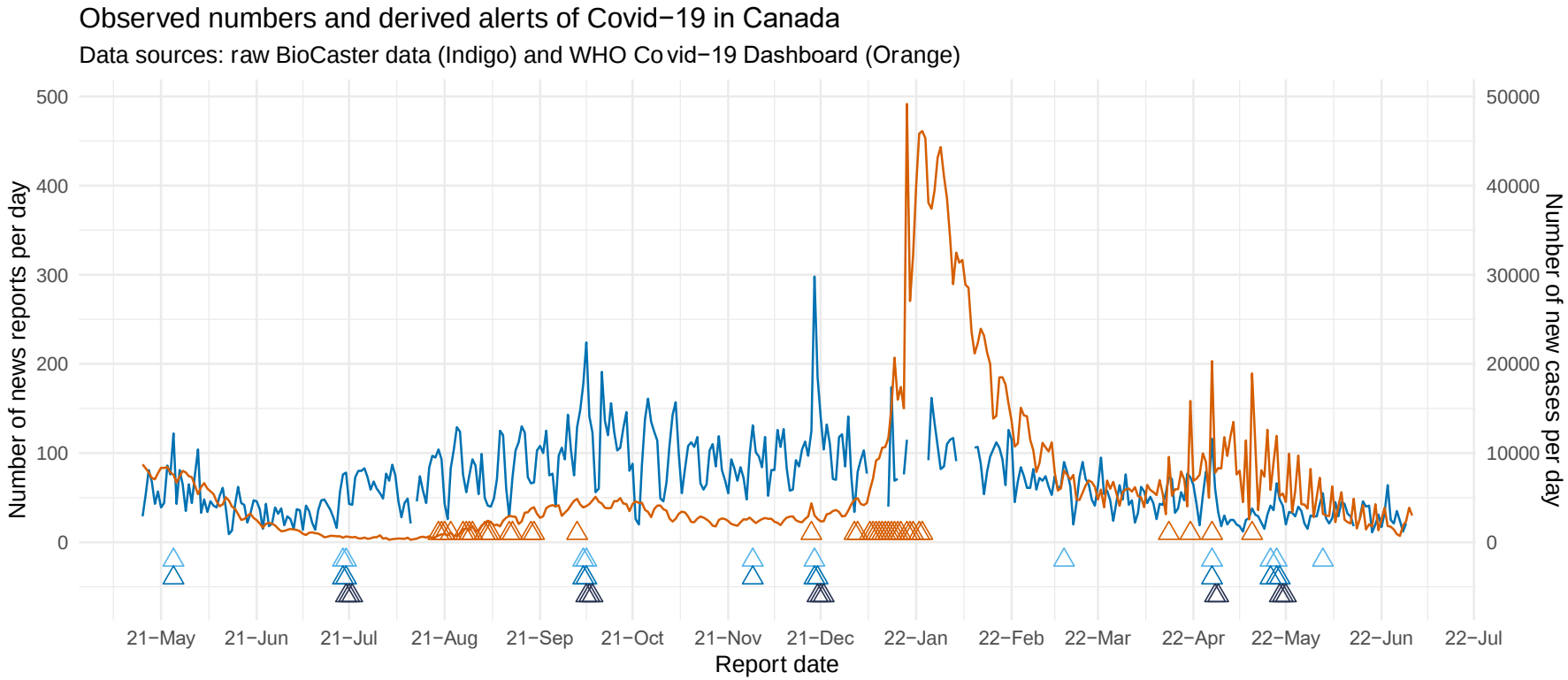
Flag known biases where there is no technical solution

Account for Background Variation in Frequency of Reports

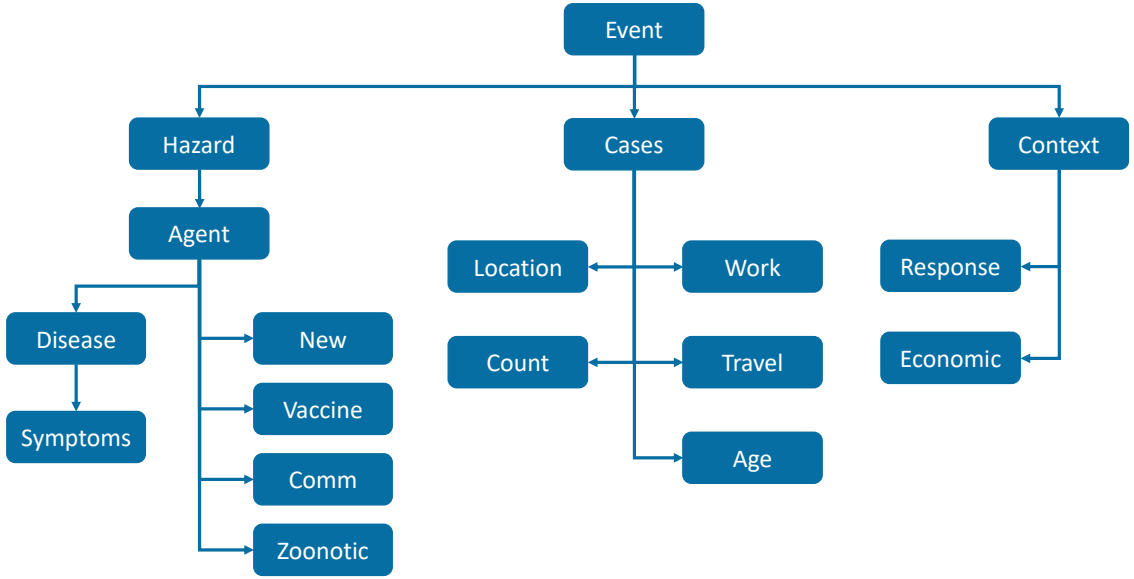
Source variation

Day-of-week

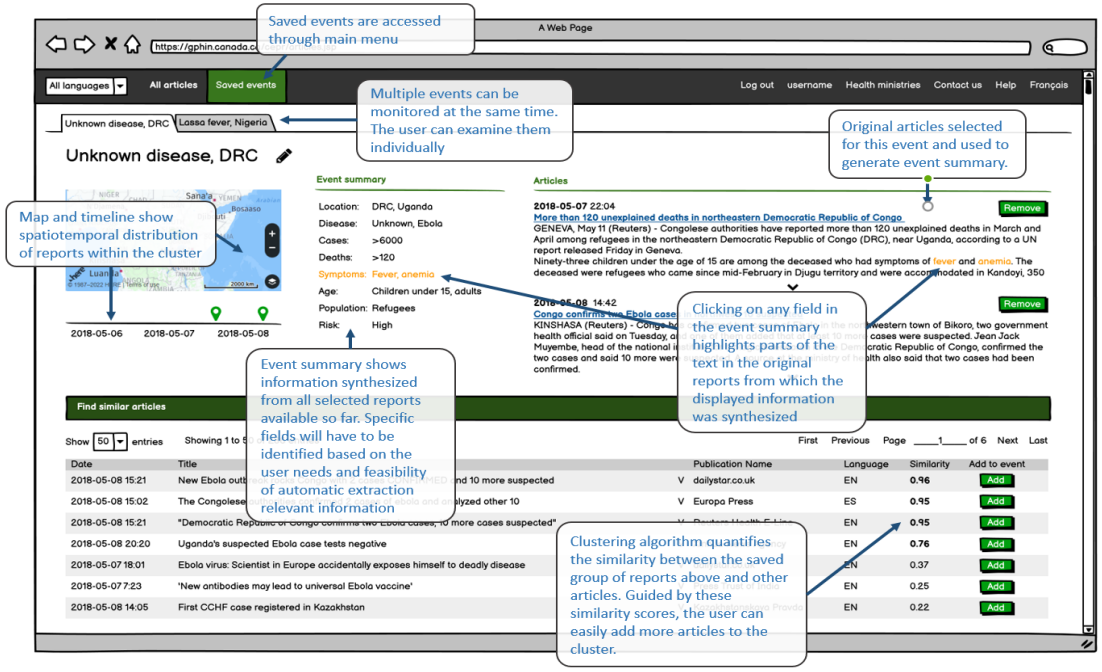
Seasonal



Connect Information Across Reports



Information model (ontology) for an event



Clustering reports and communicating information

Flag Potential Biases for Users

- Technical solutions are only available for some of the known factors that influence detection
- Informing EBS users of the unaddressed factors could limit the negative effect of these factors on detection
- Machine translation as an example
  - Accuracy of machine translation varies by language (as does NLP in general)
    - Influences completeness of information extraction
    - Influences available information and detection by geographical region
  - Accuracy for each language can be measured and communicated to users in the context of a report of event

Summary

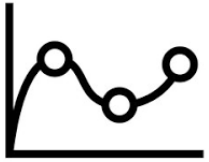
Multiple factors influence aberration detection in EBS

Reports

Disease

Context

These factors can be exploited to improve detection



Adjust for Background

Connect Reports

Flag Biases

Acknowledgements

- EPI-AI Project
  - Nigel Collier (Cambridge), Nicholas King (McGill), Maxime Polleri (McGIll), Zaiqiao Meng (Cambridge), Yannan Shen (PhD student), Meiru Zhang (Cambridge), Zihao Fu (Cambridge)
- Global Variation in Detection
  - Iris Ganser (McGill), Rodolphe Thiebault (Bordeaux), HealthMap, EIOS
- Clustering in GPHIN
  - Anya Okhmatovskaia, Yannan Shen, Erin Rees, Victoria Ng, GPHIN