Master of Science

Public Health

Master Thesis

Syndromic surveillance using emergency department data for the monitoring of unspecific acute gastrointestinal infections

---

Author:

Madlen Schranz

224230

Supervisors:

PD Dr. Linus Grabenhenrich, MPH

Dr. Hendrik Wilking

Submission Date:

04.10.2019

Abstract

Background

Timely detection is a prerequisite for the successful prevention and management of gastrointestinal infections. Complimentary to already existing surveillance systems, syndromic surveillance can be used for that purpose. This study analyses data from a piloted syndromic surveillance system in Germany, aiming at the definition of syndromes and the exploration of the system's ability to monitor seasonal patterns and short-time aberrations of gastrointestinal infections.

Methods

Routinely collected data from emergency departments were analysed. Within this paper, two Syndromic Surveillance Health Indicators (SySHI) were created in order to combine complaint information. Syndromes were defined based on a combination of SySHI and a set of ICD-10 diagnoses. Identified cases were stratified by emergency department, age and isolation status. Time series were used to analyse seasonal patterns and combined with an algorithm to perform aberration detection.

Results

Between 2012 and 2019 data on 935,032 visits of ten emergency departments were analysed. Of those, 2.7% were identified as cases according to the Syndromic Surveillance Health Indicator. The syndrome definition “unspecific GI without bleeding” identified 3,329 and the syndrome definition “bloody diarrhoea” 589 cases. Time series analysis showed seasonal patterns with peaks in the winter months for the first syndrome. Exemplary for one hospital a total of six signals was created by the algorithm.

Conclusion

Emergency department data can be used to define appropriate syndromes for the detection of seasonal patterns and aberrations in case numbers. Syndromes incorporating information on chief complaint and diagnosis were able to identify seasonal fluctuations of gastrointestinal infections.

Eidesstattliche Erklärung

Hiermit erkläre ich eidesstattlich, dass ich die vorliegende Arbeit selbstständig und ohne unzulässige Hilfe Dritter angefertigt und keine anderen als die angegebenen Quellen und Hilfsmittel verwendet habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.

Die eingereichte schriftliche Fassung der Arbeit entspricht der auf dem elektronischen Speichermedium.

Weiterhin versichere ich, dass diese Arbeit weder in gleicher, noch in ähnlicher Form bereits als Leistungsnachweis an anderer Stelle vorgelegen hat.

Berlin, 02.10.2019

Table of Contents

List of Abbreviations

AKTIN	Aktionsbündnis zur Verbesserung der Kommunikations- und InformationsTechnik in der Intensiv- und Notfallmedizin
CEDIS-PCL	Canadian Emergency Department Information System – Presenting Complaint List
DALY	Disability-Adjusted Life Years
ESEG	Erkennung und Sicherung epidemischer Gefahrenlagen
EDIS	Emergency Department Information System
ED	emergency department
GI	gastrointestinal infection
IfSG	Infektionsschutzgesetz
ICD-10	International Classification of Disease – 10th Revision
MTS	Manchester Triage System
RKI	Robert Koch Institute
SySHI	Syndromic Surveillance Health Indicator
SynDef	syndrome definition
WHO	World Health Organization

Table of Tables

Table of Figures

1 Background

1.1 Gastrointestinal Infections

Globally, diarrhoeal diseases corresponded to 3.25% of all Disability-Adjusted Life Years (DALYs) in 2017. (1) In 2016, diarrhoea was responsible for an estimate of 1.6 billion deaths. The burden of gastrointestinal disease is especially high in low-income countries, with children being the population most affected. (2) According to the World Health Organisation (WHO), diarrhoeal disease is one of the leading causes of death in children younger than five years. (3)

The burden of infections of the gastrointestinal tract, even though much lower in high-income countries like Germany, is of public health concern. A cross-sectional study conducted between 2008 and 2009 identified the annual incidence of acute gastrointestinal illness in German adults to be 0.95 episodes per person year. This corresponds to around 64.9 million episodes yearly in Germany. (4) A smaller study conducted in 75 day-care centres with households including children younger than six years found the incidence of gastrointestinal infections per person year to be between 1.15 and 2.46, with the highest number in the group up to three years of age. (5)

Most commonly gastrointestinal infections present with clinical symptoms like diarrhoea, abdominal pain or vomiting. Causative agents can be viruses or bacteria and in rare cases protozoa. Especially in high-income countries, mortality due to gastrointestinal disease is low. Most infections are very well treatable, though usually only symptomatically. With the pathogen transmitted person-to-person through the contact with faeces or contaminated food or water, there are generally good prevention opportunities. Good personal hygiene including regular washing of hands and appropriate handling of food and adequate contact to infected people are important measures to prevent transmission. In low- and middle-income countries, improving sanitation, water supply and clean food (chains) are the key factors to prevent diarrhoeal disease besides individual prevention measures. (3)

Timely public health surveillance is a prerequisite for the identification of cases of gastrointestinal disease and the initiation of appropriate prevention measures. It can help identify risk populations, monitor seasonal clusters and outbreak situations and is an important factor in the guidance of public health action.

In Germany, the Infection Protection Act (IfSG) defines a list of diseases and pathogens which are mandatorily notified through a nationwide surveillance system. Physicians and laboratories are required to inform local health authorities about cases that fit certain case definitions. (6)

Pathogens causing gastrointestinal infections are amongst those with the highest yearly incidences in Germany looking only at disease notifiable by law (IfSG). In 2018, norovirus, campylobacter and rotavirus were responsible for more than 150,000 cases, corresponding to the highest incidences of notified infectious diseases besides influenza. In total, gastrointestinal infections accounted for 35.8% of all notified disease in 2018.(7)

Many gastrointestinal infections present with similar symptoms, usually too unspecific to differentiate pathogens clinically. Extensive diagnostics can be necessary to identify a specific pathogen. Especially in the case of mild progressions however, complete diagnostic work-ups are often omitted, leaving actually notifiable infections unknown. Furthermore, only specific pathogens are notifiable. These two aspects likely lead to an underrepresentation of the burden of gastrointestinal diseases within the German population. Another shortcoming of the monitoring of gastrointestinal infections through the notifiable disease system is the time lag in notifications. Physicians are required to notify local health authorities within 24 hours, laboratories have up to 2 weeks, depending on the type of notification. As some of the notifications are still conducted paper-based, case information can arrive at the responsible authority even later. Especially in outbreak situations, fast public health measures are required.

Complementary to the already existing surveillance system, syndromic surveillance could support the monitoring of the burden of specific (defined pathogen and/or disease entity) and unspecific gastrointestinal infections by providing a real-time image of the situation.

1.2 Syndromic Public Health Surveillance

Many surveillance systems use laboratory or clinically confirmed diagnostic information to detect health outcomes. In certain areas of the health system however, pre-diagnostic information is available that can aid in the monitoring and detection of health threats. Syndromic surveillance systems have the ability to collect, analyse and utilize health-related data, ideally in real-time, to enable timely public health action. (8) Lacking a standardised definition, syndromic surveillance can make use of various data sources and be applied for different use cases. In line with the WHO aims of public health surveillance (9), syndromic surveillance can serve as an early warning system to detect infectious disease outbreaks and seasonal patterns. Combined with information on specific exposures, syndromic surveillance can further be used to monitor health conditions (e.g. during extreme weather events) or to evaluate the impact of public health measures.

Countries like France, the UK and the USA have been successfully using nation-wide syndromic surveillance for several years, exploring different use cases for their systems. Besides applications related to communicable disease in the form of outbreak detection and seasonality monitoring (10, 11), syndromic surveillance also proved useful for the monitoring of non-communicable disease, mostly in combination with external exposures like heatwaves or cold-weather events (12, 13) and the evaluation of public health interventions, for example vaccination programs or the introduction of new traffic regulations. (14, 15)

In Germany, a nation-wide syndromic surveillance system is currently piloted. As one of the first points of contact with the health system, emergency department data were chosen as the basic data source, with the potential of including further data sources in the process. Within the piloting phase, routinely collected patient data from selected hospitals, including information about demographics, hospital administrative and health related information are available retrospectively starting from 2012. The piloting phase aims at the implementation of a near-real time data-provision and the definition of a long-term strategy for a continuously running surveillance system with nation-wide coverage.

1.3 Aims and Research Question

One major challenge while developing the new German syndromic surveillance system is the management of data originally collected for a different purpose (i.e. patient data collected for routine documentation in emergency departments). Useful information need to be extracted and combined into reasonable syndromes that are able to identify cases within the data. Previous research from other existing surveillance systems proved variables like chief complaints and diagnoses to be valuable information. Depending on the health entity that needs to be monitored and the construction of the system, other variables could be used in addition.

The main aim of this paper is the definition of syndromes (data patterns, case definitions) to identify unspecific (i.e. all) gastrointestinal infections, by selecting and appropriately combining routinely collected patient data. It furthermore explores the ability of syndromic surveillance using emergency department data to describe seasonal fluctuations of unspecific gastrointestinal infections as well as the potential for the application of algorithm-based signal detection for the identification of timely aberrations (e.g. outbreaks). The following work describes the exploration and analysis of the first available data from a newly piloted surveillance system.

2 Methods

2.1 Setting and Population

Within the ESEG (Erkennung und Sicherung epidemischer Gefahrenlagen) and AKTIN (Aktionsbündnis zur Verbesserung der Kommunikations- und InformationsTechnik in der Intensiv- und Notfallmedizin) projects, a network of emergency departments provided retrospective data. Hospitals were included based on their voluntary participation in one of the two projects, spread across Germany with the aim to cover a representative sample of the German population. All patient visits recorded in one of the partnering emergency departments between the hospital-individual start date and August 2019 were included, completeness of complaint information was the only criteria for inclusion into the presented analysis.

2.2 Data Source

A set of routinely collected, fully anonymised patient data was exported from different Emergency Department Information Systems (EDIS), transferred to a central database (separately for the two projects) and passed on to the Robert Koch Institute (RKI). Raw data elements in the EDISes contained different value sets and were stored using different coding standards. A standardised mapping strategy was used to project data into a generic data standard, which was developed at the RKI as part of the ESEG project, improving comparability and homogeneity of the data. Variables that diverged too much from the standard and could not be mapped automatically were manually transferred after reviewing the raw values.

2.3 Data Elements (Variables)

Each observation in the dataset corresponded to one emergency department visit and was given a unique identifier. Repeated visits from the same patient could not be distinguished in the data. Basic information on each observation included the identifier for the treating emergency department, day and time of the visit, and further administrative information. For each visit, age (in 5-year categories), gender and part of the residential postal code were collected. Furthermore, health-related information included triage severity, chief complaints, preliminary diagnosis and whether an isolation was mandated. Triage information was either reported through the Emergency Severity Index (16) or the Manchester Triage System (17), chief complaints through either the Canadian Emergency Department Information System – Presenting Complaint List (10) codes or a combination of Manchester Triage System (MTS) presentation diagrams and indicators. Diagnoses were provided as codes using the International Classification of Diseases 10th Revision (18), considering multiple diagnoses per visit.

The data standard used for the collection of routine emergency department data for syndromic surveillance was designed to provide information that is completely anonymised. It was reviewed and approved by the Data Protection Officer at Robert Koch Institute and by the Data Protection Officer of the federal state Hessen. Due to the anonymised nature of the data, an ethics vote was not necessary as disclosed by the ethics committee of the physician's chamber Hessen.

2.4 Data Processing (Syndrome Definition)

Information from the Canadian Emergency Department Information System – Presenting Complaint List (CEDIS-PCL) and the MTS triage were chosen as the primary source to identify cases of unspecific gastrointestinal infections. All participating hospitals provided data on either MTS or CEDIS-PCL. To properly handle the challenge of using two different sources of information for symptoms and complaints, a new class of derived variables, called Syndromic Surveillance Health Indicator (SySHI) was introduced, capturing core information from different sources in a standardized terminology. Those SySHI variables allowed the identification of a given health entity, regardless of the available information source. For the identification of unspecific gastrointestinal infections (GI), two health indicators were defined: "SySHI – diarrhoea, vomiting, nausea" and "SySHI – gastrointestinal bleeding".

The list of CEDIS-PCL codes was screened for complaints, indicating one of the two SySHI variables. The codes "254 – Diarrhoea" and "257 – Vomiting and/or nausea" where classified as "SySHI – diarrhoea, vomiting, nausea", the code "260 – Blood in stool/melena" as "SySHI – gastrointestinal bleeding". For the classification of MTS information, presentation diagrams that would be chosen for patients visiting with symptoms of unspecific gastrointestinal infections were identified. The diagrams "Abdominal pain in adults", "Abdominal pain in children", "Diarrhoea and vomiting" and "Gastrointestinal bleeding" were identified and corresponding indicators screened according to the order specified in the triage process. The classification was based on the assumption of a correctly performed triage process: indicators were screened in the given order with the first accurate being picked. When conditions positioned at the end of the decision tree were picked it was assumed that all positioned above were not applicable. The indicators “Persistent vomiting” and “vomiting” were, for example, allocated to the SySHI variable “diarrhoea, vomiting, nausea”, even though belonging to the presentation diagram “Gastrointestinal bleeding”. Following MTS triage rules, a patient with indications of gastrointestinal bleeding would have gotten assigned to a more severe indicator.

Including only those with complete information on either CEDIS-PCL complaint or MTS triage, visits flagged by one of the two syndromic surveillance health indicators were identified. In order to evaluate the information content of the SySHI variables, corresponding International Classification of Diseases (ICD-10) diagnoses and isolation information were screened, to see if the majority of visits flagged by one of the two health indicator categories were given diagnoses suitable for gastrointestinal infections or were isolated for a suspected infectious gastroenteritis. In case of diagnosis information, only codes that were given to more than one percent of all SySHI-identified cases were considered.

To further narrow down the number of flagged visits, ICD-10 diagnoses were included as a second information layer, leading to the final syndrome definitions (SynDef). For each of the Syndromic Surveillance Health Indicators, a single corresponding syndrome was defined based on both the SySHI and ICD-coded diagnoses: “SynDef – unspecific GI without bleeding” and “SynDef – bloody diarrhoea”. The list of diagnoses included in the German Modification of the ICD-10 catalogue was screened for those either describing a specific or unspecific gastrointestinal infection or symptoms that were consistent with gastroenteritis. Diagnoses from the groups “A00-A09 Intestinal infectious diseases”, “K92 Other diseases of the digestive system”, “P54 Other neonatal haemorrhages” and “R10 Abdominal and pelvic pain” as well as the ICD-10 codes “P92.0 Vomiting in newborn” and “R11 Nausea and vomiting” were considered relevant for the syndrome definitions. Furthermore, cases identified by the syndrome definition receiving diagnoses form the groups “A00-A09” and “K92” were analysed separately. This leads to the classification of cases into three layers of information depth:

- visits with a certain matching chief complaint (SySHI)

- visits with a certain matching chief complaint, including only those with a GI-related ICD-10 code (SynDef)
- visits identified by SynDef, including only those with a specific ICD-10 code (from the groups “A00-A09” or “K92”)

Distinction between the domains “unspecific GI without bleeding” and “bloody diarrhoea” was made solely based on the categorisation into one of the two SySHI categories, the assigned ICD-10 code was not reviewed to that effect.

2.5 Statistics including Time Series Analyses

For the report of the coverage of the collected emergency department visits, a frequency measure relating the number of visits per year and per 1,000 inhabitants was used, in two-digit precision area code strata. Reference data for the number of inhabitants was drawn from the 2011 census, according to the German Federal Statistical Office. (19)

Descriptive reporting of cases identified by one of the SySHI categories was stratified by emergency department, age group, gender, triage severity and isolation reason.

For the identification of seasonal fluctuations of gastrointestinal infections, time series of the relative monthly frequencies were analysed. Visual trend analysis support used the “loess” method with a span of 0.3 within the R package ggplot2. (20) The analysed interval for time series was restricted from May 2016 to July 2019, including only weeks with more than 500 reported ICD-10 coded diagnoses over all eligible hospitals. To account for the different starting dates of data provision (some hospitals started reporting only in 2017), relative frequencies were used. For the SySHI variables, total visits per month were used as a denominator. In case of the syndrome definitions and the specific diagnoses, all visits receiving at least one ICD-10 code were used as a denominator.

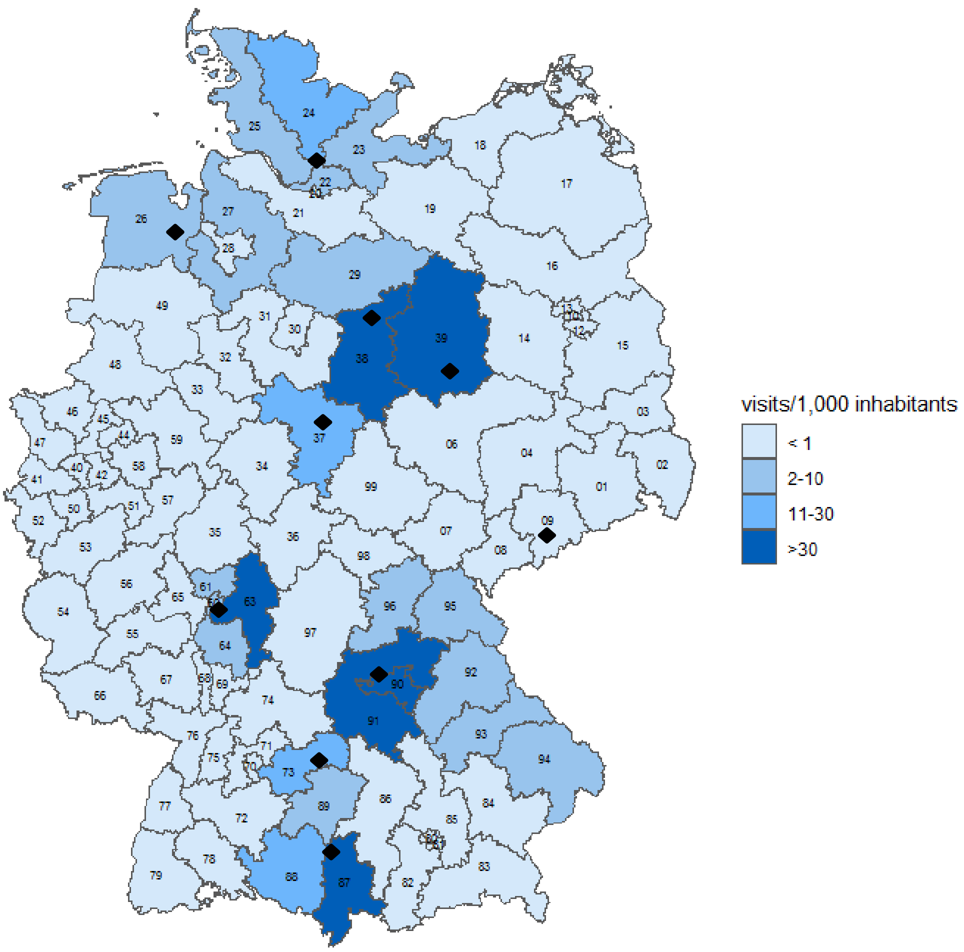
To examine aberrations in the relative frequency of syndromes, the “earsC” algorithm of the R package “surveillance” was used. The algorithm uses a customisable number of time points as baseline, which is especially helpful for data that do not cover large timespans. (21) Signal detection was performed using the C1 method on weekly frequencies, with a baseline set to 11.

For the selection of an appropriate timeframe providing enough data, the same criteria as for the seasonality analyses were applied. As GI outbreaks are often limited to smaller areas, the aberration detection algorithm was applied to each hospital separately. Emergency departments not providing data from the set timeframe or counting less than 10.000 cases in total were not included. Aberration detection was explored separately for “GI unspecific without bleeding” and “bloody diarrhoea”.

3 Results

3.1 Population

Figure 1 – Number of emergency department visits in 2018 with valid presenting complaint data per 1,000 inhabitants, by two-digit precision area code (as labels on the map);
dots marking locations of included emergency departments.



Between February 2012 and August 2019 data on 1,243,598 visits in 12 emergency departments was available. Two emergency departments, though partnering in one of the projects, were fully excluded for not reporting any complaint information (valid CEDIS-PCL code or MTS triage). This led to a final analysis sample of 935,032 (75.2%) visits from ten emergency departments with varying regional coverage throughout Germany (Figure 1).

25.3% (236,355) received at least one ICD-10-coded diagnosis. For five out of ten emergency departments (EDs) diagnosis coding completeness lay between 19.1% and 55.2%, the other EDs did not report diagnoses at all. Triage severity was reported for 90.6%, stratified by hospital between 10.8% and 100%. Information on a mandated isolation was available for 54.0% of visits, with complete reporting in four hospitals, completeness between 33.0% and 55.9% in two EDs and four emergency departments providing under 10% completeness for isolation. Information on age, gender and postal code was reported for over 99% of visits in all emergency departments. Stratified by hospital, between 33 and 142 people were treated on average every day. The highest number of visits occurred between 10 and 11 a.m., the lowest between 3 and 4 a.m..

3.2 Cases (Gastrointestinal Infections)

Visits presenting with chief complaint information indicating gastrointestinal infections were defined as cases. The majority of case-defining indicators chosen from the MTS catalogue came from the presentation diagram “Diarrhoea and vomiting”. Most cases classified into the “gastrointestinal bleeding” category, received indicators from the diagram “Gastrointestinal bleeding”. Within the selected CEDIS-PCL codes, around half of the “SySHI – diarrhoea, vomiting, nausea” cases received the code “254 – Diarrhoea”, the other half was identified through the code “257 – vomiting and/or nausea”. (Table 1)

24,879 (2.7%) visits were classified as cases by one of the two Syndromic Surveillance Health Indicators. 19,436 cases met the criteria of “SySHI – diarrhoea, vomiting, nausea” and 5,435 those of “SySHI – gastrointestinal bleeding”. The frequency of gastrointestinal infections was similar across all hospitals: around 2% (1.2-2.5, by ED) were classified as “diarrhoea, vomiting, nausea”, and less than 1% (0.3-0.9, by ED) presented with “gastrointestinal bleeding”. The highest frequency of patients with “diarrhoea, vomiting, nausea” could be seen in the age group zero to four years (6.9%). (Table 2)

Table 1 – Number of visits with MTS triage and CEDIS-PCL chief complaint categories used to define Syndromic Surveillance Health Indicators (SySHI).

MTS Presentation Diagram	MTS Indicator	Triage*	SySHI diarrhoea, vomiting, nausea n	SySHI gastrointestinal bleeding n
Abdominal pain in adults				
	Passing fresh or altered blood PR	2	-	46
	Black or redcurrant stools	3	-	347
	Persistent vomiting	3	310	-
	Vomiting	4	691	-
Abdominal pain in children				
	Black or redcurrant stools	3	-	50
	Persistent vomiting	3	3	-
	Vomiting	4	354	-
Diarrhoea and vomiting				
	Floppy child	2	8	-
	Altered consciousness level	2	11	-
	Passing fresh or altered blood PR	2	-	21
	Very hot	2	8	-
	Severe pain	2	10	-
	Signs of dehydration	3	884	-
	Black or redcurrant stools	3	-	311
	Persistent vomiting	3	795	-
	Hot	3	517	-
	Moderate pain	3	413	-
	Vomiting	4	3,232	-
	Warm	4	170	-
	Recent mild pain	4	436	-
	Recent problem	4	1,096	-
Gastrointestinal bleeding				
	Passing fresh or altered blood PR	2	-	191
	Black or redcurrant stools	3	-	1,229
	Persistent vomiting	3	30	-
	Vomiting	4	40	-
CEDIS-PCL				
254 - Diarrhoea		-	5,239	-
257 - Vomiting and/or nausea		-	5,189	-
260 - Blood in stool/melena		-	-	3,240

* Triage severity values according to MTS triage decision trees (MTS REF)

With gastrointestinal bleeding, most cases were found in the group of 65 or older (0.9%). Females more often presented with diarrhoea, vomiting and/or nausea compared to males (2.4% in women versus 1.8% in men), with similar frequencies of gastrointestinal bleeding comparing women and men. Out of all patients that were isolated due to a suspected gastroenteritis, 58.4% were categorised into one of the two health indicators.

Out of the 19,436 cases in the first SySHI category, 5,145 (26.5%) received at least one ICD-10-coded diagnosis. 1,051 (19.3% of 5,435) cases that presented with chief complaints of gastrointestinal bleeding were assigned a diagnosis. From the most frequent (> 1%) ICD-10 diagnoses given to the SySHI-categorised visits, 64.1% were indicative of gastrointestinal infections.

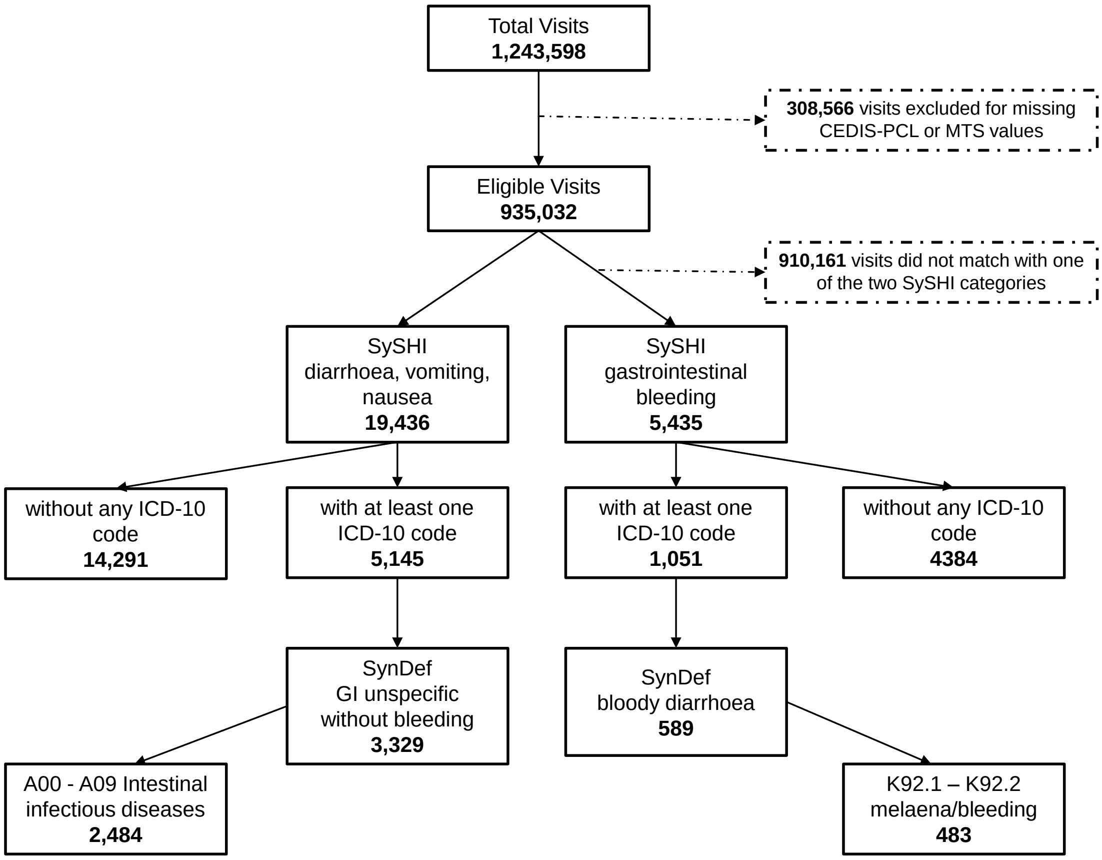
Table 2 – Characteristics of patients visiting an ED, by health indicator (SySHI) category.

		Total (n = 935,032)	SySHI diarrhoea, vomiting, nausea (n = 19,436)		SySHI gastrointestinal bleeding (n = 5,435)		not matching SySHI criteria (n = 910,161)
		n	n	(% of Total)	n	(% of Total)	n
Emergency Department	**Reporting Start**						
Paracelsus-Klinik Henstedt-Ulzburg	2017	44,048	836	(1.90)	212	(0.48)	43,000
Universitätsklinikum Magdeburg	2016	54,475	990	(1.82)	469	(0.86)	53,016
Uniklinikum Göttingen	2016	32,128	689	(2.14)	108	(0.34)	31,331
Pius-Hospital Oldenburg	2016	33,359	814	(2.44)	308	(0.92)	32,237
Ostalb-Klinikum Aalen	2016	71,200	1,374	(1.93)	380	(0.53)	69,446
Klinikum Fürth	2016	89,315	1,548	(1.73)	604	(0.68)	87,163
Klinikum Wolfsburg	2017	97,675	1,904	(1.95)	636	(0.65)	95,135
Klinikum Chemnitz	2016	88,874	1,587	(1.79)	302	(0.34)	86,985
Klinikum Memmingen	2017	58,374	686	(1.18)	221	(0.38)	57,467
Sana Klinikum Offenbach	2012	365,584	9,008	(2.46)	2,195	(0.60)	354,381
Age-Group							
0-4 years		38,896	2,682	(6.90)	81	(0.21)	36,133
5-19 years		100,511	3,295	(3.28)	153	(0.15)	97,063
20-64 years		467,983	7,534	(1.61)	2,145	(0.46)	458,304
65+ years		325,321	5,797	(1.78)	3,043	(0.94)	316,481
Gender *							
Male		481,412	8,460	(1.76)	2,877	(0.60)	470,075
Female		453,301	10,972	(2.42)	2,555	(0.56)	439,774
Triage Severity	**Time to contact**						
1	immediately	5,319	12	(0.23)	18	(0.34)	5,289
2	10 min	102,989	1,261	(1.22)	712	(0.69)	101,016
3	60 min	348,492	7,274	(2.09)	3,783	(1.10)	337,435
4	120 min	358,686	9,154	(2.55)	385	(0.11)	349,147
5	240 min	32,078	240	(0.75)	26	(0.08)	31,812
Isolation *							
No isolation		482,878	7,608	(1.58)	2,521	(0.52)	472,749
Reverse isolation		112	2	(1.79)	1	(0.89)	109
Multiresistant pathogen		3,291	160	(4.86)	47	(1.43)	3,084
Infectious gastroenteritis		2,378	1,269	(53.36)	120	(5.05)	989
Influenza-like illness		345	11	(3.19)	1	(0.29)	333
Other		15,770	1,908	(12.10)	201	(1.28)	13,661

* Categories excluded from this presentation: Gender – "Other" (340 visits total), Isolation – "Tuberculosis" (116 visits total), "Meningitis" (187 visits total), "Herpes viruses" (73 visits total)

Syndromes where defined using a selected set of GI-related ICD-10 codes. 63.2% of all SySHI-categorised cases with at least one ICD-10 code received a matching diagnosis and were included into the syndrome definitions. 3,329 cases were identified by the syndrome definition "GI unspecific without bleeding", 589 through the definition "bloody diarrhoea". (Figure 2)

Figure 2 – Sample and case numbers.



Out of all cases classified by the syndrome definition "GI unspecific without bleeding", 2,484 (74.6%) were given a diagnosis of the ICD-10 group "A00-A09 Intestinal infectious diseases". The majority of those got the unspecific diagnosis "A09.0 Other and unspecified gastroenteritis and colitis of infectious origin" or "A09.9 Gastroenteritis and colitis of unspecified origin". Diagnoses corresponding to specific pathogens were rare. The symptom specific code "R11 – Nausea and vomiting" was given the second most frequent with 18.4%, diagnoses describing symptoms of abdominal pain placed third. Amongst those included in the syndrome definition for bloody diarrhoea, 483 (82.0%) received a diagnosis indicating gastrointestinal bleeding ("K92.1 Melaena" or "K92.2 Gastrointestinal haemorrhage, unspecified"). (Table 3)

Table 3 – ICD-10 codes indicating gastrointestinal infection, by syndrome definition

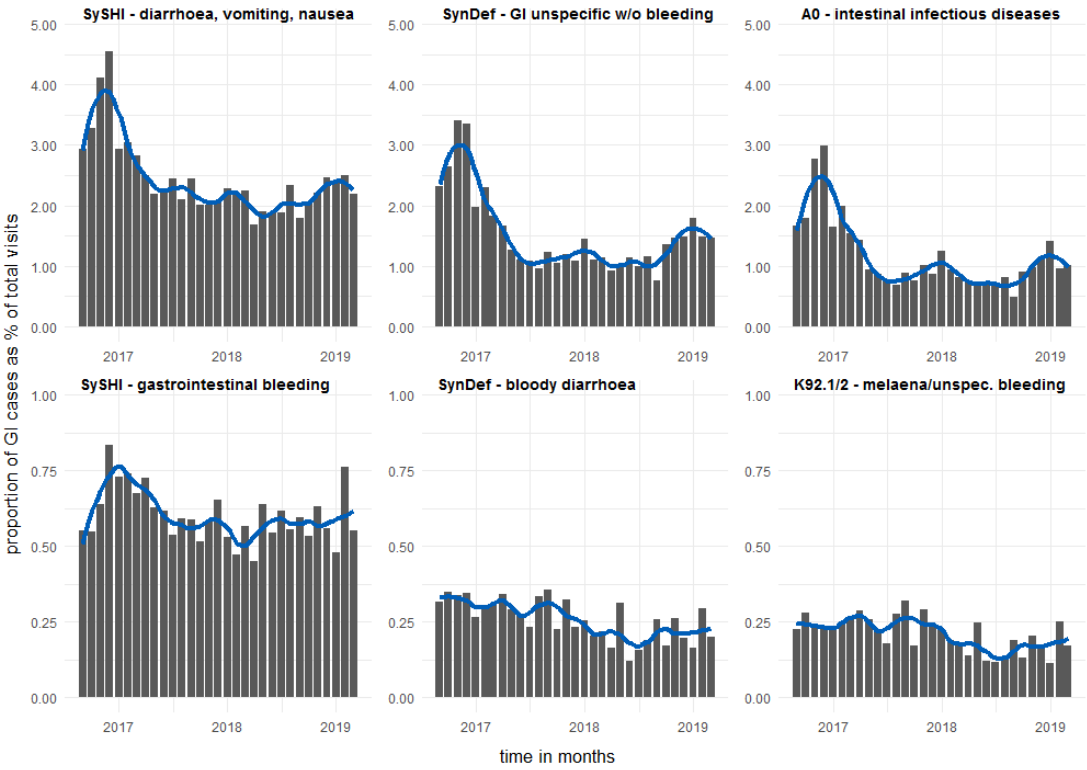
ICD-10 Code	GI unspecific w/o bleeding (n = 3329)		Bloody diarrhoea (n = 589)		Total (n = 15163)	
	n	(%)	n	(%)	n	(%)
A00 - A09 Intestinal infectious diseases	**2484**	**(74.62)**	**69**	**(11.71)**	**3992**	**(26.33)**
A01.0	1	(0.03)	-	-	3	(0.02)
A02.0	2	(0.06)	-	-	3	(0.02)
A04.5	5	(0.15)	-	-	5	(0.03)
A04.7	4	(0.12)	-	-	13	(0.09)
A04.8	1	(0.03)	1	(0.17)	3	(0.02)
A04.9	11	(0.33)	-	-	19	(0.13)
A05.0	2	(0.06)	-	-	2	(0.01)
A05.8	3	(0.09)	-	-	8	(0.05)
A05.9	4	(0.12)	-	-	11	(0.07)
A07.1	1	(0.03)	-	-	1	(0.01)
A08.0	5	(0.15)	-	-	10	(0.07)
A08.1	17	(0.51)	-	-	27	(0.18)
A08.2	3	(0.09)	-	-	6	(0.04)
A08.3	127	(3.81)	2	(0.34)	198	(1.31)
A08.4	208	(6.25)	2	(0.34)	314	(2.07)
A08.5	99	(2.97)	9	(1.53)	164	(1.08)
A09.0	689	(20.70)	17	(2.89)	1025	(6.76)
A09.9	1311	(39.38)	38	(6.45)	2191	(14.45)
K92 Other diseases of digestive system	**28**	**(0.84)**	**483**	**(82.00)**	**910**	**(6.00)**
K92.1	6	(0.18)	98	(16.64)	196	(1.29)
K92.2	22	(0.66)	398	(67.57)	740	(4.88)
P54 Other neonatal heamorrhages	**-**	**-**	**-**	**-**	**-**	**-**
P54.1	-	-	-	-	-	-
P54.3	-	-	-	-	-	-
P92.0 Vomiting in newborn	**5**	**(0.15)**	**-**	**-**	**9**	**(0.06)**
R10 Abdominal and pelvic pain	**284**	**(8.53)**	**42**	**(7.12)**	**9189**	**(60.60)**
R10.1	69	(2.07)	5	(0.85)	1783	(11.76)
R10.3	54	(1.62)	9	(1.53)	3361	(22.17)
R10.4	162	(4.87)	28	(4.75)	4117	(27.15)
R11 - Nausea and vomiting	**614**	**(18.44)**	**2**	**(0.34)**	**1217**	**(8.03)**

3.3 Seasonality

Five of the ten emergency departments provided ICD-10-coded diagnoses from May 2016. For visits categorised as GI cases without gastrointestinal bleeding by SySHI, monthly frequencies varied between 1.7% and 4.5% of all eligible visits, with the highest peak in winter 2016/2017. For the SynDef-identified cases (i.e. with matching ICD-coded diagnoses), frequencies ranged between 0.7% and 3.4%. The general pattern of variation over time was similar, with the biggest increase in the first winter period. Smaller increases in winter 2017/2018 and 2018/2019 became more pronounced in the syndrome definition with only “A00-A09” ICD-10 codes included.

Within those categorised as having gastrointestinal bleeding, no seasonal fluctuations were evident. While the winter peak in 2016/2017 was still visible in the SySHI-based definition, the other two, more differentiated categories (accounting for ICD-coded diagnoses) didn’t show clear patterns of seasonality. Monthly relative case number ranged from 0.5% to 0.8% in the SySHI and 0.1% to 0.4% in the syndrome definition group. (Figure 3)

Figure 3 – Monthly relative frequency of gastrointestinal infection, by syndrome/case definition, visual aid smoothing based on loess interpolation.



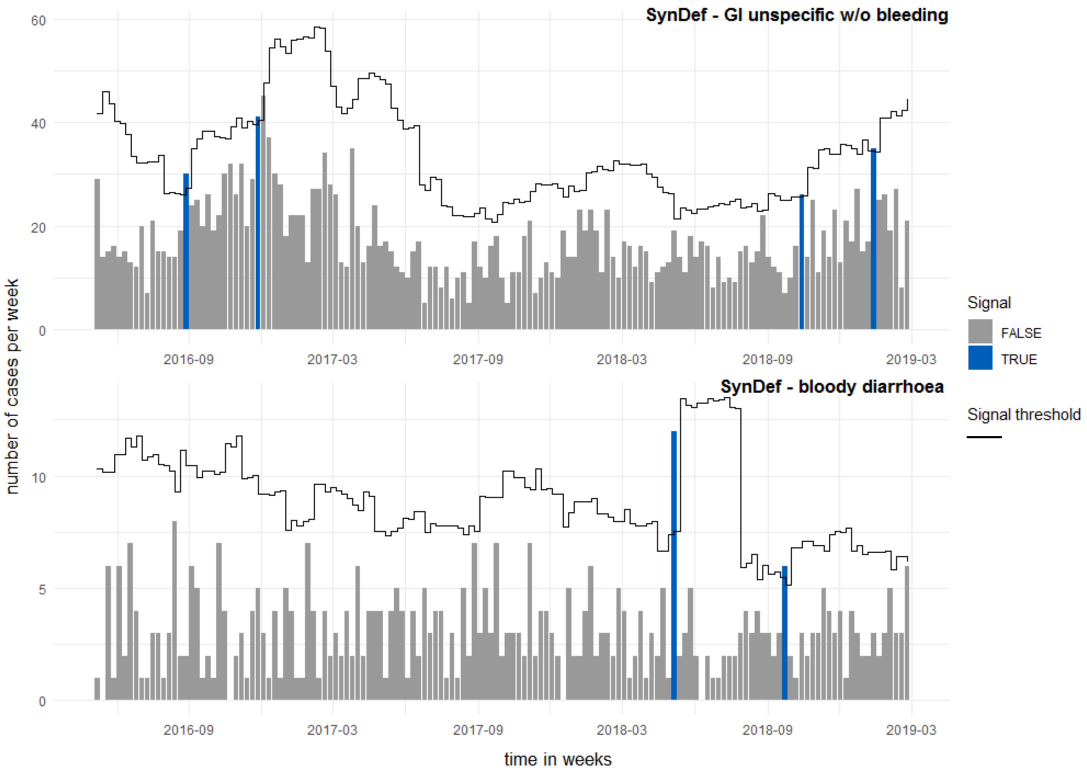
3.4 Short-term Aberration Detection

After reviewing weekly case counts and choosing a timeframe, only one emergency department was eligible for this analysis, data are shown in Figure 4. The earsC algorithm produced a total of six signals for the time period from May 2016 to July 2019.

Within the syndrome definition for “GI unspecific without bleeding”, two signals were generated at the beginning and two at the end of the timeframe. The first appearing in third term 2016 in a week with 30 detected cases, followed by several weeks with half the case numbers. The second signal at the turn of the year 2016/2017 was consistent with the peak seen in the seasonality time series. The third and fourth signals appeared in the third and fourth term 2018 and were both followed by several weeks of lower case numbers.

For the syndrome definition of “bloody diarrhoea”, two alarms were generated in the second and fourth term of 2018. The first signal marked a week with 12 cases, being the highest number in the total time period. The week marked by the second signal had six identified cases, which was not higher than several other weeks before but followed a period of around half the weekly case count. (Figure 4)

Figure 4 – Weekly case counts, by syndrome definition; threshold calculated using the earsC algorithm; data shown exemplary for one emergency department.



4 Discussion

Out of all eligible visits, 2.7% met the criteria of one of the two SySHI categories. Frequencies were homogeneous when stratified by ED. Around half of all isolated cases were identified by the SySHI. More than 60% of cases received a diagnosis indicating gastrointestinal infection. Adding ICD-10 diagnoses as a second layer of information, syndromes were defined identifying 3,329 cases as “GI unspecific without bleeding” and 589 as “bloody diarrhoea”. Within each of the syndromes, a subgroup of cases receiving a specific ICD-10 diagnosis (“A00-A09” and “K92”) was analysed. The ability of routinely collected emergency department data to detect seasonality of unspecific gastrointestinal infections could be demonstrated. For cases without gastrointestinal bleeding, frequencies peaked in the winter months within the SySHI, the syndrome definition and a selected group of diagnoses. The algorithm used for short-time aberration detection created six signals, four within the syndrome definition “GI unspecific without bleeding” and two in the group categorised as “bloody diarrhoea”.

The newly created Syndromic Surveillance Health Indicator helped combine two variables collecting information for different purposes (MTS and CEDIS-PCL) to categorise emergency department visits into two groups. The high proportion of the

SySHI cases being isolated for infectious gastroenteritis as well as the amount of ICD-10 diagnoses relating to gastrointestinal infections within one of the two SySHI groups proved the internal validity of this variable. In line with notification data for pathogens responsible for unspecific GI in 2018, highest case numbers were seen for the age group of children younger than four years. (7)

Classification into the groups with and without gastrointestinal bleeding for the final syndrome definition was defined solely based on the assignment into a SySHI category. With 74% cases in the syndrome definition for “unspecific GI without bleeding” receiving diagnosis of the group “A00-A09 Intestinal infectious diseases” and 82% of the syndrome definition “bloody diarrhoea” receiving diagnosis related to gastrointestinal bleeding, the distinction between those groups seemed to be working.

Seasonal fluctuations could be seen for those receiving an ICD-10 code relating to “Intestinal infectious diseases”, with peaks around the winter months. This is in line with what is known about the seasonality of many gastrointestinal infections. Especially norovirus gastroenteritis is known for seasonal clusters between November and March. (22) The highest peak of infection with “unspecific GI without bleeding” was evident in the winter 2016/2017. This is consistent with a high number of notifications of norovirus gastroenteritis in that timeframe. (7) Accompanied by the fact that norovirus is the second most frequently notified pathogen, it is plausible that a substantial fraction of cases identified as GI unspecific without bleeding through our syndrome definition corresponds to that increase in norovirus cases. Even though seasonal fluctuations could be seen in all steps of information depth, the pattern became clearer with more differentiated information (towards those with only a specific ICD-10 code). This indicates that adding more information layers other than chief complaints might allow the system to detect seasonal changes with higher precision.

For “bloody diarrhoea”, no seasonality was evident from the data. Highest monthly case numbers were seen in winter 2016/2017 within the SySHI-category gastrointestinal bleeding. This peak was completely removed within the syndrome definition, indicating that part of the cases that were recorded with symptoms of gastrointestinal bleeding, were in fact misclassified into that group. Looking at the aggregated notification data of pathogens responsible for gastrointestinal bleeding (EHEC, HUS, Campylobacter, Salmonella, Shigella) (23) the two signals generated for the syndrome definition could not be explained satisfactory. (7)

The data used in this paper comes from hospital routine documentation, intended for a different purpose. As shown in the analyses it is possible to reuse this data in a meaningful way. However, there are several limitations that need to be considered when interpreting the results. All information is collected within the routine documentation processes. A misclassification of variables or incorrect information collection cannot be traced or verified leaving some uncertainty with the majority of data elements. Another problem resulting from the way of data collection is the high number of missing values. While certain data elements like age and sex were almost complete for all eligible visits, other important variables had a substantial amount of missing values. This is especially problematic in case of ICD-10 codes, as that information was the basis for the syndrome definitions. But also information about a mandated isolation, vital parameters or certain hospital administrative data was reported with lots of missing values. Having a more complete data set with regard to those variables would possibly allow an even more distinct definition of syndromes. Within the ten participating emergency departments, not all provided data starting the same day. Also within the timeframes of data provision for each hospital, the completeness of collection of certain variables (like ICD-10 codes) was not constant. This restricted many analyses to a certain timeframe, which was especially challenging for the detection of aberrations in the data. Many signal detection algorithms require a certain amount of time points as a learning period. Even though, the earsC method works with a customizable baseline, the detection ability would increase with more time points available. For the purpose of seasonality monitoring it would be useful to look at patterns across all emergency departments, which was not possible due to the high amount of missing ICD-10 diagnoses in certain hospitals and timeframes.

A successfully running syndromic surveillance with emergency department data can provide crucial contributions to the improvement of population health. Whilst in many cases information on the morbidity and mortality of certain health conditions is collected retrospectively in epidemiological studies, syndromic surveillance systems can picture situations in real time. Especially in the context of communicable diseases timely public health response, either in the form of prevention or treatment can be crucial. With appropriate syndrome definitions applied, syndromic surveillance with emergency department data can monitor cases of illness when they first enter the health system and is in some cases able to detect threatening situations before other surveillance systems. (24) However, syndromic surveillance systems are not only successful in cases were immediate response is needed.

When looking at seasonal patterns of infectious diseases, syndromic surveillance may be able to detect season starts early. (25) By communicating that to local health authorities and emergency departments, timely preventive measures and hospital administrative actions can be taken. It can also help in the identification of target groups for public health interventions, by identifying risk populations of certain health outcomes.

The results of this paper show that the system currently being piloted in Germany is able to reuse routinely collected emergency department data for the syndromic surveillance of health outcomes. By further developing syndrome definitions and including more information layers, the syndromic surveillance system can increase its ability to detect cases. With an increase in the number of participating hospitals and a thereof resulting increase in coverage, the system could play a major role in the monitoring of seasonal patterns and outbreak situations.

5 Conclusion

This first exploration and analysis of routinely collected emergency department data from a piloted syndromic surveillance system in German provides a successful approach for the definition of syndromes to monitor unspecific gastrointestinal infections. The presented analyses show that the provided data can be used to look at seasonal fluctuations and detect aberrations in the case numbers of gastrointestinal infections.

6 References

1. **Institute for Health Metrics and Evaluation IHME (2018):** GBD Compare Data Visualization 2018. URL: http://vizhub.healthdata.org/gbd-compare, last access: 20.09.2019.
2. **GBD Diarrhoeal Disease Collaborators:** Estimates of the global, regional, and national morbidity, mortality, and aetiologies of diarrhoea in 195 countries: a systematic analysis for the Global Burden of Disease Study 2016. Lancet Infect Dis. 2018; 18(11):1211-28.
3. **World Health Organization (2017):** Diarrhoeal Disease. URL: https://www.who.int/news-room/fact-sheets/detail/diarrhoeal-disease, last access: 28.09.2019.
4. **Wilking H, Spitznagel H, Werber D, Lange C, Jansen A, Stark K:** Acute gastrointestinal illness in adults in Germany: a population-based telephone survey. Epidemiol Infect. 2013; 141(11):2365-75.
5. **Schlinkmann KM, Bakuli A, Mikolajczyk R:** Incidence and comparison of retrospective and prospective data on respiratory and gastrointestinal infections in German households. BMC Infect Dis. 2017; 17(1):336.
6. **Bundesamt für Justiz:** Gesetz zur Verhütung und Bekämpfung von Infektionskrankheiten beim Menschen. URL: http://www.gesetze-im-internet.de/ifsg/index.html, last access: 28.09.2019.
7. **Robert Koch Institute (2019):** Infektionsepidemiologisches Jahrbuch meldepflichtiger Krankheiten für 2018.
8. **Triple-S Projekt:** Assessment of syndromic surveillance in Europe. Lancet. 2011; 378(9806):1833-4.
9. **World Health Organization (2019):** Public health surveillance. URL: https://www.who.int/topics/public_health_surveillance/en/, last access: 20.09.2019
10. **Grafstein E, Bullard MJ, Warren D, Unger B, CTAS National Working Group:** Revision of the Canadian Emergency Department Information System (CEDIS) Presenting Complaint List Version 1.1. Canadian Journal of Emergency Medicine. 2008; 10(2), 151-161.
11. **Hughes HE, Morbey R, Hughes TC, Locker TE, Pebody R, Green HK, Ellis J, Smith GE, Elliot AJ:** Emergency department syndromic surveillance providing early warning of seasonal respiratory activity in England. Epidemiol Infect. 2016; 144(5):1052-64.

12. **Elliot AJ, Bone A, Morbey R, Hughes HE, Harcourt S, Smith S, Loveridge P. Green HK, Pebody R, Andrews N, Murray V, Catchpole M, Bickler G, McCloskey B, Smith G:** Using real-time syndromic surveillance to assess the health impact of the 2013 heatwave in England. Environ Res. 2014; 135:31-6.
13. **Hughes HE, Morbey R, Hughes TC, Locker TE, Shannon T, Carmichael C, Murray V, Ibbotson S, Catchpole M, McCloskey B, Smith G, Elliot AJ:** Using an Emergency Department Syndromic Surveillance System to investigate the impact of extreme cold weather events. Public Health. 2014; 128(7):628-35.
14. **Bawa Z, Elliot AJ, Morbey RA, Ladhani S, Cunliffe NA, O'Brien SJ, Regan M, Smith GE:** Assessing the Likely Impact of a Rotavirus Vaccination Program in England: The Contribution of Syndromic Surveillance. Clin Infect Dis. 2015; 61(1):77-85.
15. **Chini F, Farchi S, Ciaramella I, Antoniozzi T, Giorgi Rossi P, Camilloni L, Valenti M, Borgia P:** Road traffic injuries in one local health unit in the Lazio region: results of a surveillance system integrating police and health data. Int J Health Geogr. 2009; 8:21.
16. **Agency for Healthcare Research Quality (2012):** Emergency Severity Index (ESI): A Triage Tool for Emergency Departments. URL: https://www.ahrq.gov/professionals/systems/hospital/esi/index.html. last access: 28.09.2019.
17. **Mackway-Jones K, Marsden J, Windle J, Krey J (2018):** Ersteinschätzung in der Notaufnahme: das Manchester-Triage-System. Bern; hogrefe.
18. **Deutsches Institut für Medizinische Dokumentation und Information (2019):** Internationale statistische Klassifikation der Krankheiten und verwandter Gesundheitsprobleme 10. Revision German Modification Version 2019. URL: https://www.dimdi.de/static/de/klassifikationen/icd/icd-10-gm/kode-suche/htmlgm2019/, last access: 27.09.2019.
19. **Statistische Ämter des Bundes und der Länder (2018):** Ergebnisse des Zensus 2011 zum Download. URL: https://www.zensus2011.de/DE/Home/Aktuelles/DemografischeGrunddaten.html?nn=3065474, last access: 20.09.2019.
20. **Wickham H (2016):** ggplot: Elegant Graphics for Data Analysis. New York; Springer-Verlag.

21. **Salmon M, Schumacher D, Höhle M:** Monitoring Count Time Series in R: Aberration Detection in Public Health Surveillance. Journal of Statistical Software. 2016; 70(10):1–35.
22. **Lubbert C, Mutters R:** Gastrointestinale Infektionen. Internist 2017; 58(2):149-69.
23. **Kist M, Ackermann A, Autenrieth IB, von Eichel-Streiber C, Frick J, Fruth A, Glocker EO, Gorkiewicz G, von Graevenitz A, Hornef M, Karch H, Kniehl E, Mauff G, Mellmann A, von Müller L, Pietzcker T, Reissbrodt R, Rüssmann H, Schreier E, Stein J, Wüppenhorst N:** MIQ 09: Gastrointestinale Infektionen. Qualitätsstandards in der mikrobiologisch-infektiologischen Diagnostik. 2013; 11/2013(2).
24. **Kwan-Gett TS, Baer A, Duchin JS:** Spring 2009 H1N1 influenza outbreak in King County, Washington. Disaster Med Public Health Prep. 2009; 3 Suppl 2:S109-16.
25. **Zheng W, Aitken R, Muscatello DJ, Churches T:** Potential for early warning of viral influenza activity in the community by monitoring clinical diagnoses of influenza in hospital emergency departments. BMC Public Health. 2007; 7:2