CLASSIFYING EMERGENCY DEPARTMENT DATA TO IMPROVE SYNDROMIC SURVEILLANCE:

FROM MIXED DATA TYPES TO ICD CODES AND SYNDROMES

Birte Wagner
Robert Koch-Institut

15. JAHRESTAGUNG DER DGEPI
@DGEPI2020 - ONLINE

2. Oktober 2020

Background: ED data for syndromic surveillance

Data Sources

Application

Standardisation

Research

Epidemiological Classification

www.rki.de/sumo

Background: ED data for syndromic surveillance

ID	age	gender	date time	complaint presentation	complaint indicator	department	temperature	diagnosis	...
1	65	female	2019-01-01 16:00	Headache	Moderate pain	Neurology	37.0	R51	...
2	3	female	2019-01-01 17:00	Ear problems	Child temp. >38.5 °C	Paediatrics	39.3	J06	...
3	80	male	2019-01-02 06:00	Shortness of breath in adults	Respiratory infection	Internal medicine	38.0	J09	...

Background: defining influenza-like illness

Based on diagnostic information

specific and unspecific influenza/ILI ICD-10 codes:

e.g. J06, J09, J10, J11, J12, J18, J22

Based on pre-diagnostic information

e.g. temperature >= 38 AND

((complaint_presentation = ‘Shortness of breath in adults‘ AND

complaint_indicator = ‘Respiratory infection‘)

Background: surveillance of influenza-like illness

ID	age	gender	date time	complaint presentation	complaint indicator	department	temperature	diagnosis	ILI rule
1	65	female	2019-01-01 16:00	Headache	Moderate pain	Neurology	37.0	R51 →	FALSE
2	3	female	2019-01-01 17:00	Ear problems	Child temp. >38.5 °C	Paediatrics	39.3	J06 →	TRUE

Background: surveillance of influenza-like illness

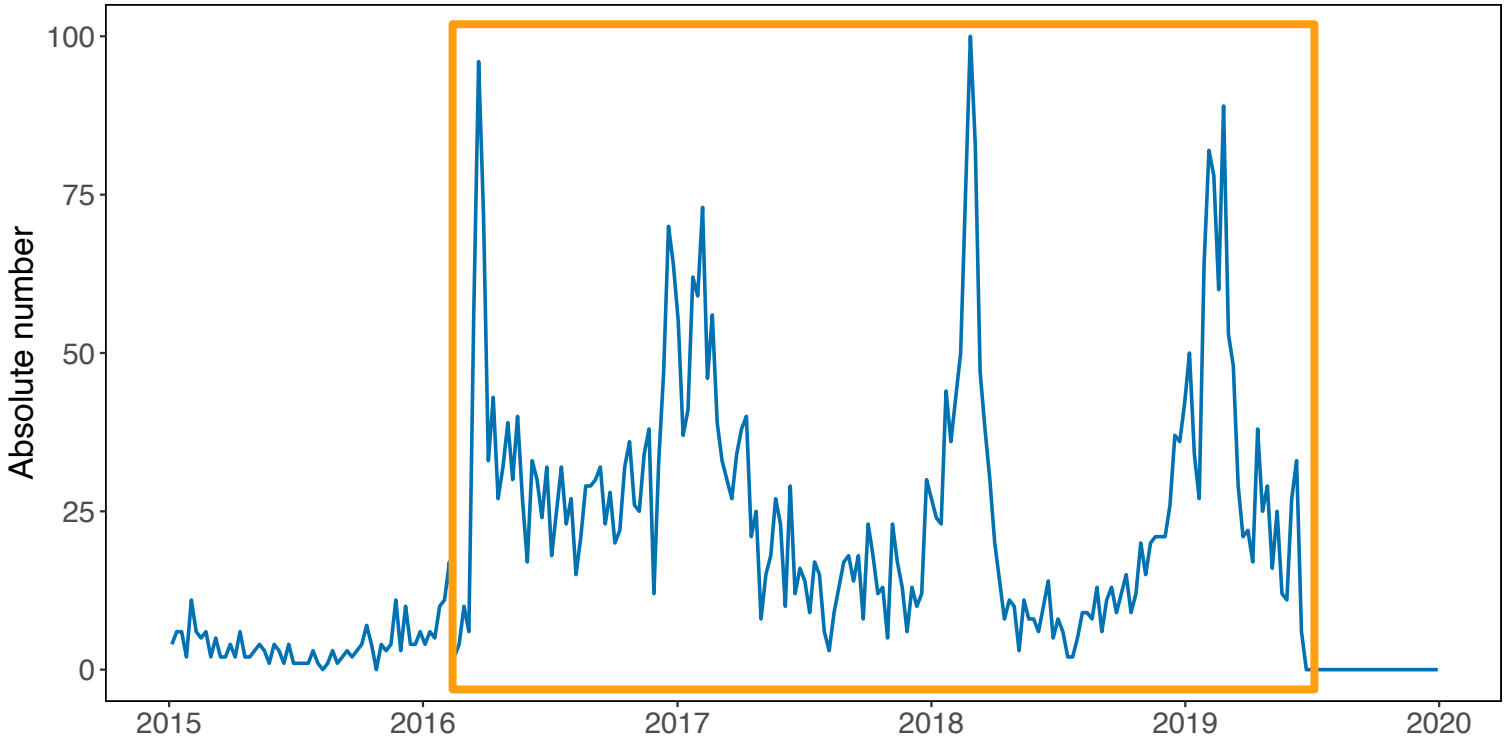


Figure 2: Absolute number of daily ILI cases in one emergency department, based on reference case definition (*internal reference*)

Background: surveillance of influenza-like illness

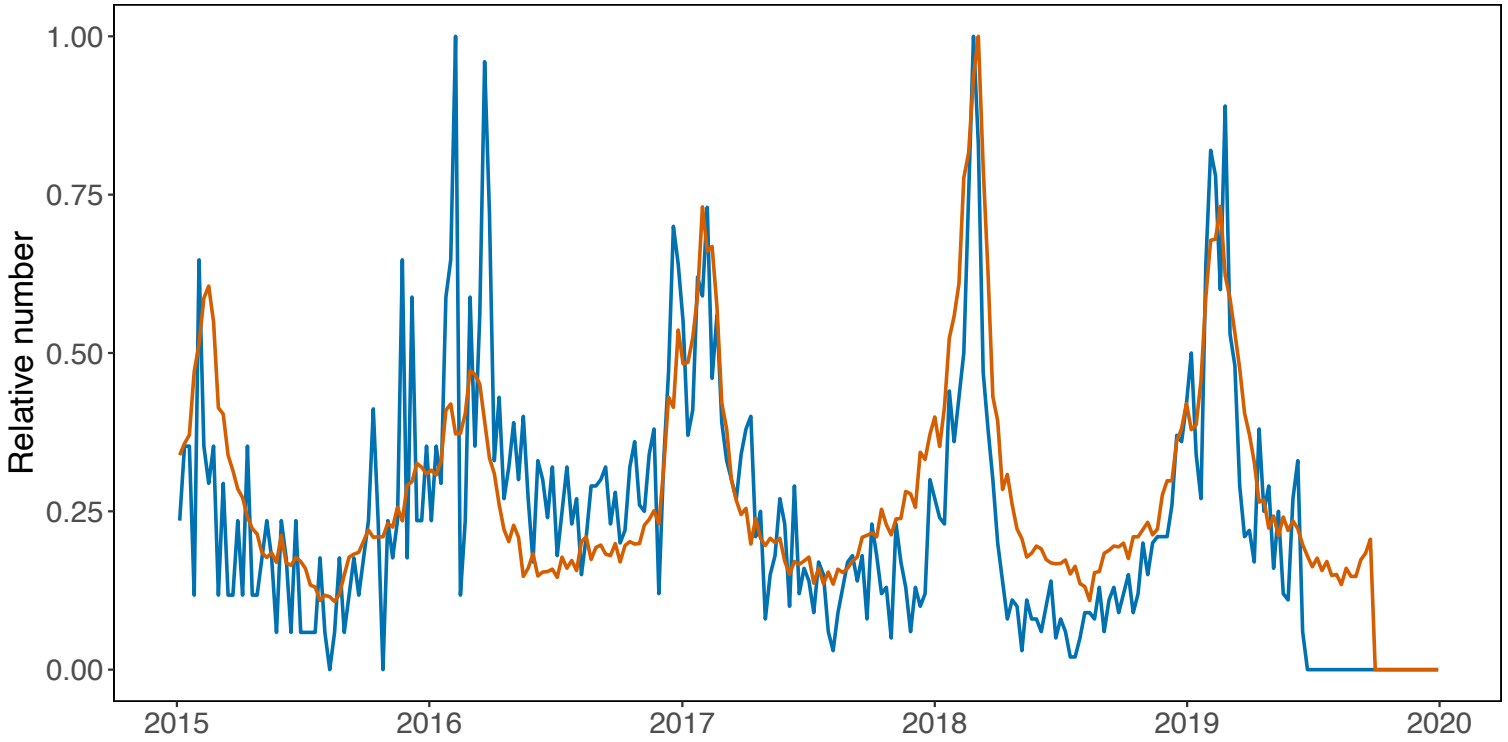


Figure 3: Relative number of daily ILI cases vs. daily SARI cases (*external reference*)
(relative to maximum of each time series)

Background: surveillance of influenza-like illness

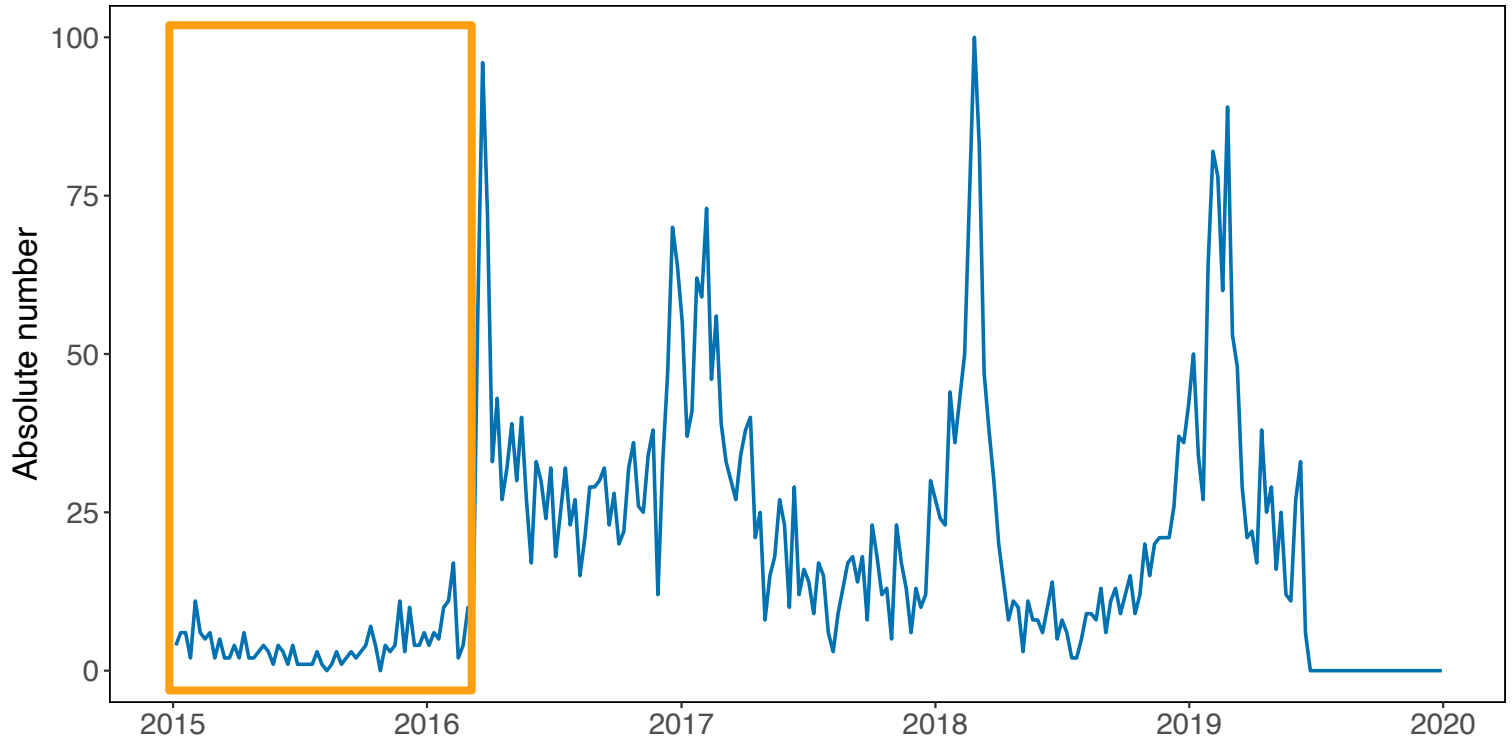


Figure 4: Absolute number of daily ILI cases in one emergency department, based on reference case definition (*internal reference*)

Problem: missings in diagnosis codes

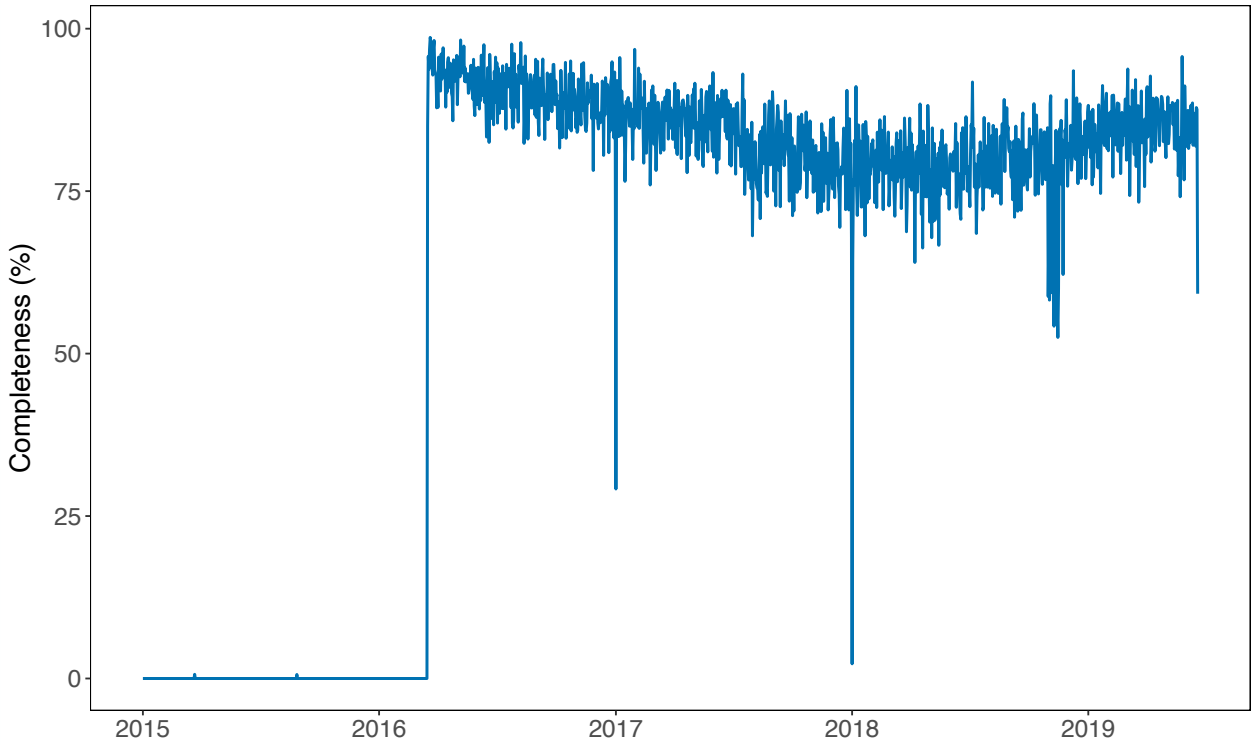


Figure 5: Amount of available ICD-10 diagnosis codes per day (%)

Problem: missings in diagnosis codes

ID	age	gender	date time	complaint presentation	complaint indicator	department	temperature	diagnosis	ILI rule
1	65	female	2019-01-01 16:00	Headache	Moderate pain	Neurology	37.0	R51	**FALSE**
2	3	female	2019-01-01 17:00	Ear problems	Child temp. >38.5 °C	Paediatrics	39.3	J06	**TRUE**
3	80	male	2019-01-02 06:00	Shortness of breath in adults	Respiratory infection	Internal medicine	38.0	NA →	?

Objective: impute diagnoses and predict syndromes

ID	age	gender	date time	complaint presentation	complaint indicator	department	temperature	diagnosis	ILI rule	ILI prob
1	65	female	2019-01-01 16:00	Headache	Moderate pain	Neurology	37.0	R51	FALSE	
2	3	female	2019-01-01 17:00	Ear problems	Child temp. >38.5 °C	Paediatrics	39.3	J06	TRUE	
3	80	male	2019-01-02 06:00	Shortness of breath in adults	Respiratory infection	Internal medicine	38.0	NA →	?	

1. impute diagnosis codes → 3-digit ICD-10 code (e.g. J10)

- multi-class naive Bayes classifier
- deep learning imputation approach DataWig[1]

[1] Biessmann, F., Salinas, D., Schelter, S., Schmidt, P., & Lange, D. (2018)

Objective: impute diagnoses and predict syndromes

ID	age	gender	date time	complaint presentation	complaint indicator	department	temperature	diagnosis	ILI rule	ILI prob
1	65	female	2019-01-01 16:00	Headache	Moderate pain	Neurology	37.0	R51	FALSE	
2	3	female	2019-01-01 17:00	Ear problems	Child temp. >38.5 °C	Paediatrics	39.3	J06	TRUE	
3	80	male	2019-01-02 06:00	Shortness of breath in adults	Respiratory infection	Internal medicine	38.0	J09 →	TRUE	

1. impute diagnosis codes → 3-digit ICD-10 code (e.g. J10)

- multi-class naive Bayes classifier
- deep learning imputation approach DataWig[1]

[1] Biessmann, F., Salinas, D., Schelter, S., Schmidt, P., & Lange, D. (2018)

Objective: impute diagnoses and predict syndromes

ID	age	gender	date time	complaint presentation	complaint indicator	department	temperature	diagnosis	ILI rule	ILI prob
1	65	female	2019-01-01 16:00	Headache	Moderate pain	Neurology	37.0	R51	FALSE	
2	3	female	2019-01-01 17:00	Ear problems	Child temp. >38.5 °C	Paediatrics	39.3	J06	TRUE	
3	80	male	2019-01-02 06:00	Shortness of breath in adults	Respiratory infection	Internal medicine	38.0	NA →		?

2. predict syndromes → ILI yes/no

- binary naive Bayes classifier
- logistic regression

Objective: impute diagnoses and predict syndromes

ID	age	gender	date time	complaint presentation	complaint indicator	department	temperature	diagnosis	ILI rule	ILI prob
1	65	female	2019-01-01 16:00	Headache	Moderate pain	Neurology	37.0	R51	FALSE	0.13
2	3	female	2019-01-01 17:00	Ear problems	Child temp. >38.5 °C	Paediatrics	39.3	J06	TRUE	0.87
3	80	male	2019-01-02 06:00	Shortness of breath in adults	Respiratory infection	Internal medicine	38.0	NA →		0.91

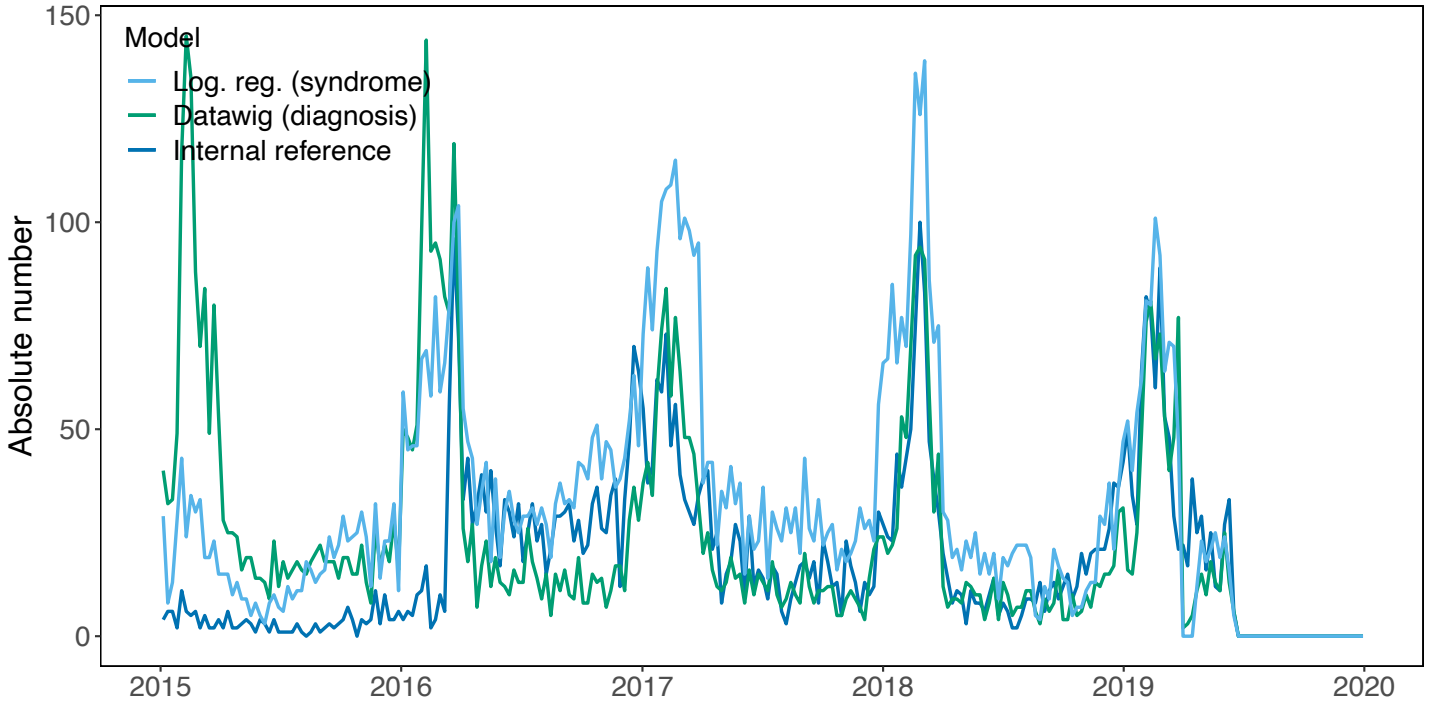
2. predict syndromes → ILI yes/no

- binary naive Bayes classifier
- logistic regression

Results: evaluation metrics

	Balanced accuracy	F1	Sensitivity (Recall)	Specificity	Precision
impute diagnosis codes					
naive Bayes	0.54	0.30	0.09	0.99	0.26
datawig	**0.55**	**0.34**	0.11	0.99	0.32
predict syndromes					
naive Bayes	0.74	**0.32**	0.54	0.94	0.23
logistic regression	**0.83**	**0.32**	0.77	0.89	0.20

Results: resulting cases follow the seasonal pattern of the internal reference cases



Correlation with internal reference cases (2016 – 2019)

Log. reg.: r = 0.793***
Datawig: r = 0.839***

*** p < 0.001

Figure 6: Weekly aggregated ILI cases by imputation and prediction models

Results: both approaches resemble the external data source of SARI cases

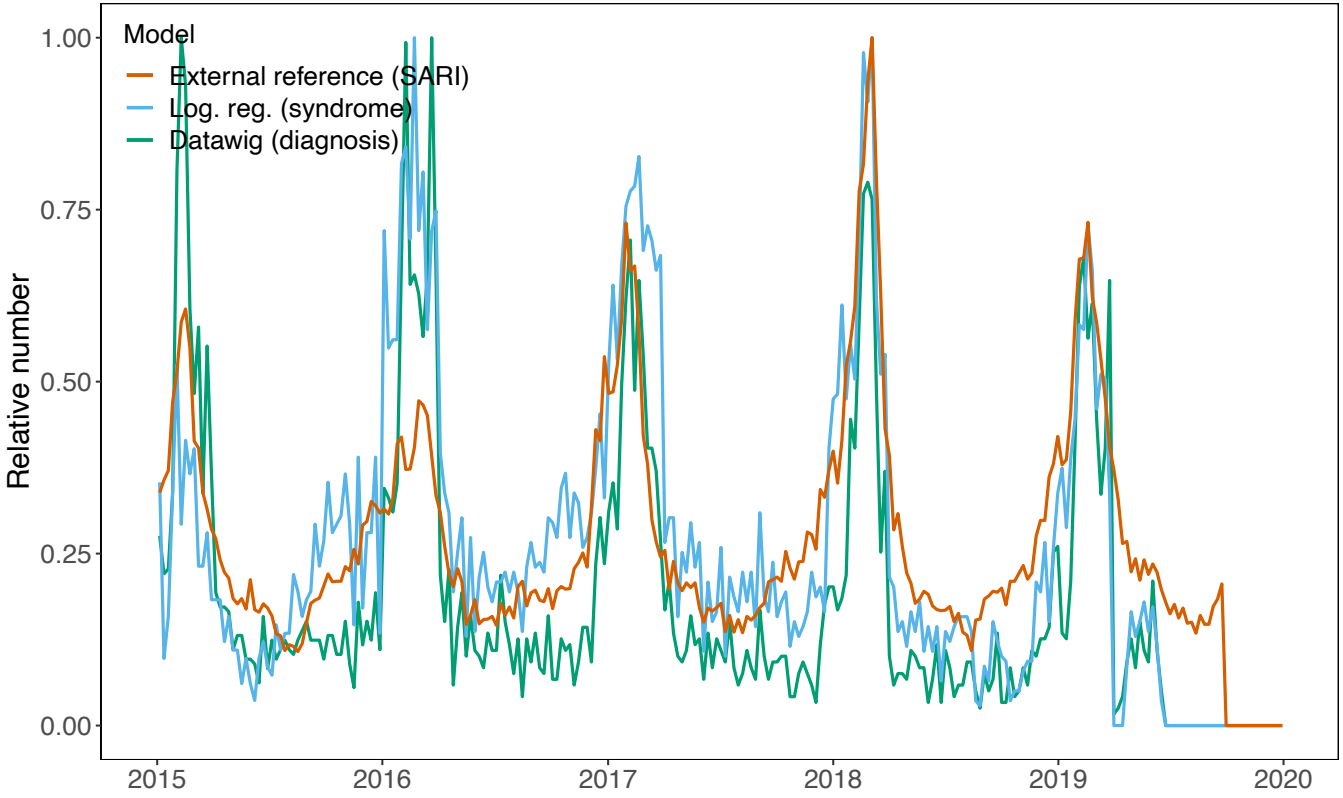


Figure 7: Relative number of daily ILI cases by models vs. daily SARI cases

Correlation with external reference cases (2016 – 2019)

LogReg: r = 0.821***
Datawig: r = 0.831***
Internal reference cases: r = 0.786***

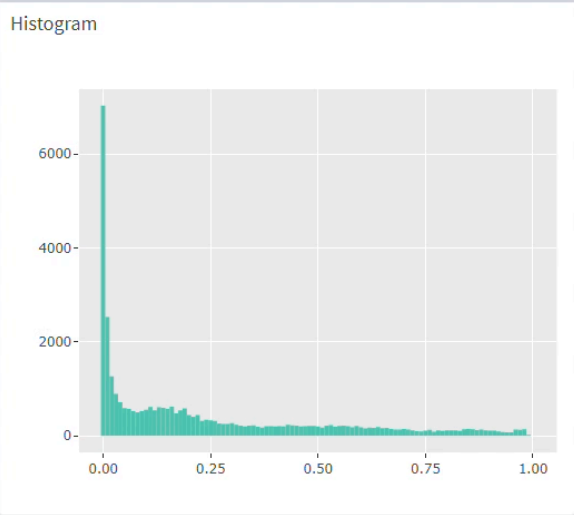
*** p < 0.001

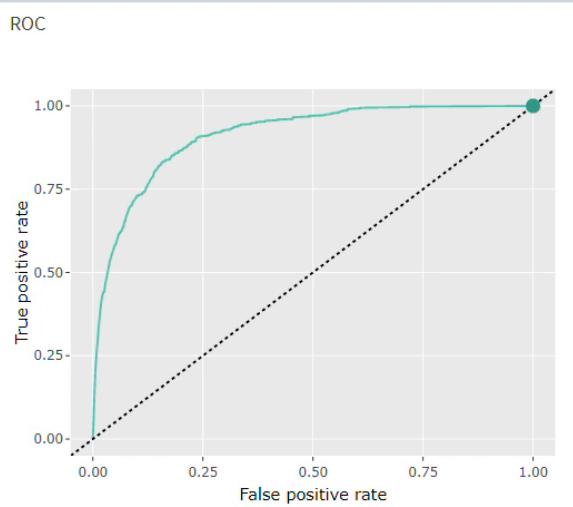
Conclusion

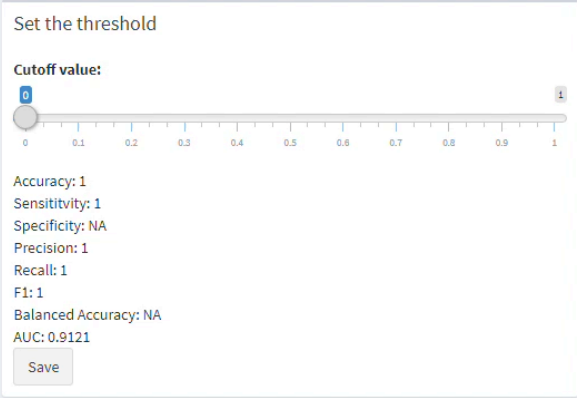
ID	age	gender	date time	complaint presentation	complaint indicator	department	temperature	diagnosis	ILI rule	ILI prob
3	80	male	2019-01-02 06:00	Shortness of breath in adults	Respiratory infection	Internal medicine	38.0	J09 →	TRUE	0.93
4	79	female	2019-01-02 05:00	Shortness of breath in adults	Respiratory infection	Internal medicine	38.0	NA	NA	0.87

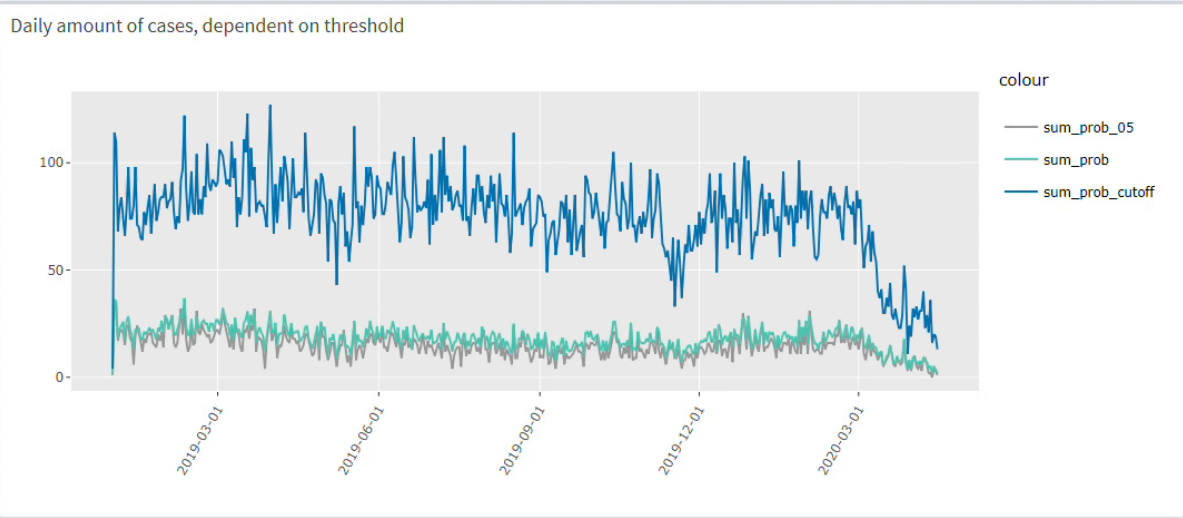
- imputing missing diagnosis codes allowes the labelling of ILI cases
- with the prediction models a data-driven case definition can be made for ILI cases, relying on symptoms, vital parameters and other available information

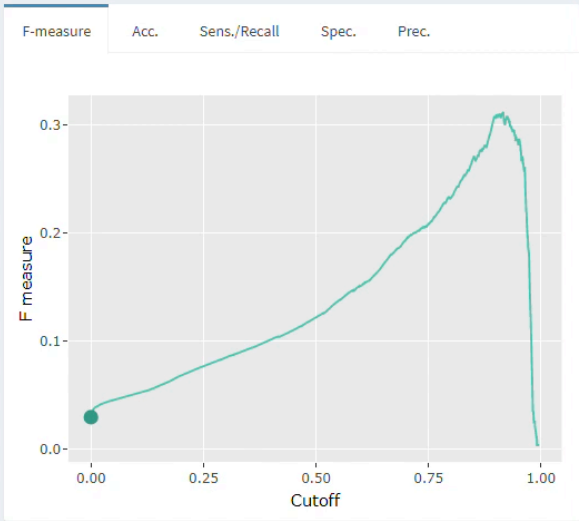
Outlook: evaluation of syndrome models











Contact us

Birte Wagner
wagnerb@rki.de

Robert Koch Institute
Dep. 3 – Infectious Disease Epidemiology

SUMO Team
Theresa Kocher
Alexander Ullrich
Linus Grabenhenrich
Madlen Schranz
Sonia Boender
Micheala Diercke

sumo@rki.de
www.rki.de/sumo