

A system for automated outbreak detection of communicable diseases in Germany

M Salmon¹, D Schumacher¹, H Burmann¹, C Frank¹, H Claus¹, M Höhle^{1,2}

1. Department for Infectious Disease Epidemiology, Robert Koch Institute, Berlin, Germany

2. Department of Mathematics, Stockholm University, Stockholm, Sweden

Correspondence: Maëlle Salmon (maelle.salmon@yahoo.se)

Citation style for this article:

Salmon M, Schumacher D, Burmann H, Frank C, Claus H, Höhle M. A system for automated outbreak detection of communicable diseases in Germany. *Euro Surveill.* 2016;21(13):pii=30180. DOI: <http://dx.doi.org/10.2807/1560-7917.ES.2016.21.13.30180>

Article submitted on 15 May 2015 / accepted on 23 November 2015 / published on 31 March 2016

We describe the design and implementation of a novel automated outbreak detection system in Germany that monitors the routinely collected surveillance data for communicable diseases. Detecting unusually high case counts as early as possible is crucial as an accumulation may indicate an ongoing outbreak. The detection in our system is based on state-of-the-art statistical procedures conducting the necessary data mining task. In addition, we have developed effective methods to improve the presentation of the results of such algorithms to epidemiologists and other system users. The objective was to effectively integrate automatic outbreak detection into the epidemiological workflow of a public health institution. Since 2013, the system has been in routine use at the German Robert Koch Institute.

Introduction

In recent years, more and more data have been collected for the routine surveillance of infectious diseases. For instance in Germany, the Robert Koch Institute (RKI) implemented a national electronic surveillance system (SurvNet@RKI) in 2001 in response to the newly enacted Protection against Infection Act that requires regular collection of data on a number of notifiable diseases [1]. Cases are first reported by laboratories or physicians to local health authorities that may perform further investigations, and then transmitted to the RKI via the federal state health authorities. Collected information about cases includes sex, age, and subtype of the pathogen.

In addition to increasing data collection, a multitude of different outbreak detection algorithms for routinely collected public health data have been published [2]. Nonetheless, the added value of applying statistical methods for aberration detection at public health institutions is still subject to discussion because of several challenges, one of which is automating the data analysis and identifying signals without producing

a plethora of signals. For instance, on October 2015, the SurvNet@RKI database contained ca 6.0 million case notifications in 88 different reporting categories such as *Salmonella* or norovirus, while outbreaks often become apparent when inspecting certain subsets of the data, e.g. within a specific geographical area or even a specific age group [3]. The problem is therefore to promptly identify these relevant subsets in the haystack of data. One statistical approach to this problem is to regularly analyse the data as multiple univariate time series in order to detect unexpected aberrations in specific subsets.

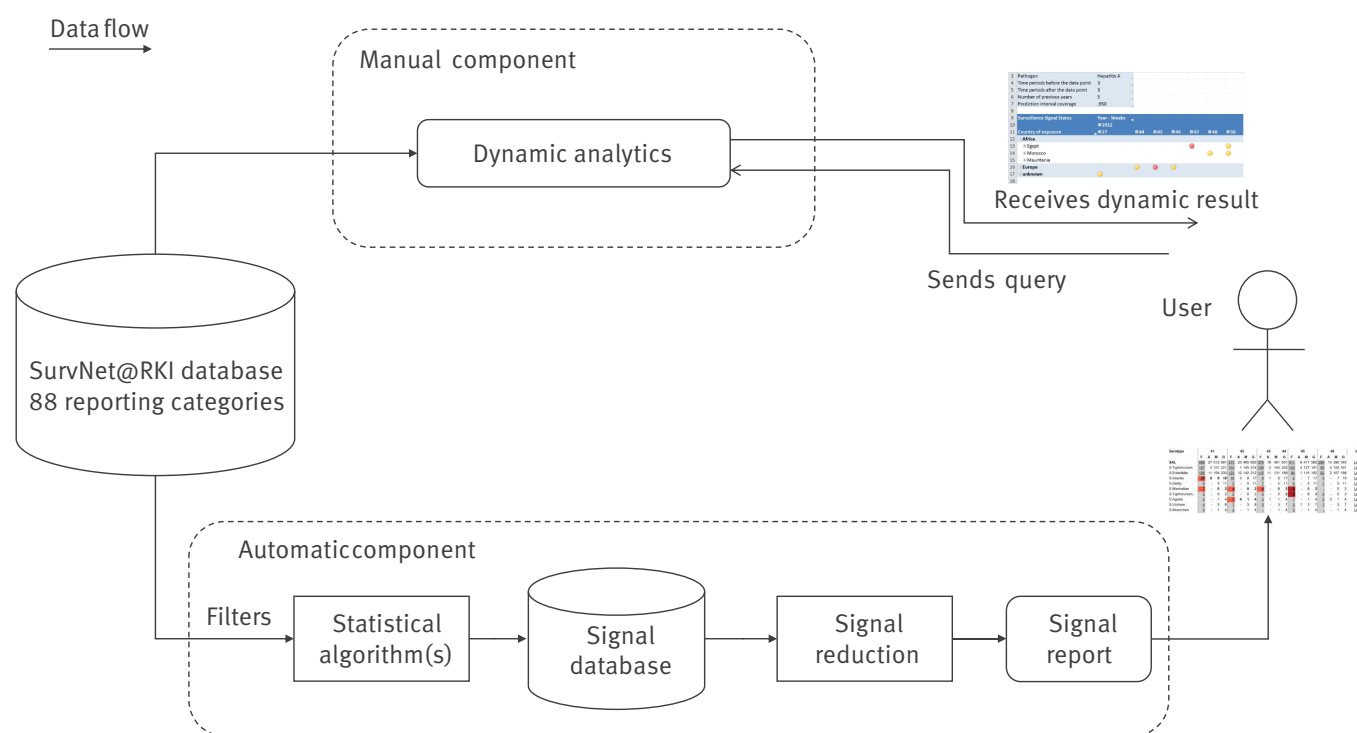
Nowadays, a semi-automatic monitoring system is in operation in many public health institutions (for examples in Europe see [4]). But because of too many signals or a misalignment between users' needs and signal presentation, the system output often has little impact on the practical work of these institutions. First attempts to focus more on the user perspective of monitoring systems are presented in Cakici et al. [4] and Kling et al. [5]. Our goal was to develop and establish an automatic information system that supports epidemiologists at the RKI in the timely detection of potential outbreaks of communicable diseases.

In this article, we present the implementation of a novel automated monitoring system at the RKI in Germany. The new system is now in routine use at the RKI for many reporting categories. Here, we describe the architecture of the system and our design decisions as well as first results and planned improvements. In sharing our experiences we aim to provide valuable information to others working on similar surveillance systems.

System design

Defining features of the system

We wanted to obtain results of a consistent quality as well as a standard procedure for the routine surveillance

FIGURE 1**Structural overview of the automatic surveillance system**

The user can receive output from the automatic component of the system which consists of reports generated with predefined settings. Excerpts of such output are shown in Table 1 and Table 2. The user can also actively make ad-hoc queries to the manual component of the system the output of which is illustrated in Table 3.

workflow in our organisation. This objective led to specific requirements for the system that were largely in line with the checklist for computer-supported outbreak detection systems formulated by Hulth et al. [6]; that article contains recommendations such as user friendliness and tight integration with the database. The development of the system and the refinements of the requirements were conducted iteratively. Based on the rapid prototyping philosophy, we initially focused on building a first prototype for one reporting category, namely *Salmonella* with its many serotypes.

Once the prototypes of the components had produced first results, we started discussing with two users, the epidemiologists in charge of *Salmonella*, the output of the system for *Salmonella*. The experiences from the first prototype led to the design of a weekly automated report sent by email to the two epidemiologists. Once the system produced satisfactory results for this reporting category, we progressively scaled up the system to 48 reporting categories which account for roughly 80% of all received cases. Our goal has always been to create a general system for a variety of diseases instead of highly disease-specific solutions. In addition to the one-on-one discussions with the system users, we received more feedback and feature requests as the system grew.

System design

The system consists of two components: an automatic component routinely monitoring the data and a manual component which enriches data queries with ad hoc aberration detection (Figure 1). The first component automatically produces surveillance reports according to pre-defined settings. The second component allows the user to make customised queries for any time series they wish.

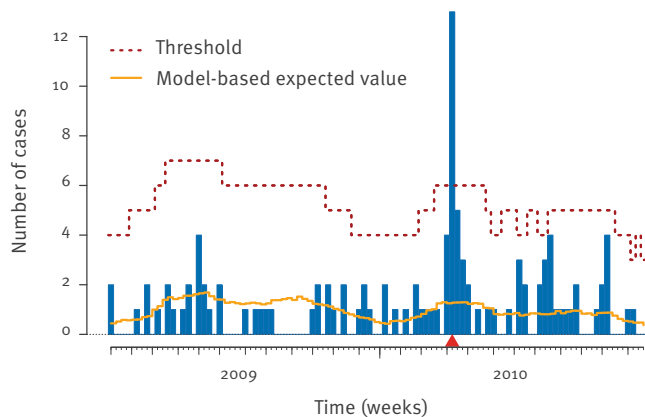
Automated analytical process

As shown in Figure 1, the automatic component consists of three subsystems: an analytical process, a signal database and a signal interface. The analytical process analyses the data with aberration detection algorithms and, in case of an unusually high number of cases, produces a signal which are stored in the signal database and communicated to the user through the signal interface.

The analytical process monitors the SurvNet@RKI case counts of the current and the six previous weeks on a daily basis for all reporting categories selected for aberration detection. Since outbreaks can occur in specific subsets of the population, e.g. at a specific location and in a specific age group, we monitor in parallel numerous time series corresponding to the respective subsets of the population in order to detect

FIGURE 2

Illustration of the time series detection algorithm applied to weekly cases of *Salmonella* Montevideo infection in Germany, 2009–2010



The algorithm is described in [7] and its open-source implementation in [8]. Blue bars indicate the observed number of cases. The red arrow indicates a signal.

signals that would be invisible when analysing the whole population. In particular, we stratify the time series by pathogen subtype (e.g. *Salmonella* serotype such as *S. Infantis*) or symptom (e.g. pneumonia), location (federal state, county), age group, sex, place of exposure. This stratification yields a set of univariate time series for each reporting category aggregated per week or month. The number of diagnostic tests performed is not a variable collected in the German mandatory reporting system. Therefore, the analysis of the numbers is sensitive to variations due to, for example, changes in laboratory procedures or in healthcare-seeking behaviour, e.g. during an outbreak with much media attention.

The system applies the implementation of the algorithm of Noufaily et al. [7] as described in Salmon et al. [8] to each time series in order to get a threshold for each observed count. The last four years of historic data are used as reference values for the algorithm. The algorithm uses an overdispersed Poisson generalised linear model with log link. The linear predictor accounts for seasonality through a 10-level factor variable, includes a time trend and uses a re-weighting scheme for taking past outliers into account. The estimates from the regression model are used to compute a threshold specific for each monitored week, defined as a quantile from the predictive distribution of the current count. A signal is generated for time t_0 if the observed number of cases exceeds the threshold. As an example, Figure 2 illustrates the detection algorithm applied to a single time series of *S. Montevideo* in Germany in 2009 and 2010 [9]. To address reporting delay, we monitor the current week and the six weeks before, i.e. it is possible to obtain a signal for one of the six previous weeks given the current data. This could mean getting a signal in week 5 of 2015 for the

number of *Salmonella* infections reported during week 3 of 2015.

The automated analytical process was initially implemented solely in the statistical programming language *R* [10], using the *surveillance* package [8,11] for the detection part and other *R* packages [12] for the data pre- and post-processing steps, as well as for support for behaviour-driven software development. As in other systems [4,13], the automated component was built in a modular way so that the detection component can incorporate different detection algorithms. *R* was chosen over other programming languages as it allowed us to directly use a variety of statistical detection algorithms and visualisation procedures out of the box and because of its ability to rapidly prototype statistical procedures. During subsequent developments we ported large parts of the data management components to Microsoft C#/.NET to harmonise the system with existing information technology infrastructure at the RKI.

Signal database

The signal database stores signals generated by the analytical process. A signal corresponds to statistical evidence that the case count in a given subset of the data is higher than we would expect it to be based on historic data. A signal combines information about that data segment in which case counts were detected by a statistical algorithm, i.e. a filter on the data with a set of attributes (e.g. ‘Hepatitis A; week 25 of 2013’) and about the algorithm itself, its configuration and its output (e.g. the detection threshold).

This definition can be used directly to store the signals in the signal database and enables subsequent processing of the signals. This has direct advantages over analysis and communication as a combined step: the signals can have an age, they can be more or less important, they can be similar to each other and they can disappear over time when new data are received. In addition, signals can be communicated differently based on aspects such as user preferences.

Signal interface and communication

Signals are communicated to the user through predefined report templates for each reporting category. The reports display relevant signals found for a given category within a given time period. In addition to these main reports, several other reports display new signals found recently (for instance a signal at week 46 for a number of cases reported during week 45, which did not give a signal at week 45 or because of transmission delays), line lists and a spatial visualisation of the cases. The main reports are archived as Microsoft Excel files once a day and are sent by email to epidemiologists in charge of specific reporting categories once a week. Such a push/pull principle of communication was inspired by other monitoring systems such as the one described by Reis et al. [13].

TABLE 1

Automated outbreak detection, excerpt of the *Salmonella* report for weeks 41–46 of 2013: time series analysis at the national level

Serotype	Week 41				Week 42				Week 43				Week 44				Week 45				Week 46			
	y_t	o_t	μ_t	U_t	y_t	o_t	μ_t	U_t	y_t	o_t	μ_t	U_t	y_t	o_t	μ_t	U_t	y_t	o_t	μ_t	U_t	y_t	o_t	μ_t	U_t
<i>Salmonella</i> , all serotypes	466	27	512	691	373	23	485	650	370	16	461	620	356	15	439	601	411	8	417	580	290	14	390	540
<i>S. Typhimurium</i>	107	2	151	221	103	1	145	214	108	2	140	208	106	5	134	202	142	4	127	191	90	4	120	181
<i>S. Enteritidis</i>	158	11	154	230	123	12	142	212	115	11	131	194	84	4	124	189	80	1	116	182	62	2	107	168
<i>S. Infantis</i>	25	6	9	18	16	3	8	17	8	1	8	18	10	-	8	17	2	-	7	17	5	-	7	16
<i>S. Derby</i>	4	NA	5	11	2	NA	5	11	7	NA	5	11	3	NA	5	11	4	NA	5	11	1	-	5	11
<i>S. Manhattan</i>	7	NA	0	2	4	NA	0	2	4	NA	0	2	3	NA	0	2	3	NA	0	2	NA	NA	0	2
<i>S. Typhimurium</i> , monophasic	2	NA	0	2	2	NA	0	2	2	NA	0	2	6	NA	0	2	5	NA	0	3	3	NA	0	3
<i>S. Agona</i>	2	NA	1	4	7	4	1	4	2	1	1	4	3	2	1	4	1	NA	1	4	3	2	1	4
<i>S. Virchow</i>	4	NA	3	8	1	NA	3	8	3	NA	3	7	1	NA	3	7	5	1	3	7	1	NA	3	7
<i>S. Muenchen</i>	3	NA	1	4	3	NA	1	4	NA	NA	1	4	3	NA	1	4	2	NA	1	4	NA	NA	1	4

NA: not applicable.

Columns shaded in grey: observed count of the week. Bold: signals. Yellow: signals detected seven or more days before the current week (here, week 46 of 2013); red: newer signals.

TABLE 2

Automated outbreak detection, excerpt of the *Salmonella* report for weeks 41–46 of 2013: cluster analysis

Serovar	Region	Data filter	Week 41				Week 42				Week 43				Week 44				Week 45				Week 46			
			y_t	o_t	μ_t	U_t	y_t	o_t	μ_t	U_t	y_t	o_t	μ_t	U_t	y_t	o_t	μ_t	U_t	y_t	o_t	μ_t	U_t	y_t	o_t	μ_t	U_t
<i>S. Agona</i>	Germany	Male	1	NA	1	3	5	3	1	3	1	1	1	3	2	2	1	3	NA	NA	1	3	2	2	1	3
	Baden-Württemberg	LK Gernersheim, LK Karlsruhe, LK Rastatt	NA	NA	NA	NA	6	4	0	2	1	1	0	2	2	2	0	2	NA	NA	0	1	2	2	0	1
<i>S. Manhattan</i>	Germany	Age 50–59 years	3	NA	0	1	NA	NA	NA	NA	NA	NA	NA	NA	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
		Male	3	NA	0	2	4	NA	0	2	1	NA	0	2	2	NA	0	2	2	NA	0	2	NA	NA	0	2
		Female	4	NA	0	2	NA	NA	0	2	3	NA	0	2	1	NA	0	2	1	NA	0	2	NA	NA	0	2
	Schleswig-Holstein	NA	3	NA	0	1	NA	NA	NA	NA	1	NA	0	1	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
<i>S. Schwarzengrund</i>	Germany	Male	NA	NA	NA	NA	2	1	0	1	NA	NA	NA	NA	NA	NA	NA	NA	2	NA	NA	NA	NA	NA	NA	NA
<i>S. Typhimurium</i> , monophasic	Germany	Male	2	NA	0	2	1	NA	0	3	NA	NA	NA	NA	4	NA	0	3	3	NA	0	4	1	NA	0	3

LK: Landkreis; NA: not applicable.

Columns shaded in grey: observed count of the week. Bold: signals. Yellow: signals detected seven or more days before the current week (here, week 46 of 2013); red: newer signals.

The signal interface uses Microsoft SQL Server Reporting Services [14], mainly because it is already used at the RKI. It allows quick development of the reports that can be accessed from the Intranet through a web browser and supports the exportation of the reports as Microsoft Excel files. Furthermore, in order to support the decision on whether a signal is relevant, the user can click on any case count in the report to see the associated list of cases from the SurvNet@RKI database (line list).

Signal abstraction

During reporting, a problem arises due to the monitoring of the many time series aggregated in different ways for a reporting category: given a set of closely related signals, what signals should be shown to the user? Closely related signals could be signals for *Salmonella* in week 22 in Bavaria, for *Salmonella* in week 22 in Munich or for *Salmonella* in week 22 in Munich for male cases. Therefore, we developed a method to reduce the number of signals for reporting categories with a high number of signals, such as *Salmonella*.

TABLE 3

Automated outbreak detection, dynamic data query with aberration detection of hepatitis A cases associated with the country of exposure in 2012

Country of exposure	Week 27	Week 44	Week 45	Week 46	Week 47	Week 48	Week 50
Africa							
Egypt					Signal		Signal
Morocco						Signal	Signal
Mauritania							
Europe		Signal	Signal	Signal			
unknown	Signal						

The procedure makes use of the fact that each signal is associated with a filter for a set of attributes, e.g. geographical location, temporal location, sex and age group. Given a set of signals available for reporting, we first determine similar signals by partitioning the original set of signals into a set of signal groups. All signals within a specific group have equal values for a number of filter attributes. For example, we could group all signals by week so that each signal group consists of signals with the same reporting week; e.g. 2013 week 42. In the system at the RKI, we group all attributes except sex, age group and reporting location of the signal. Thus the signals within a group will not necessarily have the same values for sex, age group and location.

In a second step, we filter out some signals in each of these groups, while other similar signals are not filtered, to avoid presenting information which is not considered relevant for users. This is done by so-called filter relations which allow us to rank and compare signals according to a predefined metric. We use three different relations: ‘more specific than’, ‘more general than’ and ‘more specific on the location and more general on age and sex’. The user can select between having no reduction, one of the three relations or a combination of the first two relations. In our example, the most general signal would be the signal for *Salmonella* in week 22 in Bavaria, whereas the most specific signal would be the signal for *Salmonella* in week 22 in Munich for male cases. It is therefore possible to focus the analysis of the signals on specific aspects, e.g. locating the centre of a possible outbreak by displaying only the most specific signals in terms of their filter attributes.

Manual analytical component

In addition to the automatic tool for outbreak detection, we also prepared a detection tool that can be applied to almost any subset of the data defined by the user, allowing users to screen very specific time series on demand, which was a wish expressed during meetings conducted with future users before the design of the system. This component monitors specific subsets of the data, for example case counts of hepatitis A in Berlin within the last six weeks, by comparing the current counts with past data, using a method similar to the algorithm of Stroup et al. [15].

System use

Report interface

As at October 2015, 62 users at the RKI and federal state health authorities received weekly reports from the automated component and interacted with the reports. Table 1 and Table 2 correspond to an excerpt of the Excel-based report for cases of *Salmonella* infection reported in weeks 41 to 46 in 2013. The report contains two data tables with a similar structure. For each week t , we report the number of cases y_t , the estimated expected case count μ_t , the threshold U_t and the number of cases o_t that were manually marked as being part of an outbreak in the SurvNet@RKI database. Cases are sometimes identified as a cluster by local health authorities, e.g. a cluster of cases of norovirus infection after a shared meal. Coloured cells in the Tables indicate signals for the respective week. Signals that were detected seven or more days before the current week are marked yellow, newer signals are marked red. Table 1 corresponds to the reported number of cases per serotype for the six weeks before the current week, in this example with a signal for *S. Infantis* in week 41. Table 2 displays the results of a stratified analysis as described in the previous section. In this example, we see a cluster of female cases of *S. Manhattan* infection in week 41. Some of these signals prompt further checks by epidemiologists, helped by a direct link between the signal and the corresponding cases (line list). The number of signals in a report is an interplay between the number of time series formed by the considered subgroups of sex, age and geographical location, the algorithm settings for the disease, and whether signal reduction is performed. From January to October 2015, the median number of signals over all filters in the weekly *Salmonella* report was 62.

Table 3 shows an excerpt of an output of the manual component for a query of hepatitis A cases for the year 2012 by country of exposure, which is not a time series routinely analysed by the automatic component but could be of interest in particular situations. The table displays weeks of 2012 and countries where the number of cases exceeds the upper limit of the prediction interval.

Experiences from operation

Since 2013, the monitoring system has been widely adopted at the RKI. Although it has not been formally evaluated yet, we can observe a positive user acceptance, for example supported by an increasing number of users and feedback in discussions. Furthermore, the system has contributed to several outbreak investigations. For example, it detected a large local outbreak of cryptosporidiosis in August 2013 [16]. Apart from outbreak detection, the tool has provided awareness to epidemiologists, especially those monitoring trends in frequently notified infections prone to causing outbreaks: the number of cases for various aggregations of the data can now easily be visualised. Moreover, the aberration detection tool for dynamic data queries on case counts is appreciated because it is not always straightforward to visually assess whether the numbers of a time series plot are higher than usual. The manual component of our system provides a statistically informed decision for this.

We developed a system that provides results that are easy to understand and use, while based on sound statistical methods, with disease- and user-specific adjustments. The system was the result of an interdisciplinary collaboration between computer scientists, statisticians and epidemiologists combining user-focused system design, correct treatment of uncertainty and infectious disease knowledge to obtain a decision support tool useful for everyday practice.

Although the system already produces valuable results for routine work at the RKI, a number of improvements are possible. We are working on the problem of comparing frequently incomplete first-version data (e.g. where a pathogen subtype and a possible travel history of the case may now be known yet) to historic, more complete, last-version data (e.g. where subtype and probably country of infection have been added); each version is automatically numbered by the system each time a change is made to a case report. Moreover, it may be possible to add specific detection algorithms for dealing with reporting delays [17,18]. Furthermore, we are currently only able to detect outbreaks when case numbers are above the threshold in at least one week, i.e. if an outbreak emerges very slowly over several weeks it might not be detected quickly. Here, cumulative sum (CUSUM)-oriented procedures could be better at picking up the signal [19] because they add evidence over several timepoints. On a geographical level, only a fixed set of regions is monitored: Germany as a whole, federal states, counties and each county with its adjacent neighbours (which may overlap state borders). Thus we are only able to geographically detect outbreaks that are visible in one of these predefined county clusters. However, the architecture of the system would allow us to include more sophisticated space-time methods into the surveillance process such as those used by Kulldorff, Tango et al. and Neill [20-22]. In addition, performing tests on any time series is a classical case of multiple testing and thus

leads to false alarms. Currently, we offer the epidemiologists the linelist to delve deeper into the data generating the signals in order to better understand the context, so that they can easily navigate the different signals of a report. A framework for controlling overall false alarm rates individually for each user in combination with the signal abstractions could further improve user acceptance.

We see these accounts of successful implementation at public health institutions an important contribution because automatic detection systems are much needed in the current big data environments arising from routine surveillance data collection. Our aim here was to explain the RKI development strategy and user focus of the system. A more technical article describes the algorithmic functionality of the R *surveillance* package [17].

The amount of data held by public health institutes will certainly continue to grow. As a consequence, automatic outbreak detection systems such as the one presented here, will become increasingly important. At the same time, care is needed when integrating such a system into a workflow and taking further steps towards user acceptance. From an organisational point of view, a challenge is to design effective guidelines on how the generated signals are to be handled in a standardised way. This could range from considering signals only as an additional resource for surveillance to having each signal checked by an epidemiologist. Now that the system is in place, one could in the future tailor the detection even more to the needs of the users, e.g. by actively including user feedback in the statistical detection algorithms. Including user feedback could start by collecting appropriate data about the users' reaction to each signal. We think that our experience with an automatic surveillance system will motivate the development and maintenance of similar decision support tools in other European countries.

Acknowledgements

We thank the data scientists at the RKI who have supported and still support the development and operation of the system: Doris Altmann, Susanne Behnke, Göran Kirchner, Eberhard Pape and Thomas Pfoch. We also thank the epidemiologists that have kept providing us with their feedback, in particular Mirko Faber, and all system users whose list would be too long. We are also grateful to Justus Benzler for reviewing an early draft of this manuscript. We would also like to thank the reviewers for their useful feedback.

Conflict of interest

None declared.

Authors' contributions

Based on the initial analysis and design by DS, MH and MS, DS implemented the system that HB now maintains and develops, while MS implemented the statistical algorithm in R.

DS, MS, MH drafted the first version of the manuscript. All authors critically revised the manuscript and approved the final version.

References

1. Faensen D, Claus H, Benzler J, Ammon A, Pfoch T, Breuer T, et al. SurvNet@RKI--a multistate electronic reporting system for communicable diseases. *Euro Surveill.* 2006;11(4):100-3. PMID: 16645245
2. Unkel S, Farrington C, Garthwaite PH, Robertson C, Andrews N. Statistical methods for the prospective detection of infectious disease outbreaks: a review. *J R Stat Soc [Ser A]*. 2012;175(1):49-82. DOI: 10.1111/j.1467-985X.2011.00714.x
3. Koch J, Schrauder A, Alpers K, Werber D, Frank C, Prager R, et al. Salmonella agona outbreak from contaminated aniseed, Germany. *Emerg Infect Dis.* 2005;11(7):1124-7. DOI: 10.3201/eid1107.041022 PMID: 16022796
4. Cakici B, Hebing K, Grünewald M, Saretok P, Hulth A. CASE: a framework for computer supported outbreak detection. *BMC Med Inform Decis Mak.* 2010;10(1):14. DOI: 10.1186/1472-6947-10-14 PMID: 20226035
5. Kling AM, Hebing K, Grünewald M, Hulth A. Two years of computer supported outbreak detection in Sweden: the user's perspective. *J Health Med Informat.* 2012;3(1):108. DOI: 10.4172/2157-7420.1000108
6. Hulth A, Andrews N, Ethelberg S, Dreesman J, Faensen D, van Pelt W, et al. Practical usage of computer-supported outbreak detection in five European countries. *Euro Surveill.* 2010;15(36):8-13. PMID: 20843470
7. Noufaily A, Enki DG, Farrington P, Garthwaite P, Andrews N, Charlett A. An improved algorithm for outbreak detection in multiple surveillance systems. *Stat Med.* 2013;32(7):1206-22. DOI: 10.1002/sim.5595 PMID: 22941770
8. Salmon M, Schumacher D, Höhle M. Monitoring Count Time Series in R: Aberration Detection in Public Health Surveillance. *J Stat Softw.* Forthcoming.
9. Stöcker P, Rosner B, Werber D, Kirchner M, Reinecke A, Wichmann-Schauer H, et al. Outbreak of Salmonella Montevideo associated with a dietary food supplement flagged in the Rapid Alert System for Food and Feed (RASFF) in Germany, 2010. *Euro Surveill.* 2011;16(50):20040. PMID: 22221497
10. R Core Team. R: A Language and Environment for Statistical Computing. 2013: Vienna, Austria.
11. Höhle M. surveillance: An R package for the monitoring of infectious diseases. *Comput Stat.* 2007;22(4):571-82. DOI: 10.1007/s00180-007-0074-8
12. Ripley B, Lapsley M. RODB: ODBC Database Access. 2012. R package version 1.3-10
13. Reis BY, Kirby C, Hadden LE, Olson K, McMurphy AJ, Daniel JB, et al. AEGIS: a robust and scalable real-time public health surveillance system. *J Am Med Inform Assoc.* 2007;14(5):581-8. DOI: 10.1197/jamia.M2342 PMID: 17600100
14. Microsoft. Microsoft SQL Server Analysis Services. 2012.
15. Stroup DF, Williamson GD, Herndon JL, Karon JM. Detection of aberrations in the occurrence of notifiable diseases surveillance data. *Stat Med.* 1989;8(3):323-9. DOI: 10.1002/sim.4780080312 PMID: 2540519
16. Gertler M, Dürr M, Renner P, Poppert S, Askar M, Breidenbach J, et al. Outbreak of *Cryptosporidium hominis* following river flooding in the city of Halle (Saale), Germany, August 2013. *BMC Infect Dis.* 2015;15(1):88. DOI: 10.1186/s12879-015-0807-1 PMID: 25879490
17. Salmon M, Schumacher D, Stark K, Höhle M. Bayesian outbreak detection in the presence of reporting delays. *Biom J.* 2015;57(6):1051-67. DOI: 10.1002/bimj.201400159
18. Noufaily A, Farrington P, Garthwaite P, Enki DG, Andrews N, Charlett A. Detection of infectious disease outbreaks from laboratory data with reporting delays. *J Am Stat Assoc.* Published ahead of print. DOI: <http://dx.doi.org/10.1080/01621459.2015.1119047>
19. Höhle M, Paul M. Count data regression charts for the monitoring of surveillance time series. *Comput Stat Data Anal.* 2008;52(9):4357-68. DOI: 10.1016/j.csda.2008.02.015
20. Kulldorff M. SaTScan - Software for the spatial, temporal, and space-time scan statistics. Boston: Information Management Services; 2014
21. Tango T, Takahashi K, Kohriyama K. A space-time scan statistic for detecting emerging outbreaks. *Biometrics.* 2011;67(1):106-15. DOI: 10.1111/j.1541-0420.2010.01412.x PMID: 20374242

22. Neill DB. Fast subset scan for spatial pattern detection. *J R Stat Soc Series B Stat Methodol.* 2012;74(2):337-60. DOI: 10.1111/j.1467-9868.2011.01014.x

License and copyright

This is an open-access article distributed under the terms of the Creative Commons Attribution (CC BY 4.0) Licence. You may share and adapt the material, but must give appropriate credit to the source, provide a link to the licence, and indicate if changes were made.

This article is copyright of the authors, 2016.