FGAI4H-F-013-A01

Zanzibar, 3-5 September 2019

Source: TG-Outbreaks topic driver

Title: TDD update: Outbreaks (AI for Outbreak Detection)

Purpose: Discussion

Contact: Martina Fischer E-mail: fischerma@rki.de
Topic driver
Robert Koch Institute, Germany

Abstract: This PPT summarizes the activities for the TG-Outbreaks, for presentation and discussion during the meeting.

AI for Health
An ITU Focus Group
In collaboration with WHO

Topic Group:
Disease Outbreak Detection
Data & Benchmarking challenges

Dr. Martina Fischer
Robert Koch-Institute, Berlin, Germany

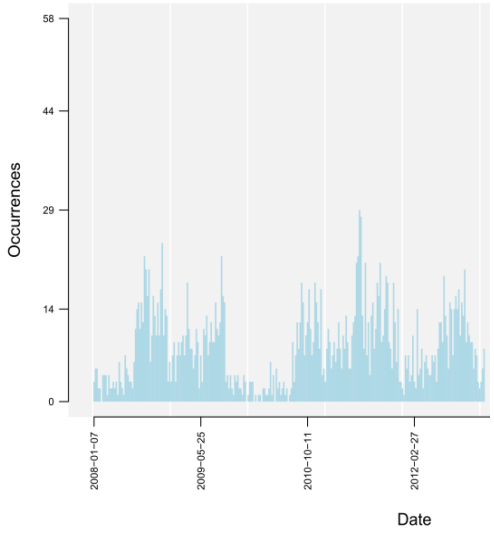
Zanzibar 3rd-5th Sept 2019

Background

- Infectious disease outbreaks pose a major risk to public health
- Early detection of outbreaks can prompt fast interventions
- Case data are collected by diverse surveillance systems
- AI algorithms can be applied to detect aberrant case numbers based on these data collections
- AI algorithms have the potential to increase the timeliness and accuracy of outbreak detection

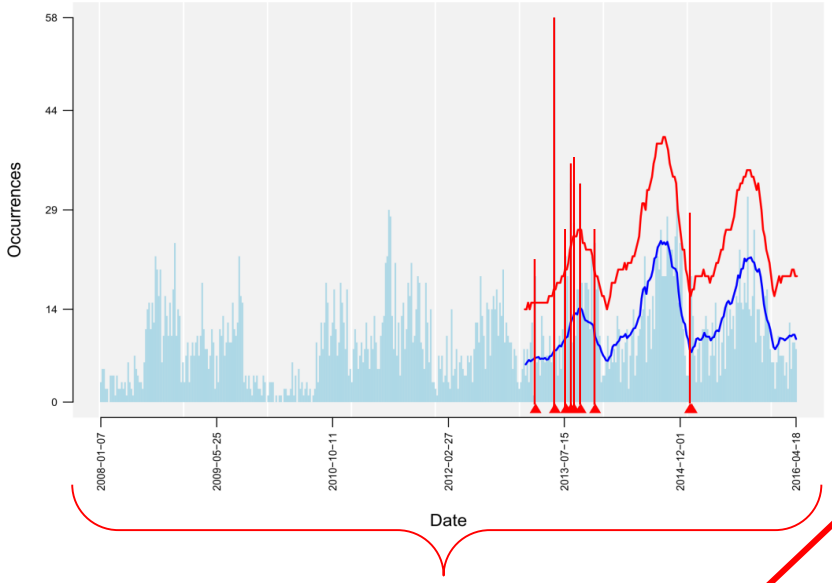
Objective

Data



(spatio-)temporal data
(reported cases of infection, symptom counts, etc.)

AI task



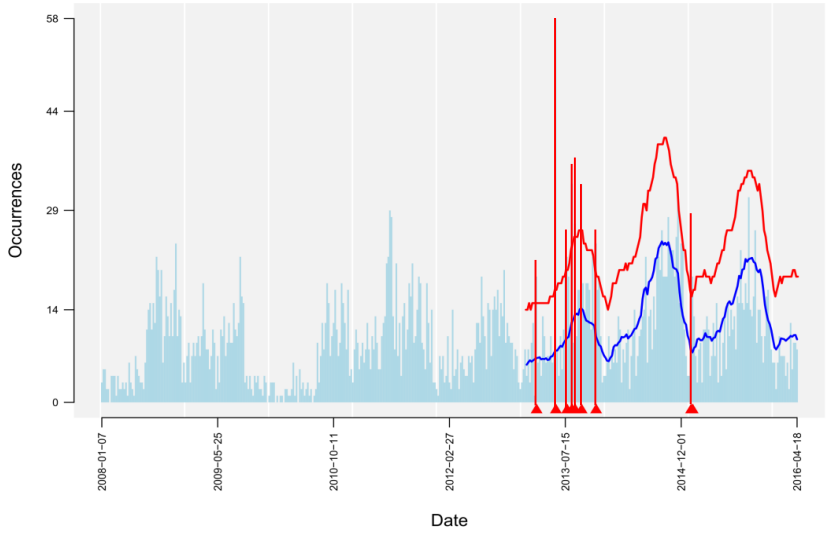
Detection of aberrant case numbers (▲)

warning signal

hint on potential outbreak event

Objective

AI task

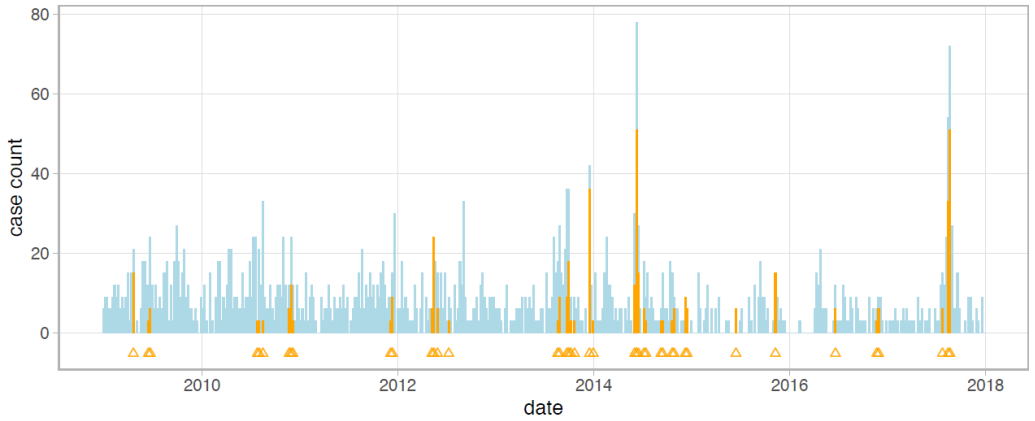


Detection of aberrant numbers (▲)

warning signal

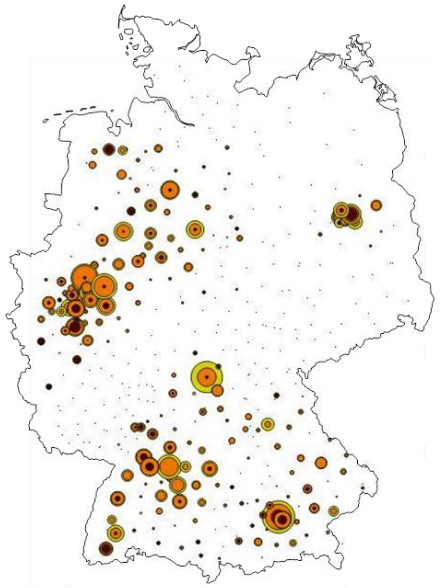
hint on potential outbreak event

Benchmarking

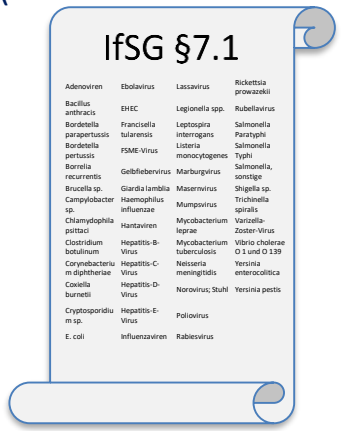


Test Data: confirmed outbreak labels ▲

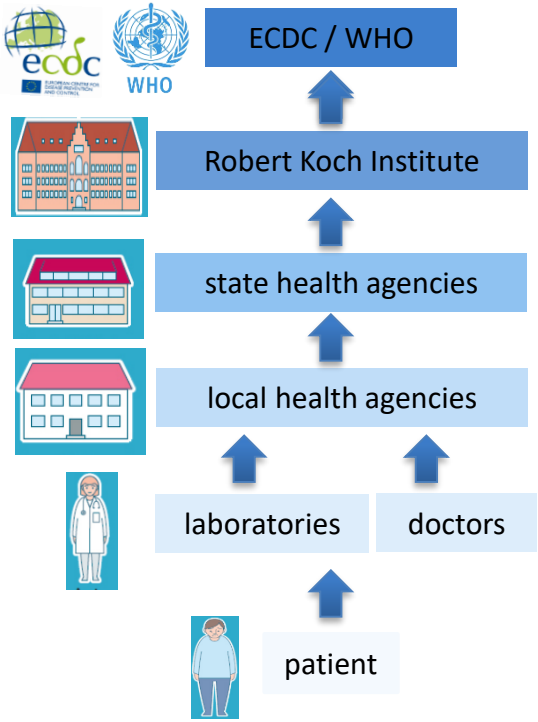
Example: German reporting system



Notifiable pathogens
(Infection Protection Act)



German Reporting System

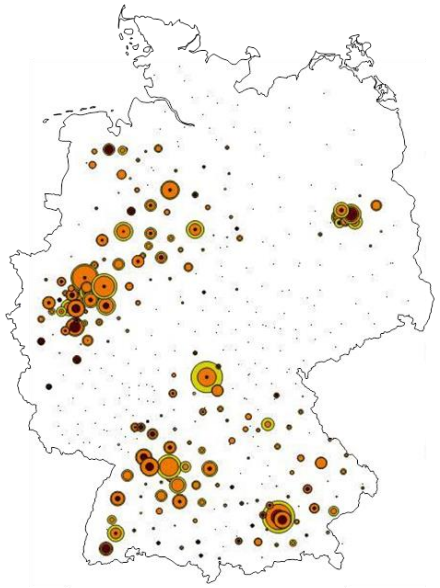


Aim:

Early outbreak detection for
fast *intervention*

Surveillance of > 80 pathogens
and > 400 counties
recording ~ 500.000 cases/year
detection ~ 20.000 outbreaks/year

Example: German reporting system

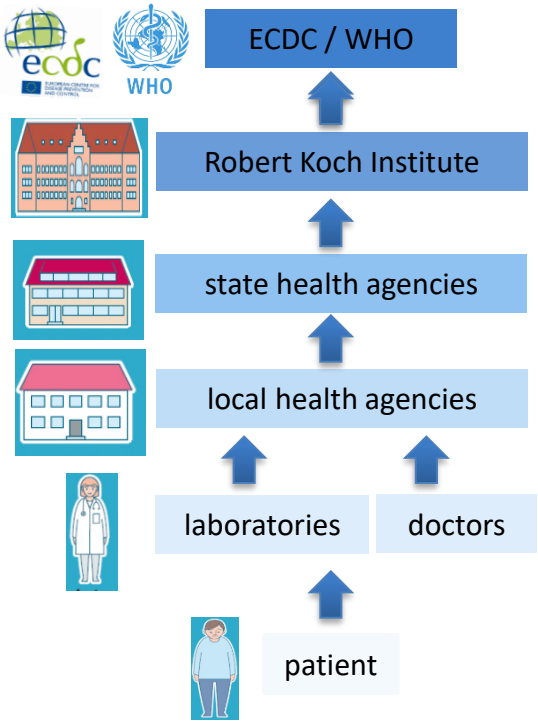


Notifiable pathogens
(Infection Protection Act)

IfSG §7.1

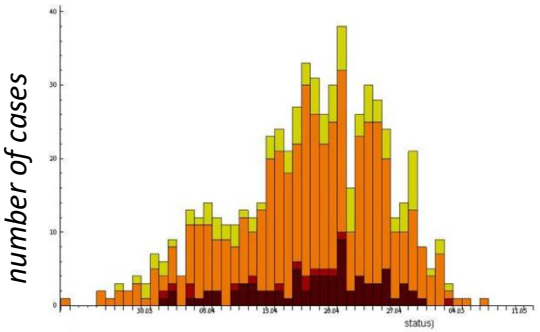
Adenoviren	Ebolavirus	Lassavirus	Rickettsia prowazekii
Bacillus anthracis	EHEC	Legionella spp.	Rubellavirus
Bordetella parapertussis	Francisella tularensis	Leptospira interrogans	Salmonella Paratyphi
Bordetella pertussis	FSME-Virus	Listeria monocytogenes	Salmonella Typhi
Borrelia recurrentis	Gelbfiebervirus	Marburgvirus	Salmonella, sonstige
Brucella sp.	Giardia lamblia	Masernvirus	Shigella sp.
Campylobacter sp.	Haemophilus influenzae	Mumpsvirus	Trichinella spiralis
Chlamydophila psittaci	Hantaviren	Mycobacterium leprae	Varizella-Zoster-Virus
Clostridium botulinum	Hepatitis-B-Virus	Mycobacterium tuberculosis	Vibrio cholerae O 1 und O 139
Corynebacterium diphtheriae	Hepatitis-C-Virus	Neisseria meningitidis	Yersinia enterocolitica
Coxiella burnetii	Hepatitis-D-Virus	Norovirus; Stuhl	Yersinia pestis
Cryptosporidium sp.	Hepatitis-E-Virus	Poliovirus	
E. coli	Influenzaviren	Rabiesvirus	

German Reporting System



Surveillance of > 80 pathogens
and > 400 counties
recording ~ 500.000 cases/year
detection ~ 20.000 outbreaks/year

Aggregate case data into timeseries:



Example: weekly case data of *Salmonella*

ROBERT KOCH INSTITUT

Signale-Bericht | Home | Krankheiten

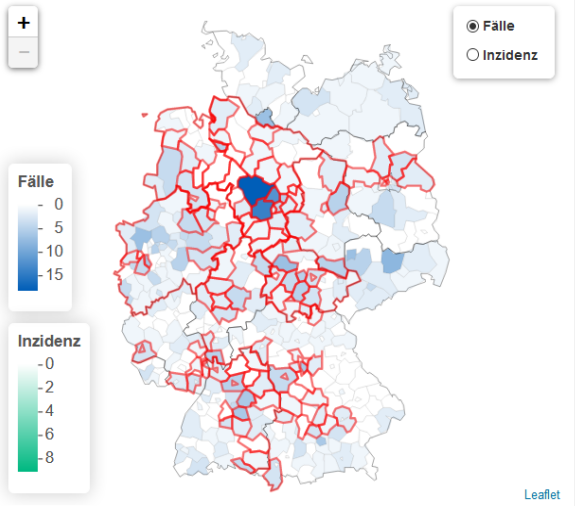
Für **Salmonellose (SAL-ST-GrB), KW33-38, Stichtag 23.09.2017**: Kein Signal wurde generiert für die 6 Subtypen, die mindestens 1 Fall haben.

CSV | Excel | Kopieren | Drucken

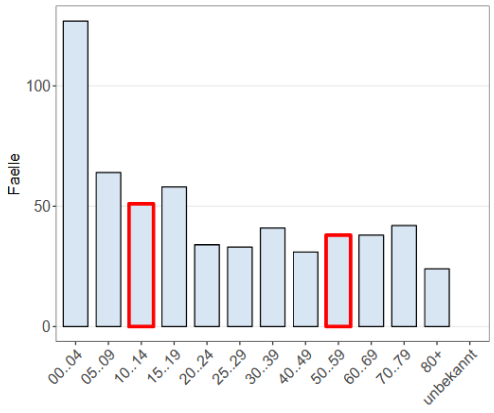
1 bis 6 von 6 Einträgen

Subtyp	gF	Woche 33				Woche 34				Woche 35				Woche 36				
		F	A	M	G	F	A	M	G	F	A	M	G	F	A	M	G	F
S.Typhimurium	429	65	6	59	94	79	7	60	93	75	0	62	95	80	3	60	95	95
Salmonella der Gruppe B	146	26	0	24	40	24	1	25	40	17	1	26	41	29	1	24	39	20
S.Typhimurium, monophasisch	6	1	0	2	8	1	0	3	8	2	0	2	8	2	0	2	8	0
Salmonella der...:4,12 H1:- H2:-	2	0	0			1	0			0	0			1	0			0
Salmonella der...:4,12 H1:i H2:-	1	0	0			1	0			0	0			0	0			0
Salmonella der...O:4,5 H1:- H2:-	1	0	0			0	0			0	0			0	0			0

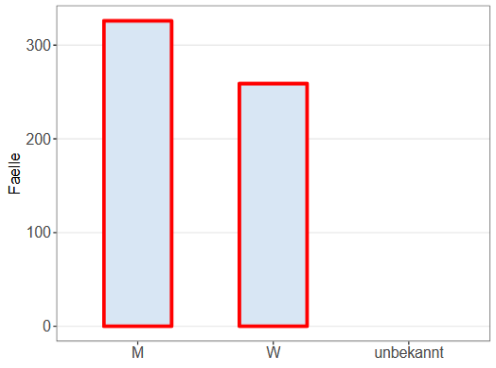
Fallkarte für SAL-ST-GrB, KW33-38, Stichtag 23.09.2017



Verteilung nach Altersgruppe für SAL-ST-GrB, KW33-38, Stichtag 23.09.2017



Verteilung nach Geschlecht für SAL-ST-GrB, KW33-38, Stichtag 23.09.2017



Data sources for outbreak detection algorithms

Different surveillance systems:

- national/ mandatory reporting systems
- (syndromic) surveillance systems:
  - Near-real-time syndromic surveillance:
    e.g. routine data from emergency departments and hospitals
    (e.g. *ESEG-project in Germany*)
  - Antibiotic Resistance Surveillance (*ARS project in Germany*)
  - …

potential other data sources:

- publicly available sources (e.g. meterological data,..)
- online data sources (wikipedia, google clicks, HealthTweets, Twitter, etc.)
- near real-time symptom data by self-assessment health apps
- …

Benchmarking challenges: (1) Data

- Definition label ‚Outbreak‘
  - Exact start/end-time point → size of outbreaks often unknown
  - Number of epidemiologically connected cases?
  - Confirmation by simple lab test or molecular analysis

- Label uncertainties
  - How to deal with unlabelled outbreak cases for benchmarking?
  - Minor peaks (with no confirmation): outbreak **or** random variation?

- Data diversity
  - Highly diverse outbreak data patterns of the <u>different pathogens</u>

  → Detection necessary per pathogen and per feature combinations (regions, risk groups,..)

- Test data
  - Needs to reflect national/international outbreak realities
  - Each country relies on individual national disease surveillance systems.

  → How do we optimally define a test set (undisclosed) to serve as a gold standard for benchmarking?

Benchmarking challenges: (2) Metrics

- Definition of epidemiologically relevant metrics for AI-algorithm evaluation:

  - Sensitivity & Specificity
    → How to strike the balance?
    - precise + early detection of outbreaks
    - minimize number of false alarms
    - accounting for case numbers: missed large outbreaks penalized more than missed small ones
    - precise outbreak size detection

  - Timeliness
    - by time (days/weeks) passed
    - by number of occurred cases

    before outbreak detection

  - Pathogen-specific
    - specific metrics according to pattern diversity?

- Metrics & test data need to be usable for evaluation comparison of AI and established statistical models for outbreak detection

Current members in topic group:

- Robert-Koch-Institute (National Public Health Institute Germany), involving members from different groups within the institute
  - Dr. Alexander Ullrich, Infection epidemiology, Signale team (RKI, Unit 31)
  - Dr. Stéphane Ghozzi, Infection epidemiology, Signale team (RKI, Unit 31)
  - Dr. Benedikt Zacher, Infection epidemiology, Nosocomial infections (RKI, Unit 37)
  - Dr. Bernhard Renard, Bioinformatics (RKI, Unit MF1)
  - Dr. Martina Fischer, Method development and research infrastructure (RKI, Dep. MF)
  - Dr. Janina Esins, Method development and research infrastructure (RKI, Dep. MF)

Interested in topic group:

Name	Affiliation
Philippe Verstraete, Dr. Giovanna Gutierrez	Co-founders of "Milan and Associates“, consultancy in epidemic intelligence, outbreak response, and more
Elaine Nsoesie, PhD	Assistant Professor of Global Health; School of Public Health; Boston University
Victor Akelo, MD, MPH	Co-lead for Child health and mortality Prevention Surveillance (CHAMPS) project in Kenya
Claire Jarashow, PhD, MPH	Chief, Epidemiology and Data Unit; Acute Communicable Disease Control; LA County Department of Public Health
Sharon Greene, PhD, MPH	Director, Data Analysis Unit; Bureau of Communicable Disease; NYC Department of Health and Mental Hygiene

Contacted by us for participation in topic group:

Name	Affiliation
Dr. Andre Charlett Alex J Elliot, PhD Roger Morbey, PhD	Public Health England; National Infection Service; * Statistics, Modelling and Economics * Real-time Syndromic Surveillance

Call for participation

Contributions by:

- Collecting **labelled test data**
  → data stream directly linked to outbreak labels (expert/lab confirmed) of high value.
- **AI models** and algorithms for outbreak detection
  → contributing to the development of a viable benchmarking framework
- **General support** on different aspects of this topic (data, methods, benchmarking, etc.)