Description of two waterborne disease outbreaks in France: a comparative study with data from cohort studies and from health administrative databases

D. MOULY[1]*, D. VAN CAUTEREN[1], N. VINCENT[1], E. VAISSIERE[1],
P. BEAUDEAU[1], C. DUCROT[2] AND A. GALLAY[1]

[1] *French Institute for Public Health Surveillance, Saint-Maurice, France*
[2] *INRA, Epidemiology Animal Unit, Clermont-Ferrand – Theix, France*

SUMMARY

Waterborne disease outbreaks (WBDO) of acute gastrointestinal illness (AGI) are a public health concern in France. Their occurrence is probably underestimated due to the lack of a specific surveillance system. The French health insurance database provides an interesting opportunity to improve the detection of these events. A specific algorithm to identify AGI cases from drug payment reimbursement data in the health insurance database has been previously developed. The purpose of our comparative study was to retrospectively assess the ability of the health insurance data to describe WBDO. Data from the health insurance database was compared with the data from cohort studies conducted in two WBDO in 2010 and 2012. The temporal distribution of cases, the day of the peak and the duration of the epidemic, as measured using the health insurance data, were similar to the data from one of the two cohort studies. However, health insurance data accounted for 54 cases compared to the estimated 252 cases accounted for in the cohort study. The accuracy of using health insurance data to describe WBDO depends on the medical consultation rate in the impacted population. As this is never the case, data analysis underestimates the total number of AGI cases. However this data source can be considered for the development of a detection system of a WBDO in France, given its ability to describe an epidemic signal.

Key words: Investigation, outbreaks, surveillance system, waterborne infections.

INTRODUCTION

Waterborne disease outbreaks (WBDO) are a public health concern in France because of the proportion of people affected when contamination of drinking water occurs. Almost all WBDO result in outbreaks of acute gastrointestinal infection (AGI) and for most of these, the attack rate in an exposed population reaches 20–50% in France [1]. Children and people with low immunity are usually the most affected. To date, detection of these events is mainly based on the reporting of clusters of AGI by general practitioners (GPs) to health authorities. Consequently, the number of WBDO is probably underestimated in France due to the absence of a specific surveillance system. Improving the detection of infections caused by contaminated drinking water regarding public health, is a challenge to improving the knowledge of risk factors, identifying the drinking water networks with high risk, and proposing appropriate preventive

* Author for correspondence: Mr D. Mouly, InVS-Dcar-Cire Midi-Pyrénées, 10 chemin du raisin 31050 Toulouse, Cedex 9, France.
(Email: damien.mouly@ars.sante.fr)

measures. In this context, the French Institute for Public Health Surveillance is exploring the possibility of using the health administrative databases from the French Health Insurance to develop a national automated detection system of WBDO.

Healthcare administrative databases, which collect data for management and medical purposes, are increasingly used for epidemiological surveillance in developed countries. Several studies using these types of databases have already highlighted their strengths and weaknesses with respect to accurate disease surveillance [2]. In France, an algorithm was specifically developed to identify AGI cases in 2011. It uses data on reimbursement for payment of prescribed drugs from the French National Health Insurance Information System (SNIIRAM; Système national d'information inter régimes de l'Assurance maladie) database [3]. The SNIIRAM database covers 98% of the French population and collects both administrative and individual medical information [4]. Therefore analysis of this data source constitutes one possible approach to develop a detection system of WBDO resulting in AGI. From this perspective, the ability of the SNIIRAM database to describe a WBDO has first to be evaluated.

The benefits of syndromic surveillance to describe WBDO, compared to pharmacy over-the-counter sales data, emergency department visits and even epidemic curves related to AGI have been well documented [5–9]. Nevertheless, to date no comparative study has been published in France to evaluate the use of the SNIIRAM database for the description of WBDO resulting in AGI.

The primary purpose of this study was to compare the SNIIRAM data with a classic epidemiological approach (population-based cohort study) for the description of WBDO. This comparison would improve our knowledge of benefits and limits of SNIIRAM data to describe WBDO with the aim of developing an automated system for their detection with this data source (in process).

MATERIAL AND METHODS

Two different WBDO which occurred in France in 2010 and 2012 were selected for this comparison. For each WBDO, retrospective cohort studies were conducted during both outbreaks and institutional reports (in French) were edited [10, 11]. In the present study, data collected from SNIIRAM will be compared to data previously collected during cohort

studies. The comparison focused on the epidemic curves, the number of cases, individual characteristics (age group, gender), and the extent of the outbreaks.

The two selected AGI WBDO occurred in three municipalities located in the Auvergne region in France, in June 2010 ('WBDO A') and April 2012 ('WBDO B'). The main characteristics of both outbreaks and affected populations are summarized in Table 1.

Data from cohort studies

Two cohort studies were conducted in the population of the municipalities served by the polluted drinking water network (40% of the total municipal population in WBDO A, i.e. 1067 inhabitants, and 100% in WBDO B, i.e. 1753 inhabitants). Three weeks after the beginning of each WBDO, self-administered questionnaires were distributed in the mailboxes of all households served by the contaminated water networks. One overall 'household' questionnaire and four 'individual' questionnaires were distributed to each household. An information letter and a self-addressed, stamped return envelope were also distributed. Data were collected on individual characteristics (age, gender), clinical symptoms (dates of symptom onset, nature and duration), the use of healthcare (medical consultation, date of consultation) and consumption habits of tap water.

For WBDO A, the circumstances that may have led to contamination of the drinking water system included 3 consecutive days of heavy rain, flooding of the system's drinking water borehole and of the mechanical chlorination system (the only treatment mechanism in place). For WBDO B, an incident with the system's sand filter followed by a malfunction of the turbidity alarm was responsible for the introduction of polluted raw water (river) into the drinking water system.

Data from SNIIRAM

SNIIRAM aims at evaluating beneficiaries' healthcare consumption and associated expenditures. It covers more than 98% of the French population and records all reimbursements to patients for out-of-pocket medical procedures, medications and payments to professionals for consultations [4]. AGI medications are included in this database if they are reimbursable, prescribed by a GP and dispensed in a pharmacy. The identification of AGI cases in the two WBDO above

Table 1. *Description of study sites, population and criteria for waterborne disease outbreaks A and B, France (own data, not previously published, available in institutional reports [10, 11])*

Type of waterborne disease outbreak	Waterborne disease outbreak 2010 (WBDO A) [10]	Waterborne disease outbreak 2012 (WBDO B) [11]	
Municipalities impacted	Pérignat les sarlième	Pleaux; Barriac les Bosquet	
Municipal population (Insee, 2010)	2696 inhabitants	1753 inhabitants	
Distribution by age group (years)	*N* (%)	*N* (%)	*P*†
0–5	134 (5%)	44 (3%)	$<10^{-3}$
6–14	393 (15%)	133 (8%)	$<10^{-3}$
15–64	1753 (65%)	968 (55%)	$<10^{-3}$
⩾65	416 (15%)	608 (35%)	$<10^{-3}$
Distribution by gender			
Male	1341 (49·7%)	864 (49·3%)	
Female	1355 (50·3%)	889 (50·7%)	
Drinking water supply of municipalities			
Number of drinking water networks impacted by the pollution	2/4	3/3	
Number of people supplied by polluted drinking water (% of all inhabitants)	1067 people (39·6%)	1753 people (100%)	
Type of source	Mountain spring + borehole in alluvial aquifer	Borehole in surface river	
Type of treatment	Disinfection (Cl₂)	Pre-oxidation + clarification + filtration + disinfection (ClO₂)	
Occurrence of pollution at the time of waterborne disease outbreak			
Circumstances of occurrence of pollution	Heavy rains; Flooding of the borehole; Cessation of chlorination	Heavy rains; River pollution; Operating incident in treatment plant + alarm malfunction	
Date of the pollution intrusion into the drinking water network	17 June 2010	7 April 2012	
Supposed duration of exposure to polluted tap water*	8 days	5 days	
Faecal contamination indicators in drinking water network	>100 c.f.u. per100/ml *E. coli*	>100 c.f.u. per100/ml *E. coli*	

c.f.u., Colony-forming units.
* Delay between pollution intrusion and restrictions on water consumption.
† *P* value of similar population in both municipalities.

required two consecutive steps: (i) data extraction from the SNIIRAM database and (ii) using the AGI algorithm developed by Bounoure *et al.* [3] for selecting AGI cases. The criterion for the data extraction step was the reimbursement for at least one prescribed target drug used to treat AGI† bought by people living in the impacted municipality. The criteria for the AGI discriminative algorithm were: the delay between the prescription and delivery of drugs (<24 h), the number of different AGI-specific drugs prescribed, the treatment duration (<8 days), and the

co-prescription of non-AGI specific drugs (e.g. anti-cancer drugs). Information on age, gender, date of consultation and place of residence was available for each AGI case.

Case definitions

In cohort studies, a case of waterborne AGI was defined as any person in the population exposed to contaminated drinking water, with ⩾3 stools in a 24-h period or vomiting [12] within 3 weeks following contamination of the water system. These cases were defined as 'cohort cases'. Of these cases, those consulting a GP were defined as 'cohort cases with GP consultation'.

† Antiemetic drugs – ATC classification: A04A, A03 F; anti-diarrhoea drugs – A07X, A07D; intestinal adsorbents drugs – A07B, A02X and oral rehydration salts.

Using the SNIIRAM data, people living in the impacted municipalities who consulted a GP within 3 weeks after contamination and who then went to a pharmacy to buy medications prescribed to treat AGI, were defined as 'SNIIRAM cases'.

Data comparison

Description of WBDO

Several epidemiological parameters were used for the description of cohort studies: the *attack rate* in the population was estimated using the ratio between cohort cases and the total number of respondents of the cohort studies. The attack rate was used to estimate the total number of AGI cases in the general population. The *consultation rate* was defined as the ratio between cohort cases who consulted a GP and the total number of cohort cases. Finally, the number of AGI cases in the general population who consulted a GP was estimated from cohort studies by applying age-based consultation rates to the number of AGI cases estimated in the general population.

For SNIIRAM data, the *medication rate* was estimated by comparing SNIIRAM cases with the total population of municipalities impacted (2696 people in WBDO A and 1753 in WBDO B).

The total number of cases assessed from cohort studies and the number of SNIIRAM cases were compared.

Additional comparisons between both data sources included:

- The duration of the epidemic, which was arbitrarily defined as the period covering 90% of the cohort or SNIIRAM cases and starting with the day when at least 5% of the cases had already occurred.
- The delay between the contamination of the water system and the peak of the epidemic curve.
- The distribution of gender and age groups (by applying Fisher's exact test).

Analysis of the correlation between SNIIRAM data and cohort studies

With the view of using SNIIRAM data for the detection of WBDO we tested the similarity of both signals (SNIIRAM *vs.* cohort) by variation of two parameters: (i) the temporal window of aggregation of AGI cases from 1 to 7 days, and (ii) the lag of the two series of AGI cases (SNIIRAM *vs.* cohort) from 0 to 7 days. A correlation coefficient between the two time series was estimated for each pair of values (aggregation level, lag).

RESULTS

Characteristics of outbreak cases

General data

The WBDO A cohort study identified 74 cases (attack rate = 18·1%) in 408 respondents (response rate = 38·2%). Of these, 27 people had consulted a GP (consultation rate = 36·5% [13]) (Table 2). The number of AGI cases in the affected population was estimated from the cohort study at 252, of whom 97 had consulted a GP. The ratios between the number of SNIIRAM cases ($n = 54$) and respectively, the number of AGI cases in the affected population who had consulted a GP (cohort-based estimation), and total AGI cases (cohort-based estimation), were 0·56 (95% confidence interval (CI) 0·42–0·81) and 0·21 (95% CI 0·16–0·31). The pathogen agent identified in the cohort study for WBDO A was *Campylobacter jejuni* in 2/12 patients' stools [13].

In WBDO B, the attack rate estimated in the cohort study was 25·4% (171 cases) for 674 respondents (response rate = 38·4%) [13]. Of these, 50 people had consulted a GP (consultation rate 29·2%). The number of AGI cases in the population was estimated at 458, of whom 123 had consulted a GP. The ratios of cases estimated (see for WBDO A above) were 0·21 (95% CI 0·17–0·28) and 0·06 (95% CI 0·05–0·07). In WBDO B, pathogen agent identified was norovirus genogroup 2 in 4/5 patients' stools [13].

By gender

Men and women were equally affected in both outbreaks and both data sources. The sex ratio (female/male) in both WBDO A and B, respectively, was 1·6 and 0·9 in the cohort studies [13], and 1·0 and 1·4 using SNIIRAM data.

By age group

In both outbreaks, the age groups most affected included children aged <15 years: those aged 6–14 years in the cohort studies [13] (attack rate = 43·1% in WBDO A and 42·9% in WBDO B) and those aged 0–5 years using SNIIRAM data (medication rate = 9·0% in WBDO A and 2·3% in WBDO B) (Table 2).

Table 2. *Description of outbreak cases from cohort studies and Health Insurance data, France, June 2010 and April 2012*

	Cohort study*				SNIIRAM data		
	WBDO A (408 respondents); WBDO B (674 respondents)				Cases with medical consultation followed by reimbursed purchase of drugs (SNIIRAM cases)		
	Cohort cases		Cohort cases with medical consultation				
	N (%)	Attack rate† % (95% CI)	*N* (%)	Consultation rate in cohort‡ % (95% CI)	*N* (%)	Medication rate in population§ % (95% CI)	*P*∥
WBDO A							
Gender							
Male	29 (39·2)	15·4% (10·3–20·6)	10 (37·0)	34·5% (17·2–51·8)	27 (50·0)	2·0% (1·3–2·8)	0·279
Female	45 (60·8)	20·5% (15·2–25·9)	17 (63·0)	37·8% (23·6–51·9)	27 (50·0)	2·0% (1·2–2·7)	
Age group (years)							
0–5	4 (5·4)	14·1% (0·2–28·0)	2 (7·4)	50·0% (1·0–99·0)	12 (22·2)	9·0% (4·1–13·8)	0·006
6–14	16 (21·6)	43·1% (28·9–57·2)	9 (33·3)	56·3% (31·9–80·6)	21 (38·9)	5·3% (3·1–7·6)	0·048
15–64	44 (59·5)	23·3% (17·7–28·8)	14 (51·9)	31·8% (18·1–45·6)	19 (35·2)	1·1% (0·6–1·6)	0·008
⩾65	10 (13·5)	4·9% (0·9–8·9)	2 (7·4)	20·0% (0·0–44·8)	2 (3·7)	0·5% (0·0–1·1)	0·071
Total cases	**74 (100·0)**	**18·1% (14·4–21·9)**	**27 (100·0)**	**36·5% (25·5–47·5)**	**54 (100·0)**	**2·0% (1·5–2·5)**	
Total cases estimated#	**252 (175–328)**		**97 (67–128)**				
WBDO B							
Gender							
Male	88 (51·5)	27·2% (22·3–32·0)	24 (48·0)	27·3% (18·0–36·6)	11 (42·3)	1·3% (0·5–2·0)	0·408
Female	83 (48·5)	23·7% (19·3–28·2)	26 (52·0)	31·3% (21·3–41·3)	15 (57·7)	1·7% (0·8–2·5)	
Age group (years)							
0–5	4 (2·3)	23·4% (3·3–43·5)	1 (2·0)	25·0% (0·0–67·4)	1 (3·8)	2·3% (0·0–6·7)	0·511
6–14	17 (9·9)	42·9% (28·5–57·4)	4 (8·0)	23·5% (3·4–43·7)	2 (11·5)	2·3% (0·0–4·8)	1·00
15–64	91 (53·2)	27·8% (23·1–32·5)	16 (32·0)	17·6% (9·8–25·4)	13 (61·5)	1·7% (0·8–2·5)	0·834
⩾65	59 (34·5)	19·9% (15·1–24·8)	29 (58·0)	49·2% (36·4–61·9)	3 (23·1)	1·0% (0·2–1·8)	0·022
Total cases	**171 (100·0)**	**25·4% (22·1–28·7)**	**50 (100·0)**	**29·2% (22·4–36·1)**	**26 (100·0)**	**1·5% (0·9–2·0)**	
Total cases estimated#	**458 (355–561)**		**123 (94–152)**				

CI, Confidence interval.

* Own data, not previously published, available in institutional reports [10, 11].

† The attack rate was estimated for respondents (408 in WBDO A and 674 in WBDO B).

‡ The consultation rate was estimated for AGI cases in cohort studies (74 in WBDO A and 171 for WBDO B).

§ The medication rate was estimated for total population of municipalities impacted (2696 in WBDO A and 1753 in WBDO B).

∥ The *P* value compare the distribution of gender and age groups in the cohort study versus the SNIIRAM cases.

Estimation of total cases in the population impacted from cohort studies (1067 people in WBDO A and 1753 in WBDO B).
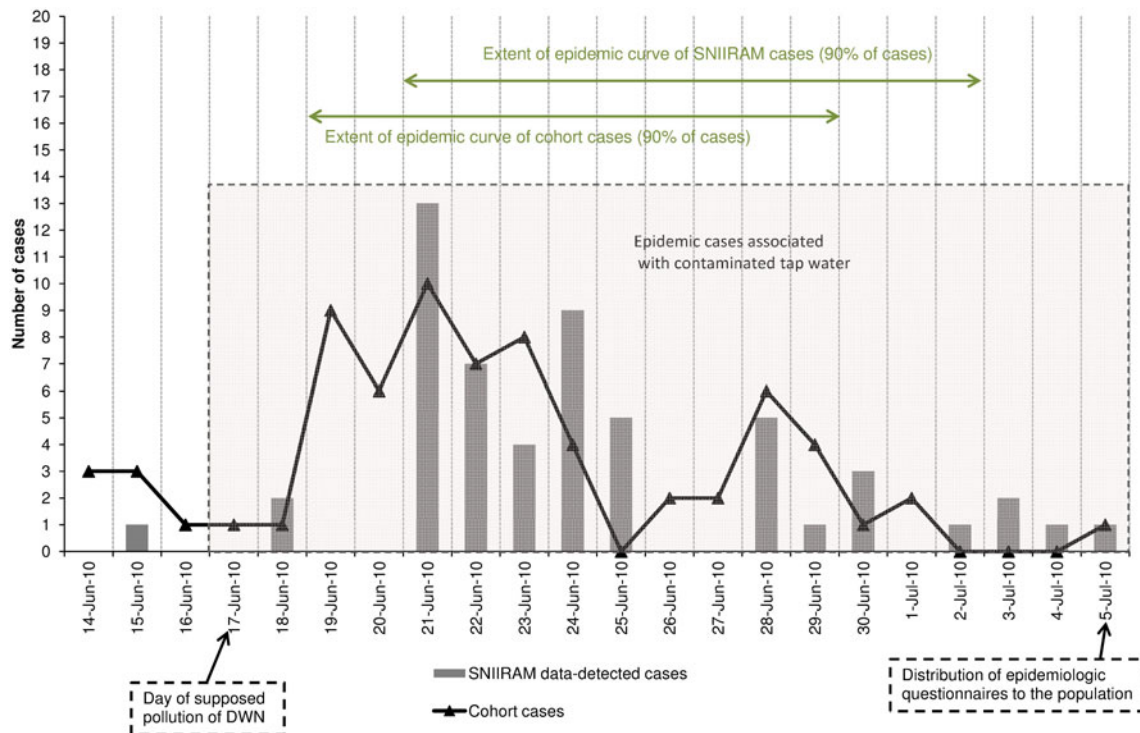
Fig. 1. Description of daily numbers of cohort cases and SNIIRAM cases 14 June 2010 to 5 July 2010 for WBDO A, Pérignat les Sarliève, France, June 2010. In the WBDO A cohort study, missing data existed for ten cases (14%) regarding the date of onset of symptoms, and consequently could not be represented. Cohort data were collected during the outbreak (own data, not previously published, available in institutional reports [10]). DWN, Drinking water network.

In both outbreaks and with both data sources, persons aged >64 years were the least affected age group. Estimated proportions of cases in this age group were from 0·5% to 1·0% using the SNIIRAM data and from 4·9% to 19·9% in the two cohort studies [13]. This age group was also characterized by a different rate of GP consultations compared to other age groups. The lower rate was observed in WBDO A (20%) and the higher in WBDO B (49%), irrespective of the estimation means used.

The 15–64 years age group represented the intermediate age group for the estimated cohort rates and for medication in the SNIIRAM data.

Comparison of epidemic curves

In WBDO A, the temporal distribution of SNIIRAM cases was similar to the distribution of cohort cases (Fig. 1). The duration of the epidemic using SNIIRAM data was 12 days (21 June to 2 July 2010), peaking on 21 June 2010. These results were similar with the cohort data [13]: an epidemic duration of 11 days (19 June–29 June 2010) with an epidemic peak on June 21 (Fig. 1), 4 days after the

contamination of the water system. A secondary peak was observed for both data sources on 28 June 2010.

In WBDO B, temporal distribution of SNIIRAM cases was different from the distribution of cohort cases, no large increase in the number of cases nor the epidemic peak being observed (Fig. 2). Using cohort data [13], the duration of the epidemic was estimated at 14 days (8 April–21 April 2012) with an epidemic peak on 12 April 2012, 5 days after the contamination of the water system (Fig. 2).

### Correlation between SNIIRAM data and cohort studies

The aggregation of cases over 3 days in WBDO A (lag = 1 day) and 5 days in WBDO B (lag = 5 days) is associated with the highest correlation coefficient (respectively 0·83 and 0·94) between epidemic curves from SNIIRAM data and cohort studies (Figs 3 and 4).

DISCUSSION

Our study evaluated the possibility of using the SNIIRAM database to describe WBDO, by comparing the results from the former's data with
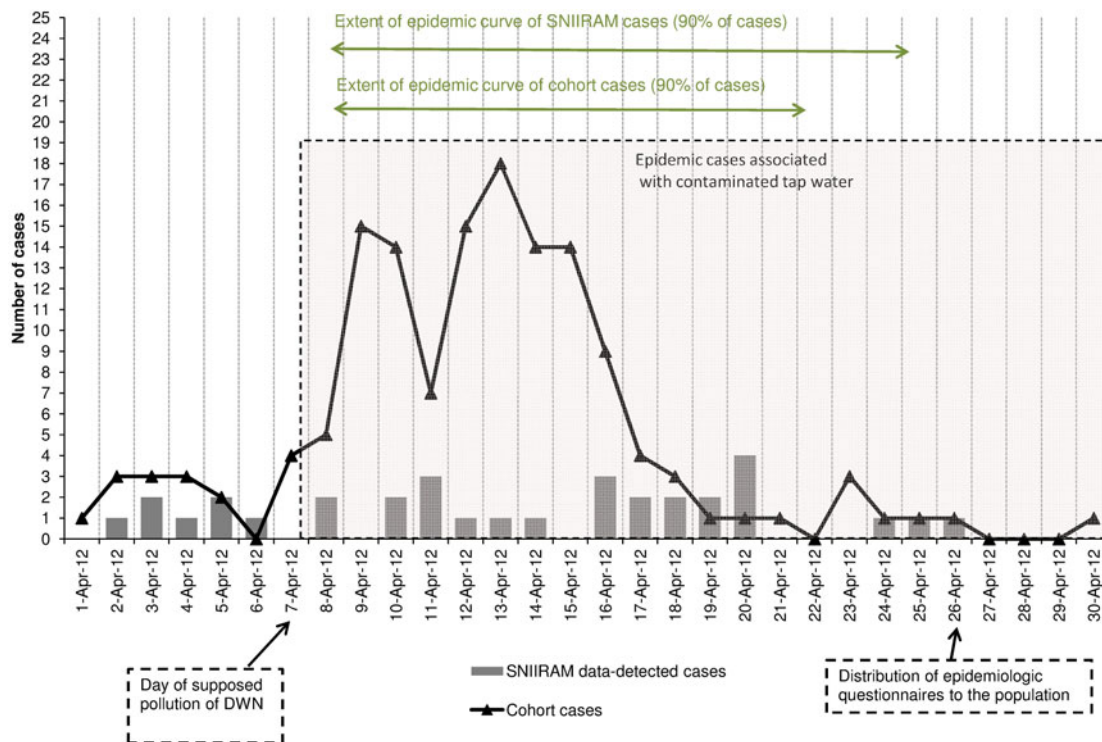
Fig. 2. Description of daily numbers of cohort cases and SNIIRAM cases 1 April 2012 to 30 April 2012 for WBDO B, Pleaux, France, April 2012. In the WBDO B cohort study, 39 (23%) cases had missing data regarding the date of onset of symptoms of cases, and consequently could not be represented. Cohort data were collected during the outbreak (own data, not previously published, available in institutional report [11]). DWN, Drinking water network.
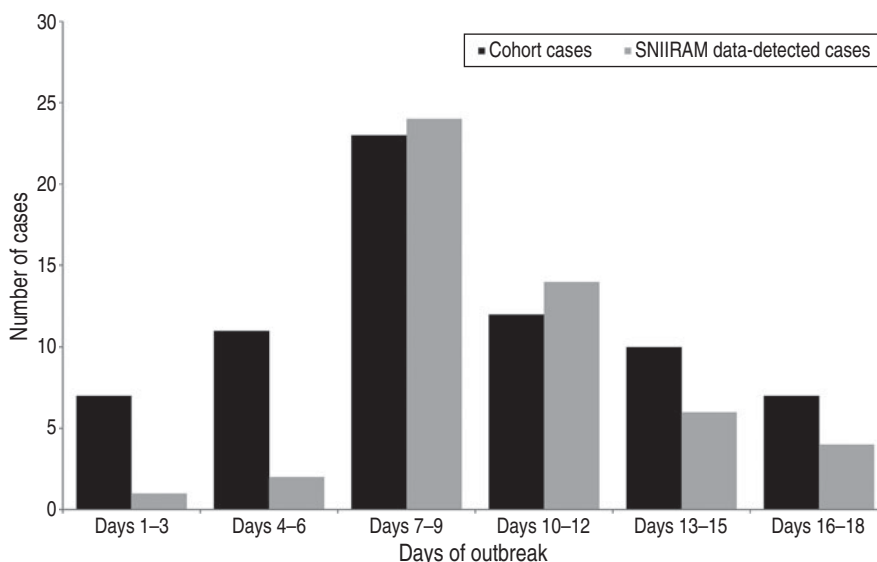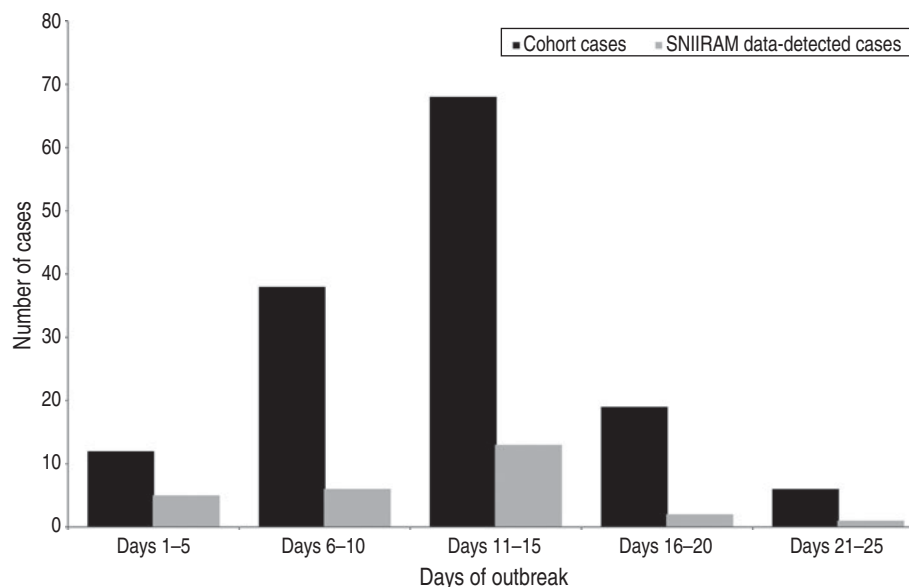

Fig. 3. Distribution of cohort cases and SNIIRAM cases aggregated over 3 days and applying a lag of 1 day on SNIIRAM cases – WBDO A, Pérignat les Sarliève, France, June 2010.

results from two population-based cohort studies. Results of our comparative study point out the benefits and limits of SNIIRAM data for their use in an automated detection system for WBDO as discussed below.

Interpretation of data comparison between SNIIRAM data and cohort studies

The comparison of the two epidemic curves created using data from the cohort studies and from

Fig. 4. Distribution of cohort cases and SNIIRAM cases aggregated over 5 days and applying a lag of 5 days on SNIIRAM cases – WBDO B, Pleaux, France, April 2012.

SNIIRAM showed an accurate representation of the epidemic by the SNIIRAM data in WBDO A. The duration of the epidemic was similar in both curves, the peak of the epidemic occurring the same day in both data sources. In France, it is estimated that more than nine out of 10 AGI cases consult within 3 days of illness onset [14]. To be detected in the SNIIRAM database the delay between the GP visit and drugs delivery in a pharmacy had to be <24 h. Therefore a delay between 0 and 4 days was expected between the cohort cases (date of illness onset) and the SNIIRAM cases (date of GP consultation) for both the outbreak duration and outbreak peak. This delay was not observed in WBDO A and may be explained by the fact that 19–20 June 2010 was on Saturday–Sunday, leading to fewer GP consultations, and reduced or no healthcare utilization. Correlation analysis shows that an aggregation of cases over 3 days allows the optimization of the epidemic signal with data from SNIIRAM (highest coefficient between SNIIRAM and cohort).

For WBDO B no peak was observed using SNIIRAM data. This may be explained by the following factors: first, WBDO B occurred the first day of the Easter weekend (7–9 April 2012). During this period health services were closed and therefore healthcare utilization was limited, with few cases being identified in the SNIIRAM data analyses. Second, school holidays continued for 2 weeks following the Easter weekend (7–22 April 2012). Third, alternative healthcare utilization (e.g. a home visit by a nurse where no prescription was written) cannot be excluded given the 3-day closure period of medical services. Finally people reported previous episodes of drinking water pollution in WBDO B. The knowledge of risk by inhabitants of the municipality (repeated pollution) may have led to more use of the family medicine chest or over-the-counter drugs without medical consultation.

However, by aggregating cases over 5 days in WBDO B, we improved the correlation level (highest coefficient) between SNIIRAM and cohort.

Overall sensitivity of the SNIIRAM data – detected cases

The SNIIRAM data accounted for 21% and 6% of all AGI cases estimated from cohort studies in the population during WBDO A and WBDO B, respectively. These proportions are lower than consultation rates observed in a national population-based study (33%) [14]. Nevertheless, the number of total cases from cohort studies could be overestimated. Indeed, it is possible that ill people were more likely to participate in cohort studies, because of the procedure involved for interviewing people (i.e. the use of a voluntary, self-administered questionnaire). This may have constituted a source of selection bias, leading to an overestimation of the number of AGI cases in the population [15, 16].

Using the attack rate usually associated with WBDO in France (from 20% to 50%) [1], the expected health impact from the SNIIRAM data analysis – percentage of medical cases – would lie between 1% and 10% of people exposed to polluted drinking water. This sensitivity could affect the capacity for detection of WBDO based on the use of SNIIRAM data, especially when a small population is served by contaminated drinking water.

Factors influencing the sensitivity of SNIIRAM indicator for AGI

The overall sensitivity of SNIIRAM data for the description of WBDO may have been influenced by algorithm discrimination of AGI cases, healthcare-seeking behaviour for AGI and access to health services, age and the nature of pathogen.

AGI algorithm

The selection of a case of AGI using SNIIRAM data was dependent on the AGI definition case implemented in the algorithm [3]. The algorithm was cross-validated on a national level with data from the National GP Sentinel Network [17] and with data from a population-based national study [14]. Results showed a good representativeness of the seasonality when using SNIIRAM data, compared to the Sentinel Network, and an estimated annual incidence rate equivalent to that obtained from the national study [14]. Furthermore, the intrinsic sensitivity and specificity of the algorithm were evaluated, each reaching almost 90% [3]. In the context of localized AGI outbreaks such as WBDO, the number of AGI cases selected using SNIIRAM data may be sensitive to the proportion of older people (>65 years) involved. For this age group, we set the selection algorithm more towards specificity, as many treatments, including anti-diarrhoeal medications, are prescribed for reasons other than AGI.

Healthcare-seeking behavior for AGI and access to health services

Several determinants of treatment for AGI in the population may affect the sensitivity of the SNIIRAM indicator as we observed in our study. Although 76% of AGI cases in France use medication, most utilize the family medicine chest (42%) [14]. For these cases, the SNIIRAM data source is blind because of the absence of consultations and prescriptions. Only cases who consulted a physician for AGI (33%) were registered in the SNIIRAM data: 31% of these consulted a GP, 1% a paediatrician and 1% visited the hospital. Alternative healthcare, such as home visits by a nurse, is not visible in SNIIRAM data because of the absence of a drug prescription. Neither does SNIIRAM data take into account over-the-counter medicines bought at a pharmacy without prescription. However, other data collecting over-the-counter information exist in France. Despite the quick availability of over-the-counter data, we considered this data source to be less appropriate for surveillance of WBDO than SNIIRAM data, because of its lack of specificity and the fact that its spatial resolution (pharmacy) did not overlap with the drinking water distribution system [18].

Age and the nature of the pathogen

Deciding to consult a GP is dependent on the person's age and the nature of the pathogen (causative agent). In both WBDO, we observed that younger people (<15 years) were those most affected by the disease (higher attack rate), irrespective of the pathogen in question (Campylobacter sp. in WBDO A and norovirus WBDO B). In WBDO B, the particularly high attack rates in young children can be explained by the greater sensitivity of children to contract AGI, and by the causative agent (virus) of the disease in which secondary transmission plays a more important role than in adults. The GP consultation rate was also higher in younger people. Consequently, relative to the cohort studies, a higher proportion of younger cases occurred using SNIIRAM data (children aged <15 years accounted for 60% of cases from SNIIRAM data in WBDO A vs. 27% from the cohort study). This implies improved sensitivity of the SNIIRAM data for AGI for these age groups.

Consultation rates following WBDO A and B were consistent with those found in a published study showing that the frequency of visits to a GP was more often associated with bacterial than viral infections [19]. Similar trends resulting from waterborne disease outbreaks of AGI have been highlighted in other cohort studies reporting behavioural differences in outbreak situations (e.g. consultation rate = 52% in Gourdon [20]).

Counting waterborne AGI cases

Waterborne outbreak cases, i.e. AGI cases resulting from the consumption of contaminated tap water,

were defined in our study as any AGI case occurring after the day polluted water was introduced into the network. This definition does not distinguish between individual AGI cases due to contaminated drinking water and the baseline of AGI cases. Taking into account the size of the respective municipalities involved, and weekly incidence of AGI reported by the National GP Sentinel Network [17], the number of cases not associated with drinking polluted water during the two WBDO would be 0·75 for WBDO A, and 1·14 for WBDO B. During annual winter outbreaks of AGI with mainly person-to-person transmission, description of a WBDO would necessitate removing cases directly related to the winter outbreak.

Implication for waterborne disease detection

Several studies which deal with the question of the implications for syndromic surveillance of AGI illness or WBDO detection have been published previously [5, 9]. A recent study has compared the ability of three sources of syndromic data (telephone triage, over-the-counter sales, web queries) for the detection of local outbreak signals [6]. Nine outbreaks, which each involved more than 100 cases, were selected. The authors concluded that four out of nine point-source outbreaks were validated in the telephone triage of AGI and two in over-the-counter sales. The three largest outbreaks detected were associated with drinking water contamination and reported between 2400 and 27 000 AGI cases.

Unlike our study, the size and duration of the detected outbreak in Andersson *et al.* [6] were much higher than WBDO A or B. Furthermore, indicators for AGI were established from pre-clinical data, i.e. without medical consultation (which was a prerequisite for SNIIRAM cases). Therefore, one can assume that telephone triage and over-the-counter sales are more sensitive and more readily available than SNIIRAM data, despite their lower specificity.

Furthermore, the challenge of WBDO detection addressed in published studies [5, 6, 9] highlights the difficulty of detecting short outbreaks involving fewer than 100 cases. For this purpose, information collected for cases has to have sufficient temporal (ideally the day) and spatial (municipality could be sufficient) resolution to allow the detection of local outbreak signals like WBDO. Correlation analysis in our study suggests taking into account the aggregation of cases over several days (e.g. 3 and 5 days in WBDO

A and WBDO B, respectively) to optimize the detection of the epidemic signal.

In addition, syndromic surveillance is useful to estimate the size, duration and health impact of detected outbreaks, as we know the consultation rate in the impacted population. This estimation should take into account factors influencing the consultation rate, in particular age and access to health services as shown in our study.

From a public health point of view, detected epidemic signals from SNIIRAM data should be followed by a set of operational measures, including field investigation. These were conducted to validate and describe the outbreak, and to understand the origin and mechanisms involved in case diffusion in order to influence decision-making for public health prevention.

CONCLUSION

We evaluated the ability of SNIIRAM data to describe a WBDO. Our work helped to provide parameters for the description of WBDO of AGI using data from SNIIRAM. It also identified benefits and limits of syndromic surveillance for the detection of WBDO. However, the results of this study, based on two well-documented WBDO, cannot be extrapolated to all WBDO situations and could only be confirmed and complemented by other comparative studies. Nevertheless, the results do allow us to conclude that the use of SNIIRAM data could improve the detection of AGI WBDO with respect to the current surveillance system which is mainly based on GP voluntary reporting. Finally, taking tap water sources of exposure into account in the method of detection of AGI WBDO requires the development of an integrated approach which ensures that data on administrative delimitation of municipalities (aggregation area of AGI cases) and delimitation of the drinking water distribution units (ecological unit of exposure to tap water) can be overlapped.

DECLARATION OF INTEREST

None.

REFERENCES

1. **Beaudeau P, et al.** Lessons learned from ten investigations of waterborne gastroenteritis outbreaks, France, 1998–2006. *Journal of Water and Health* 2008; **6**: 491–503.

2. **Saint-Laurent D, Grémy I, Therre H.** Contribution of health administrative databases to epidemiology and surveillance: crossed views from France and Quebec. *Bulletin Epidemiologique Hebdomadaire* 2013; **19**: 1–58.

3. **Bounoure F, et al.** Syndromic surveillance of acute gastroenteritis based on drug consumption. *Epidemiology and Infection* 2011; **139**: 1388–1395.

4. **Tuppin P, et al.** French national health insurance information system and the permanent beneficiaries sample. *Revue Epidemiologique de Sante Publique* 2010; **58**: 286–290.

5. **Edge VL, et al.** Syndromic surveillance of gastrointestinal illness using pharmacy over-the-counter sales. A retrospective study of waterborne outbreaks in Saskatchewan and Ontario. *Canadian Journal of Public Health* 2004; **95**: 446–450.

6. **Andersson T, et al.** Syndromic surveillance for local outbreak detection and awareness: evaluating outbreak signals of acute gastroenteritis in telephone triage, web-based queries and over-the-counter pharmacy sales. *Epidemiology and Infection* 2014; **142**: 303–313.

7. **Proctor ME, Blair KA, Davis JP.** Surveillance data for waterborne illness detection: an assessment following a massive waterborne outbreak of *Cryptosporidium* infection. *Epidemiology and Infection* 1998; **120**: 43–54.

8. **Buckeridge DL.** Outbreak detection through automated surveillance: a review of the determinants of detection. *Journal of Biomedical Informatics* 2007; **40**: 370–379.

9. **Berger M, Shiau R, Weintraub JM.** Review of syndromic surveillance: implications for waterborne disease detection. *Journal Epidemiology and Community Health* 2006; **60**: 543–550.

10. **Daures M.** Investigation of waterborne gastroenteritis outbreak, Pérignat-lès-Sarliève, France, June-July 2010 [in French]. Saint-Maurice: Institut de veille sanitaire, 2011, 46 pp.

11. **Mouly D.** Outbreak of viral gastroenteritis due to contaminated drinking water in Pleaux, France, April 2012 [in French]. Saint-Maurice: Institut de veille sanitaire, 2013, 44 pp.

12. **Majowicz SE, et al.** A common, symptom-based case definition for gastroenteritis. *Epidemiology and Infection* 2008; **136**: 886–894.

13. **Mouly D, et al.** Authors' data, not previously published, available in institutional reports in references [10, 11]).

14. **Van Cauteren D, et al.** Burden of acute gastroenteritis and healthcare-seeking behaviour in France: a population-based study. *Epidemiology and Infection* 2012; **140**: 697–705.

15. **Goldberg M, Luce D.** Selection effects in epidemiological cohorts: nature, causes and consequences. *Revue d'Epidemiologique et de Santé Publique* 2001; **49**: 477–492.

16. **Hernan MA, Hernandez-Diaz S, Robins JM.** A structural approach to selection bias. *Epidemiology* 2004; **15**: 615–625.

17. **Sentinelles Network.** (https://websenti.u707.jussieu.fr/sentiweb/?lang=en). Accessed 10 November 2014.

18. **Beaudeau P, et al.** Automated detection of gastroenteritis waterborne outbreaks from the data of sale or drug reimbursement [in French]. Saint-Maurice: Institut de veille sanitaire, 2006, 40 pp.

19. **Wheeler JG, et al.** Study of infectious intestinal disease in England: rates in the community, presenting to general practice, and reported to national surveillance. The Infectious Intestinal Disease Study Executive. *British Medical Journal* 1999 17; **318**: 1046–1050.

20. **Gallay A, et al.** A large multi-pathogen waterborne community outbreak linked to faecal contamination of a groundwater system, France, 2000. *Clinical Microbiology and Infection* 2006; **12**: 561–570.