



**HAL**  
open science

## Surveillance of gastrointestinal disease in France using drug sales data

Mathilde Pivette, Judith E. Mueller, Pascal Crepey, Avner Bar-Hen

► **To cite this version:**

Mathilde Pivette, Judith E. Mueller, Pascal Crepey, Avner Bar-Hen. Surveillance of gastrointestinal disease in France using drug sales data. *Epidemics*, 2014, 8, pp.1-8. 10.1016/j.epidem.2014.05.001 . hal-02464647

**HAL Id: hal-02464647**

**<https://ehesp.hal.science/hal-02464647v1>**

Submitted on 31 Aug 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## Surveillance of gastrointestinal disease in France using drug sales data



Mathilde Pivette<sup>a,b,c,\*</sup>, Judith E. Mueller<sup>a,d</sup>, Pascal Crépey<sup>a,e</sup>, Avner Bar-Hen<sup>b</sup>

<sup>a</sup> EHESP French School of Public Health, Sorbonne Paris Cité, Rennes, France

<sup>b</sup> Université Paris Descartes, MAP5, Paris, France

<sup>c</sup> Celtipharm, Vannes, France

<sup>d</sup> Institut Pasteur, Emerging Diseases Epidemiology Unit, Paris, France

<sup>e</sup> Aix Marseille Université, IRD French Institute of Research for Development, EHESP French School of Public Health, UMR.D 190 "Emergence des Pathologies Virales", Marseille, France

### ARTICLE INFO

#### Article history:

Received 18 December 2013

Received in revised form 2 May 2014

Accepted 6 May 2014

Available online 20 May 2014

#### Keywords:

Population surveillance

Disease outbreaks

Drug utilization

Epidemiology

Gastroenteritis

### ABSTRACT

Drug sales data have increasingly been used for disease surveillance during recent years. Our objective was to assess the value of drug sales data as an operational early detection tool for gastroenteritis epidemics at national and regional level in France. For the period 2008–2013, we compared temporal trends of drug sales for the treatment of gastroenteritis with trends of cases reported by a Sentinel Network of general practitioners. We benchmarked detection models to select the one with the best sensitivity, false alert proportion and timeliness, and developed a prospective framework to assess the operational performance of the system. Drug sales data allowed the detection of seasonal gastrointestinal epidemics occurring in winter with a distinction between prescribed and non-prescribed drugs. Sales of non-prescribed drugs allowed epidemic detection on average 2.25 weeks earlier than Sentinel data. These results confirm the value of drug sales data for real-time monitoring of gastroenteritis epidemic activity.

© 2014 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-SA license (<http://creativecommons.org/licenses/by-nc-sa/3.0/>).

### Introduction

Drug sales data have increasingly been used in recent years for outbreak detection and surveillance of acute diseases, like respiratory (Das et al., 2005; Davies and Finch, 2003; Magruder et al., 2004; Ohkusa et al., 2005; Vergu et al., 2006; Magruder, 2003) or diarrheal diseases (Das et al., 2005; Edge et al., 2004, 2006; Stirling et al., 2001; Proctor et al., 1998; Hogan et al., 2003; Pelat et al., 2010; Kirian and Weintraub, 2010; Rodman et al., 1997). Although generally unspecific, drug sales data can provide information within short delays, capture a large part of the population and may detect changes in the population health status (Henning, 2004).

Gastroenteritis is a highly infectious disease caused by viruses, bacteria and parasites. Symptoms are commonly vomiting and diarrhea. Early detection of the start of gastroenteritis epidemics could limit their spreading, since preventative measures could be implemented and public health messages delivered in a timely manner. Previous studies have shown that medications sales

for gastroenteritis are a good proxy of gastroenteritis incidence (Edge et al., 2004, 2006; Stirling et al., 2001; Proctor et al., 1998) and that outbreak onsets could be detected retrospectively 1.75 weeks–2.4 weeks before reference data (Hogan et al., 2003; Pelat et al., 2010). The objective of the present study was to assess the value of drug sales data analysis as an operational epidemic detection tool for gastroenteritis seasonal epidemics in France. We focused on gastroenteritis as it is a public health issue for which early and valid detection is of particular interest and because surveillance data of the disease, reported by a Sentinel Network of physicians, are available in France (French GPs Sentinelles Network). We refined analyses by using non-prescribed and prescribed drug sales separately, as over-the-counter (OTC) drugs have shown to be early indicators of diseases in previous studies (Davies and Finch, 2003; Magruder et al., 2004; Vergu et al., 2006; Magruder, 2003; Hogan et al., 2003; Najmi and Magruder, 2005). We first retrospectively assessed the correlation between medication sales and GP-reported cases of gastroenteritis in France. We performed a benchmark of detection models to select the one with the best sensitivity, false alert proportion and early detection, and we present a method which can be used prospectively for detecting seasonal epidemics based on weekly or daily drug sales data at both national and regional scale.

\* Corresponding author at: EHESP, Avenue du Professeur Léon Bernard, 35043 Rennes, France. Tel.: +33 6 48412957.

E-mail address: [mathilde.pivette@ehesp.fr](mailto:mathilde.pivette@ehesp.fr) (M. Pivette).

## Methods

### Data sources

Daily drug sales data were extracted from a stratified sample of community pharmacies on the French metropolitan territory set up by the company Celtipharm, Vannes, France (Celtipharm). The sample size increased from 1086 pharmacies in 2008 to 3004 (13% of the total pharmacies) in 2013. Real time sales data of these pharmacies are automatically and continuously collected, every day, since 2007. Thanks to a constant up-to-date database of all the 22,458 French active community pharmacies, Celtipharm built a stratification on sale revenue (6 levels for global revenue and 4 levels per type of sales: prescribed drugs, OTC, and other type of sales), localisation (5 geographic areas) and sales area (5 types, from rural to densely urban). Each stratum has a minimum of 30 pharmacies or is merged with neighbouring strata. Sampling rates per strata are computed with the Neyman optimal allocation algorithm (Kish, 1965). Extrapolations from this sample have been validated with data from drug manufacturers who distribute their products to pharmacies. This database has already been used in previous studies (Crépey et al., 2013a,b).

Reference surveillance data were obtained from the French Sentinel Network of physicians (French GPs Sentinelles Network; Valleron et al., 1986; Flahault et al., 2006). The network is constituted of 1300 volunteer general practitioners working in metropolitan France (2% of the total GPs). They transmit each week data from their patient consultations on 8 health indicators via Internet connections. Cases of acute diarrhea are reported since 1990. A case of acute diarrhea is defined as a patient having at least three daily, soft or watery stools in the past 14 days. Weekly data are extrapolated to the national level and are accessible online (French GPs Sentinelles Network). In France, surveillance of gastroenteritis is also based on hospital emergency data and laboratory data in case of investigation of localized outbreaks (school, nursing home. . .) (Invs French Institute for Public Health Surveillance). However, we used Sentinel Network data as reference because they are more representative of the general population.

### Indicator drugs selection

We selected groups of products likely to be prescribed or bought for gastroenteritis. Pharmaceutical products are aggregated into classes according to the Anatomical Classification of the European Pharmaceutical Marketing Research Association (EPHMRA) based on their anatomical site of action, main indication, therapeutic use and composition. We selected the following medication classes (class code): intestinal anti-infectives/antidiarrheals (A07A), intestinal adsorbents/antidiarrheals (A07B), antidiarrheal microorganisms (A07F), motility inhibitors (A07H), other antidiarrheals (A07X), other antiemetics and anti-nauseants (A04A9). Drugs for motion sickness were excluded from the antiemetic group (A04A9). We also selected two classes of parapharmaceutical products: oral rehydration solutions and alimentary products for vomiting and diarrhea commonly used for children. The total selection contains 8 classes, corresponding to 256 products (Fig. S1).

### Data

The study period lasted from January 7, 2008 (week 2 2008) to June 30, 2013 (week 26 2013). We used the weekly number of cases per 100,000 inhabitants. For drugs, we used the weekly and daily number of boxes sold per 100,000 inhabitants at the national and regional level. We differentiated products prescribed by a health practitioner and products purchased without prescription.

### Correlation analysis

Cross-correlation analyses were performed to measure the similarity of the time series at different time lags and the dates of peak in each time series were compared. Since reference data were obtained on a weekly basis, correlation analysis and peak comparisons were performed on weekly data.

### Epidemics detection method

To identify epidemic periods in drug sales data series, we applied a Serfling-type periodic regression model, a widely-used method for detecting outbreaks for diseases with a seasonal background pattern (Costagliola et al., 1991). The method was first proposed by Serfling (1963) and more recently implemented by Pelat et al. (2007) in a software application. The drug data series are assumed to be composed of a periodic baseline level and epidemic periods. We first removed a highest  $N$  percentile of the observations ( $N$  is the pruning value) in the training period (e.g. the highest 30%). A model is then fitted on the remaining data to estimate a non-epidemic baseline level. We modeled the baseline level with a linear trend and commonly used periodicities of 12, 6 and 3 months. We defined the threshold as the upper bound of the prediction interval (PI) of the non-epidemic level (Pelat et al., 2007). An epidemic is defined when a defined number of observations are above the threshold, depending on the detection rule. The onset of an epidemic corresponds to the first observation above the threshold.

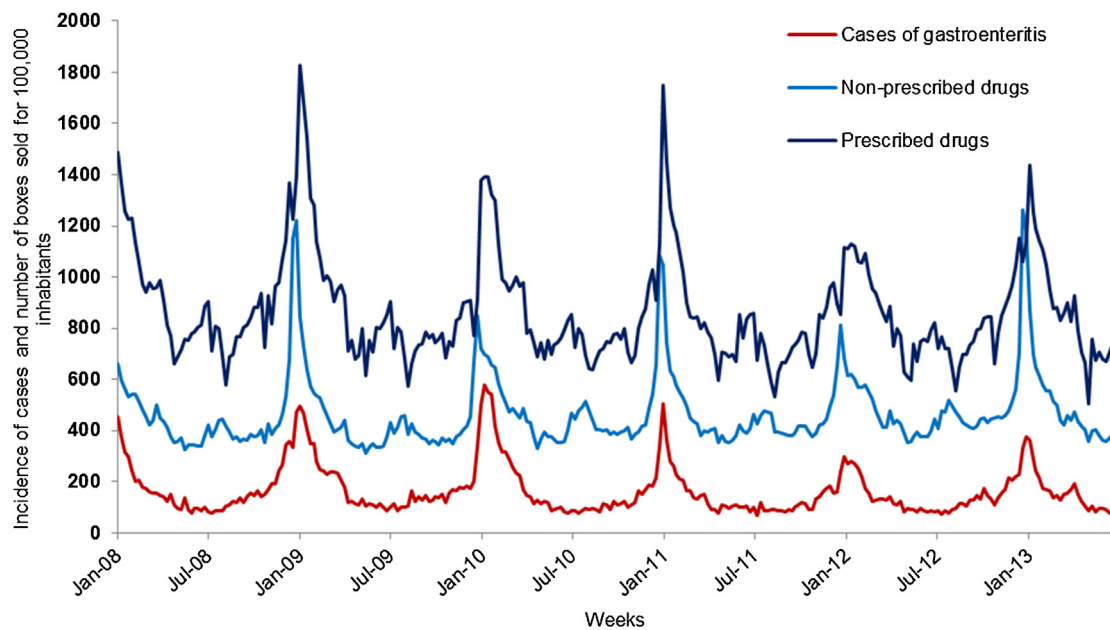
Primary and secondary analyses were performed using weeks and days, respectively, as time unit. For daily analyses, we removed Sundays and public holidays, as most pharmacies in France are closed those days. We also removed one day before and after public holidays, as sales are higher than expected. We added a weekly periodicity in the periodic regression model based on the sales of each weekday to take into account the day-of-the-week effect (Table S1) described in other studies (Ohkusa et al., 2005; Magruder, 2003). The model equations are given in appendix (Text S1).

To identify the best detection model, we tested different key parameters. We excluded between 15% and 40% of the highest observations in the training period. Thresholds were defined by taking the upper limit of the 90%, 95% and 99% prediction interval. We tested detection rules of 1 or 2 consecutive weeks above the threshold to define an epidemic for weekly analyses, and detection rules of 2–8 consecutive epidemic days for daily analyses.

### Epidemics detection evaluation

The reference epidemic periods were the epidemic periods at the national level as they had been defined and published by the Sentinel Network, using periodic regression models on acute diarrhea incidence rates (French GPs Sentinelles Network). We defined a “detection window” of 4 weeks before and 4 weeks after the onset of epidemics defined by the Sentinel Network. The choice of a wide detection window will enable the detection of very early signals of epidemics, until 4 weeks before reference data.

A true positive alert was defined as an epidemic onset computed from drug data occurring in the detection window. A false positive alert was defined as an epidemic onset computed from drug data occurring outside the detection window. The sensitivity represents the proportion of seasonal epidemics that are detected by the drug-based detection system, i.e. if at least one positive alert occurred in the detection window. False alert proportion was defined as the ratio of false positive alerts to the total number of alerts. We focused only on the start of the epidemics, regardless of their duration. Timeliness is defined as the time difference between the first true positive alert and the start of epidemic period from Sentinel Network. Timeliness is undefined when no alert is emitted in the



**Fig. 1.** Weekly sales of prescribed and non-prescribed drugs for gastroenteritis (in number of boxes sold for 100,000 inhabitants), and incidence of gastroenteritis obtained from a Sentinel Network of general practitioners (in number of cases for 100,000 inhabitants). Metropolitan France, calendar week 2 of 2008 to week 26 in 2013.

detection window. Note that a true positive alert can occur at maximum 4 weeks before the detection from Sentinel Network, given our detection window. For daily data, timeliness was calculated using as reference the first day (Monday) of the first epidemic week from Sentinel Network, as epidemic detection from Sentinel is only based on weekly data.

We calculated sensitivity, false alert proportion and timeliness for the different models by varying the key parameters (pruning value, PI and detection rule). The objective was to find the set of parameters allowing first the highest sensitivity, then the lowest false alert proportion, and we finally kept, among that selection, the parameters allowing the earliest detection.

#### Geographical analyses

To evaluate the surveillance tool at regional level and to describe geographical dynamics of gastroenteritis, we applied the selected periodic regression models on regional time series of drug sales data. In each region, every week was defined as epidemic or non-epidemic. We constructed maps representing the weekly epidemic status of the 21 administrative regions of metropolitan France (Corsica excluded).

#### Prospective detection during 2012/2013

To assess the real time performance of the detection model, we evaluated the epidemic detection performance on the last year (2012–2013) based on weekly and daily non-prescribed drug sales, after fitting a periodic regression model on the four preceding years (2008–2012). We first selected the best detection model for the period 2008–2012, using the methodology previously described. The selected model is then extrapolated to forecast an epidemic threshold for the following year (2012–2013). We evaluated sensitivity, false alert proportion and timeliness of the detection in 2012–2013.

All analyses were performed on R Statistical software ([www.r-project.org](http://www.r-project.org)) and Microsoft Office Excel 2010. Maps were prepared with the R statistical software.

## Results

### Descriptive analysis

The five main peaks in the data series of diarrhea incidence rates were also observed in the sale rates of prescribed and non-prescribed drugs (Fig. 1), and in each of the eight individual drug classes (Fig. S2). All main peaks occurred during winter months. Milder increases during summer months were also observed in drug sales series, but not with diarrheal cases. This summer increase appeared among all classes, being most pronounced for sales of oral rehydration solutions and motility inhibitors. Weekly incidences ranged from 66 to 577 cases per 100,000 inhabitants. Weekly drug boxes sold for 100,000 inhabitants ranged from 310 to 1260 for non-prescribed drugs and from 503 to 1829 for prescribed drugs. Overall, the number of prescribed drug sales was 1.9 times higher than the number of non-prescribed drug sales (in France, drugs need to be prescribed to be reimbursed by the French Health insurance).

The prescribed drug sales and the number of reported cases were highly correlated, with a maximum correlation coefficient of 0.89 at no time lag. The maximum correlation between non-prescribed drug sales and cases was observed at a time lag of  $-1$  week with a coefficient of 0.77.

During all five epidemic periods, the peak occurred first in non-prescription drug sales and on average 1.2 weeks before the peak in Sentinel cases. Peaks in prescribed drug sales lagged behind Sentinel data peaks by on average 0.7 weeks (Table 1).

### Evaluation of epidemic detection performance

#### Non prescribed drug sales

**Weekly data analyses.** Analyzing weekly sales of non-prescription drugs, the best detection performance was obtained in a model with a pruning value of 30%, a PI of 99% and with an epidemic period being defined when at least two weeks exceed the threshold (Table 2). In this model, non-prescription drugs allowed detecting all five seasonal epidemics, with no false alert and detection occurred on average 1.6 weeks earlier (median:  $-2$ ; min:  $-3$ ; max:

**Table 1**  
Timing of winter peak for each time series and peak comparison between drug sales data and consultation-based sentinel data. Metropolitan France, week 2 of 2008 to week 26 of 2013.

Winter period	Cases (calendar week of peak)	Non prescribed drugs (calendar week of peak)	Time difference between the peaks in non-prescribed drugs and in Sentinel data (weeks)	Prescribed drugs (calendar week of peak)	Time difference between the peaks in prescribed drugs and in Sentinel data (weeks)
2008/2009	2009 02	2009 01	-1	2009 02	0
2009/2010	2010 02	2009 53	-2	2010 03	+1
2010/2011	2011 01	2010 52	-1	2011 01	0
2011/2012	2012 01	2011 52	-1	2012 03	+2
2012/2013	2013 01	2012 52	-1	2013 02	+1
Mean timeliness (weeks)			-1.2		0.7

**Table 2**  
Sensitivity, false alert proportion and timeliness in detection of gastroenteritis from weekly non-prescribed drug sales. Metropolitan France, week 2 of 2008 to week 26 of 2013. The best detection model was obtained with a pruning value of 30%, and a prediction interval of 99% (gray shade).

PI (%)	Pruning value (%)	Sensitivity	False alert proportion	Mean timeliness (weeks)	Median timeliness (weeks)
90	15	1	0.17	-1.6	-2
90	20	1	0.17	-2.0	-2
90	25	1	0.29	-2.2	-3
90	30	1	0.50	-2.4	-3
90	35	1	0.55	-2.4	-3
90	40	0.8	0.67	-2.0	-2
95	15	1	0.17	-1.2	-2
95	20	1	0.17	-1.6	-2
95	25	1	0.17	-2.0	-2
95	30	1	0.38	-2.2	-3
95	35	1	0.44	-2.4	-3
95	40	0.8	0.69	-2.0	-2
99	15	1	0	-1.0	-1
99	20	1	0	-1.2	-2
99	25	1	0	-1.4	-2
99	30	1	0	-1.6	-2
99	35	1	0.38	-2.2	-3
99	40	1	0.44	-2.4	-3

+1). Detection was earlier in the four last seasonal epidemics, with a mean timeliness of 2.25 weeks (Table 3, Fig. 2A).

**Daily data analyses.** Analyzing daily data, the best detection performance was obtained in a model with a pruning value of 25%, a PI of 90% and with an epidemic period being defined when at least seven consecutive observations were higher than the threshold. The resulting sensitivity was of 100% with no false alert, and the detection occurred on average 12.4 days earlier (median: -16; min: -25; max: +5;). Detection occurred first from non-prescribed drugs in the four last epidemics with a mean timeliness of 16.75 days (Table S2, Fig. S3).

#### Prescribed drug sales

**Weekly data analyses.** Analyzing weekly sales of prescribed drugs, the best performing model used a pruning value of 30%, a PI of 95% a detection rule of two consecutive epidemic weeks (Table S3). Sensitivity was 100%, with no false alert, and a detection occurring

0.6 weeks earlier than reference data (median: 0; min: -2; max: +1) (Fig. 2B, Table 3).

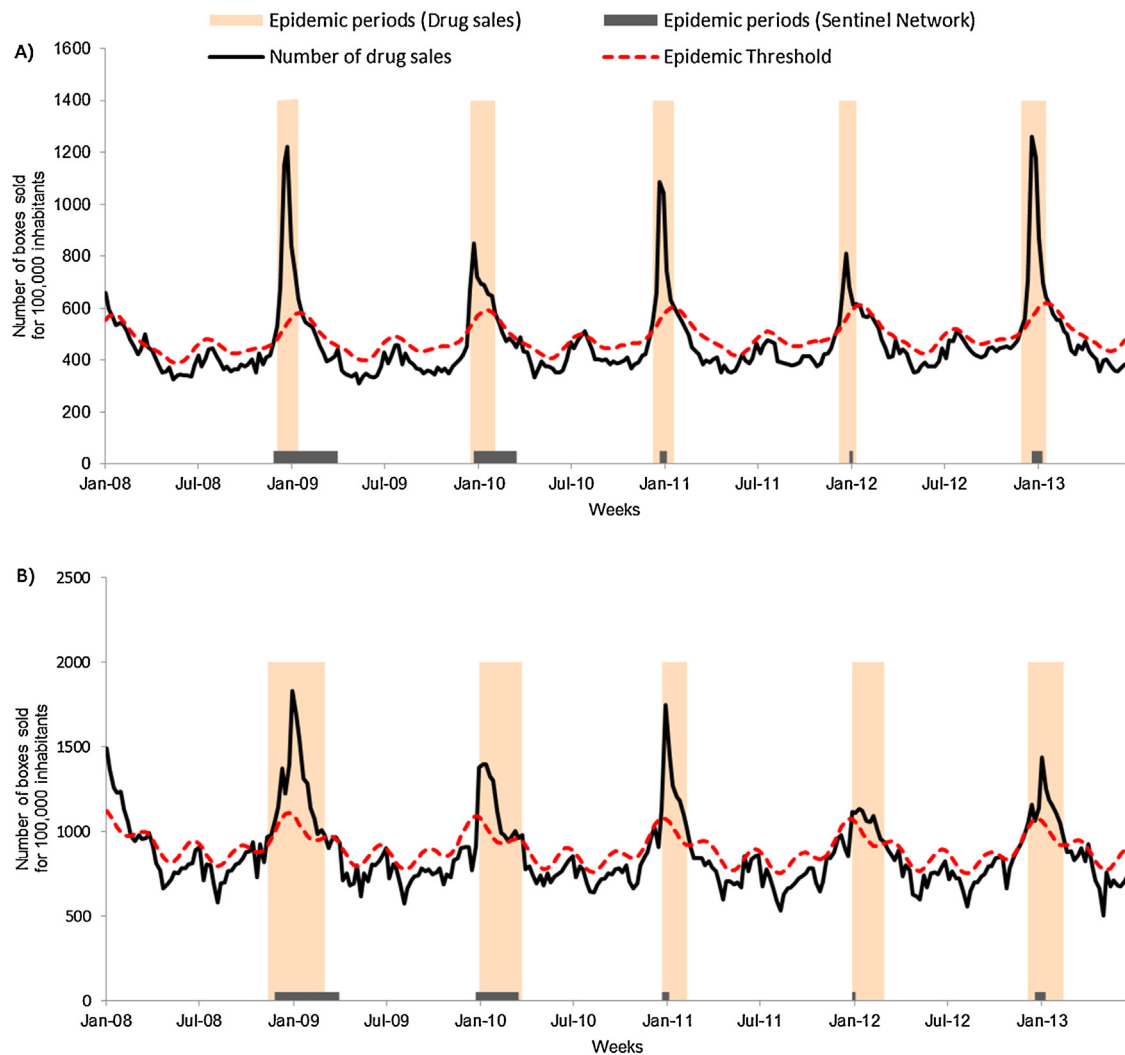
**Daily data analyses.** The best detection performance with daily data was obtained for a pruning value of 40%, a PI of 95% and a detection rule of six days. We obtained a sensitivity of 80% with 20% of false alert, and an average timeliness of +3.6 days (median: 1; min: 0; max: 14). No parameters allowed 100% sensitivity without false alert rate. The lower detection performance can be explained by high inter-week variations in daily prescribed drug sales.

#### Prospective detection in 2012–2013

Using the best performing models fitted only on the four preceding years 2008–2012 (pruning value of 20% and PI 95% for weekly data; pruning value of 20% and PI 90% for daily data), the epidemic period in 2012–2013 was detected without false alert. Based on weekly and daily data, the epidemic period started, respectively, 3

**Table 3**  
Start of epidemic periods obtained each winter from non-prescribed drugs (based on the selected model with a pruning value of 30% and a PI of 99%), from prescribed drugs (based on the selected model with a pruning value of 30% and a PI of 95%) and timeliness of detection compared to Sentinel data. Metropolitan France, week 2 of 2008 to week 26 of 2013.

Year	Sentinel Network epidemic period (calendar week of start)	Non prescribed drugs epidemic period (calendar week of start)	Timeliness of detection from non prescribed drugs compared to Sentinel data (weeks)	Prescribed drugs epidemic period (calendar week of start)	Timeliness of detection from prescribed drugs compared to Sentinel data (weeks)
2008/2009	2008 49	2008 50	+1	2008 47	-2
2009/2010	2009 53	2009 52	-1	2010 01	+1
2010/2011	2010 52	2010 50	-2	2010 52	0
2011/2012	2012 01	2011 50	-3	2012 01	0
2012/2013	2012 52	2012 49	-3	2012 50	-2
Mean timeliness			-1.6		-0.6
Median timeliness			-2		0



**Fig. 2.** Weekly sales of non-prescribed drugs (A) and prescribed drugs (B) with the corresponding epidemic periods defined by the selected detection models. Epidemic periods defined by the Sentinel Network are represented with gridded surfaces.

weeks and 19 days earlier than the period defined in the Sentinel Network (Fig. 3). As 2 consecutive epidemic weeks and 7 consecutive epidemic days were necessary to confirm an epidemic, the start of the seasonal epidemic could have been declared on December 13 using daily data and on December 16 using weekly data, whereas declaration based on the Sentinel Network (French GPs Sentinelles Network) was done on January 7.

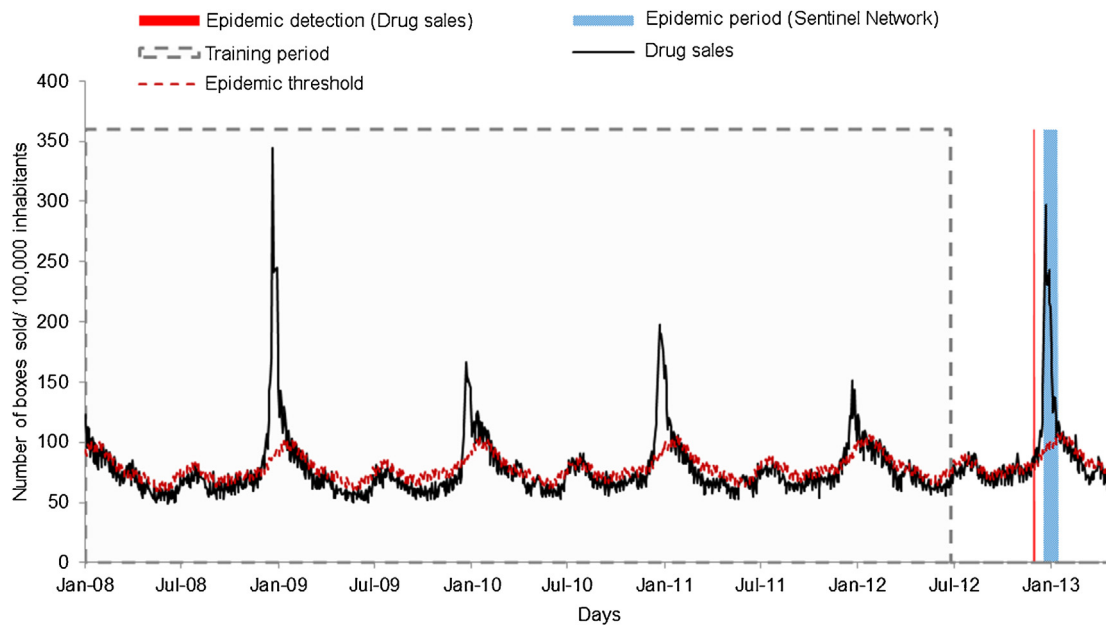
#### Geographical analyses

In Fig. 4, we represented the regions affected by gastroenteritis epidemic every week for each winter period. During the four winters 2009–2013, we observed that some regions were detected as epidemic even before the national detection from drug sales. For example, based on non-prescribed drug sales analysis in 2011–2012 and 2012–2013, epidemic was first detected in the northeast regions of France and one week before the national detection. We did not use reference data at regional level to evaluate the detection performance. Indeed, the sample of sentinel physicians is relatively low and heterogeneous, which implies too much variability at a regional level to use them as reference.

#### Discussion

Our study presents a method for real-time detection of gastroenteritis epidemics based on drug sales data. Our results suggest that non-prescribed drug sales allow epidemic detection substantially earlier than consultation-based sentinel data.

Over-the-counter (OTC) sales of anti-diarrheals, anti-nauseants and rehydration products have proven to be good indicators of gastroenteritis illness in several previous studies (Edge et al., 2004, 2006; Stirling et al., 2001; Proctor et al., 1998; Hogan et al., 2003; Rodman et al., 1997) but only two studies have shown the potential of drug sales for early detection of gastroenteritis compared to traditional surveillance data. In Hogan et al. (2003), detection from OTC electrolytes sales occurred an average 2.4 weeks earlier than detection from hospital diagnoses of respiratory and diarrheal pediatric diseases in the eighteen outbreaks studied. In Pelat et al. (2010), four therapeutic classes correlating with diarrhea incidence were selected and an algorithm based on the selected classes allowed the detection of seasonal gastrointestinal epidemics in France with a sensitivity of 100%, a specificity of 95% and a timeliness of 1.7 weeks before official alerts. However, this study did not make any distinction between OTC and prescription



**Fig. 3.** Prospective epidemic detection in 2012–2013 using daily non-prescribed drug sales data. Metropolitan France, week 2 of 2008 to week 26 of 2013. The vertical red line indicates the epidemic detection from drug sales data. The 2012–2013 epidemic was detected nineteen days earlier than Sentinel Network detection. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

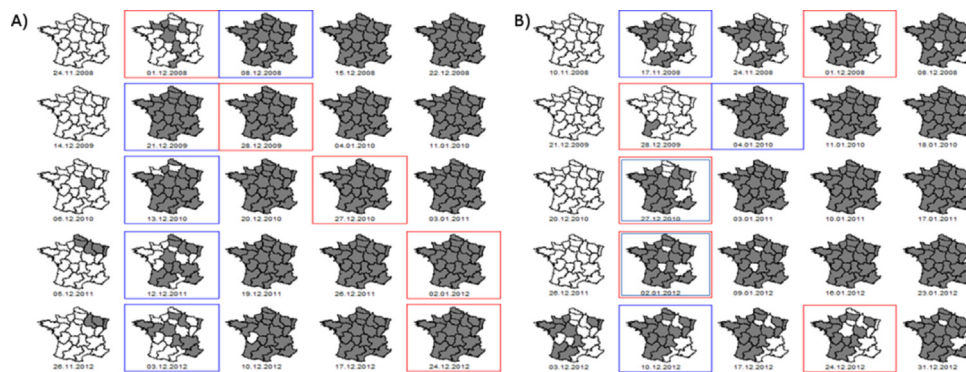
drugs and used only a retrospective analysis framework that do not allow conclusion on usefulness in real-time surveillance.

Contrary to published studies that focused on OTC drug sales (Das et al., 2005; Edge et al., 2004, 2006; Stirling et al., 2001; Proctor et al., 1998; Hogan et al., 2003; Kirian and Weintraub, 2010; Rodman et al., 1997), which can be purchased with or without a prescription, we analyzed products which were actually purchased without a prescription. Our analysis therefore reflects more precisely self-medication behavior of the population, which appears to allow more timely detection of epidemics.

We found that the correlation between non-prescribed drugs and cases was highest with a one week lag-time, indicating a different timing in patient behavior. Non-prescription drug sales peaked earlier than the number of cases every year. Non-prescribed drug sales analysis allowed detecting all seasonal epidemics without false alert and the detection was earlier than reference data in four out of five epidemics with a mean timeliness of 2.25 weeks. Using day as time unit, the detection was on average 16.75 days earlier. This result suggests that non-prescribed drug sales

have the potential to be a sensitive, specific and timely indicator of seasonal gastroenteritis epidemics. It seems that patients buy non-prescribed drugs during the early phase of illness when they become symptomatic, before or instead consulting a physician. The detection occurred later with non-prescribed drugs only during the first year (2008–2009) of the study period; however, the sample of pharmacies was smaller at the beginning of the study period and may partly explain the differences observed with the following years.

By contrast, we found that prescribed drug sales provided relatively little benefits in terms of early detection compared to Sentinel reference surveillance data (on average 0.6 weeks). We found high correlations between prescribed drug sales and the number of reported cases, indicating that both time-series peak approximately at the same time. This result confirmed the adequacy between the two data sources and is explained by the fact that patients buy drugs just after seeing a physician. Prescribed drug sales can therefore be used as a source of information to monitor gastroenteritis disease and is a good indicator of medical consultations.



**Fig. 4.** Weekly epidemic regions defined using non-prescribed (A) and prescribed drugs (B) are represented in black for each winter period. The detection weeks obtained from drug sales analysis at the national level are framed by a blue line, the detection weeks from Sentinel Network data at the national level are framed by a red line. Metropolitan France (Corsica excluded), 2008–2013 winter seasons. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Data sales in our database were obtained on a daily basis, whereas a majority of surveillance studies were based on weekly data. For non-prescribed drugs, the dates of epidemic detection did not differ between analysis of daily and weekly data. The main advantage of daily sales data analysis is the number of observations above the threshold necessary to declare an epidemic period. Contrary to weekly data for which two consecutive epidemic weeks are needed, only seven consecutive epidemic working days are necessary for epidemic detection based on daily data. Epidemic periods can therefore be declared earlier using daily data and public health response can be delivered more rapidly.

In our data collection system, data were available for analysis one day after the day of sale. This near-real-time collection allowed the use of prospective analyses to detect epidemics. Our prospective analysis of the 2012–2013 season showed that non-prescribed drug sales data would have allowed the detection of the epidemic start three weeks or 19 days earlier than the GP based sentinel system, and the warning would have been sent 25 days earlier than official alerts. This confirms the potential of non-prescribed drugs for timely real-time detection and suggests that the method may be used to detect seasonal epidemics in advance. This gain in timeliness may be of interest for the short term planning of healthcare resources (hospital wards and staff management) and for timing more effectively public health campaigns aimed at limiting the spread of the disease, but it may be of an even greater interest in case of emergence or re-emergence of specific pathogens (e.g. *E. coli*).

One limitation of our study is the selection of category of products, which was based on their indications. These products do not all have the same specificity for gastroenteritis and may have others indications. However, analysis of products categories is a simple method that limits the impact of individual drugs changes due to marketing-induced behavior or others causes of variability (e.g. sales promotion, advertising, out-of-stock issues, Hogan and Wagner, 2006). Besides, some drugs that may be used in case of gastroenteritis, such as analgesics or antipyretics, were not included as there were not specific enough of the disease. We could also not differentiate drugs bought for stockpiling in prevention of the seasonal gastroenteritis from the one immediately used. However, sudden and massive stockpiling by the population is unlikely to occur without a triggering public event, which could then be taken into account in the analysis. It is also likely that a medication purchase is more difficult to trigger than a web search query, which makes our analysis more robust than web queries based systems (Olson et al., 2013; Pelat et al., 2009). The observed seasonality cannot be due to bias on data collection. Indeed, control series, such as hormonal contraceptives, presented a stable number of sales during the study period (data not shown). The use of medications may vary by demographic factors, including age and gender (Frosst et al., 2006), and only a part of the population with acute gastroenteritis consults or buys medications. A telephone-based survey (Van Cauteren et al., 2012), conducted in France on 260 patients reporting acute gastroenteritis, found that only 30.9% of the participants purchased prescription-based medications and that 8.9% purchased OTC drugs. However, underestimation is also an issue in other types of surveillance system, as the same study showed that only 33.4% of the patients consulted a physician. The size of the source population is not precisely known. Therefore, analysis of drug sales may not be appropriate to accurately estimate the incidence of gastro-enteritis in a population but is still a relevant tool to determine dynamics of the disease.

We focused on the high-intensity national seasonal gastroenteritis epidemics for which reference clinical data were available and reliable. The detection windows of 4 weeks before and after the epidemic onset from reference data did not allow the detection of localized outbreaks occurring during other periods of the year.

For example, we observed increases in drug data sales during summer months; one hypothesis is that the increase reflects localized gastroenteritis outbreaks in summer. They are of low-intensity, localized in smaller geographical region, and not currently captured by the national network of general practitioners. Considering the large number of pharmacies in the sample (3004) even at a regional level (around 200 on average), we can assume that analyzing drug sales can be a useful alternative way to monitor epidemic spread at a local level. Data sales are available at the pharmacy level, and sales increase in a localized geographical area could therefore be detected rapidly. However to validate this approach, field investigations have to be conducted and field data are needed (laboratory data, GP consultations data, telephone-based data in the study area) to compare drug sales data to the number of cases.

We could not currently validate our regional analysis with reference data since the French GP-based sentinel network is too heterogeneous to provide reliable local estimates on a weekly basis. However, our regional analysis suggests that local epidemics could be detected even earlier than the national one, which could be another way to further improve public health alerts.

Drug sales data present many advantages in term of public health surveillance. Data can be obtained in almost real-time, which allows rapid assessment of a situation. Efforts and cost to automatically collect sales data are lower compared to other types of surveillance system. Finally, data collection is exhaustive for each pharmacy, enabling the monitoring of a large number of diseases at the same time.

## Conclusion

Drug sales data contains information necessary to detect gastrointestinal epidemics occurring in winter but a distinction between prescribed and non-prescribed drugs should be made in order to maximize the timeliness of the detection. Sales of non-prescribed drugs allow epidemic detection on average 2.25 weeks earlier than GP reported clinical data. Our analyses confirmed that drug sales data analysis is a valid and useful tool for gastroenteritis real-time and spatial surveillance. Analysis of drug sales could also be interesting for the surveillance of diseases for which clinical surveillance does not exist, is costly, or poorly reported. Therefore, future drug sales studies should be developed for other infectious diseases.

## Acknowledgments

This research was funded by Celtipharm (Vannes, France) a company specialized in the real time collection and statistical processing of healthcare data ([www.celtipharm.org](http://www.celtipharm.org) – [www.openhealth.fr](http://www.openhealth.fr)), through a doctoral thesis contract for Mathilde Pivette.

## Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at [doi:10.1016/j.epidem.2014.05.001](https://doi.org/10.1016/j.epidem.2014.05.001).

## References

- Celtipharm (Internet). Available from: [www.celtipharm.org](http://www.celtipharm.org)
- Costagliola, D., Flahault, A., Galinec, D., Gamenn, P., Menares, J., 1991. A routine tool for detection and assessment of epidemics of influenza-like syndromes in France. *Am. J. Public Health* 81 (11), 97–99.
- Crépey, P., Pivette, M., Bar-hen, A., 2013a. Quantitative assessment of preventive behaviors in France during the Fukushima nuclear crisis. *PLoS ONE* 8 (3), 1–5.
- Crépey, P., Pivette, M., Desvarieux, M., 2013b. Potential impact of influenza A/H1N1 pandemic and hand-gels on acute diarrhea epidemic in France. *PLoS ONE* 8 (10), 1–6.



- Das, D., Metzger, K., Heffernan, R., Balter, S., Weiss, D., Mostashari, F., 2005. Monitoring over-the-counter medication sales for early detection of disease outbreaks – New York city. *MMWR – Morb. Mortal. Wkly. Rep.* 54 (August (Suppl.)), 41–46.
- Davies, G.R., Finch, R.G., 2003. Sales of over-the-counter remedies as an early warning system for winter bed crises. *Clin. Microbiol. Infect.* 9 (August (8)), 858–863.
- Edge, V.L., Pollari, F., Lim, G., Aramini, J., Sockett, P., Martin, S.W., et al., 2004. Syndromic surveillance of gastrointestinal illness using pharmacy over-the-counter sales. A retrospective study of waterborne outbreaks in Saskatchewan and Ontario. *Can. J. Public Health* 95 (6), 446–450.
- Edge, V.L., Pollari, F., Ng, L.K., Michel, P., McEwen, S.A., Wilson, J.B., et al., 2006. Syndromic surveillance of norovirus using over-the-counter sales of medications related to gastrointestinal illness. *Can. J. Infect. Dis. Med. Microbiol.* 17 (July (4)), 235–241.
- EPHMRA (Internet). Available from: [www.ephmra.org](http://www.ephmra.org)
- Flahault, A., Blanchon, T., Dorléans, Y., Toubiana, L., Vibert, J.F., Valleron, A.J., 2006. Virtual surveillance of communicable diseases: a 20-year experience in France. *Stat. Methods Med. Res.* 15 (October (5)), 413–421.
- French GPs Sentinelles Network (Internet). Available from: <http://websenti.u707.jussieu.fr/sentiweb>
- Frosst, G.O., Majowicz, S.E., Edge, V.L., 2006. Factors associated with the use of over-the-counter medications in cases of acute gastroenteritis in Hamilton, Ontario. *Can. J. Public Health* 97 (6), 489–493.
- Henning, K.J., 2004. What is syndromic surveillance? *MMWR – Morb. Mortal. Wkly. Rep.* 53 (September (Suppl.)), 5–11.
- Hogan, W.R., Wagner, M.M., 2006. Sales of over-the-counter healthcare products. In: Wagner, M.M., Moore, A.W., Aryel, R.M. (Eds.), *Handbook of biosurveillance*. Elsevier Inc., Burlington, MA, pp. 321–331.
- Hogan, W.R., Tsui, F.-C., Ivanov, O., Gesteland, P., Grannis, S., Overhage, J.M., 2003. Detection of pediatric respiratory and diarrheal outbreaks from sales of over-the-counter electrolyte products. *J. Am. Med. Inf. Assoc.* 10 (6), 555–562.
- Invs French Institute for Public Health Surveillance. Surveillance of acute gastroenteritis in France (in French) (Internet). Available from: <http://www.invs.sante.fr/fr/Dossiers-thematiques/Maladies-infectieuses/Risques-infectieux-d-origine-alimentaire/Gastro-enterites-aigues-virales/Contexte-et-dispositif-de-surveillance>
- Kirian, M.L., Weintraub, J.M., 2010. Prediction of gastrointestinal disease with over-the-counter diarrheal remedy sales records in the San Francisco Bay area. *BMC Med. Inform. Decis. Mak.* January (10), 39.
- Kish, J.L., 1965. *Survey Sampling*. Wiley & Sons, New York, NY, pp. 643.
- Magruder, S.F., 2003. Evaluation of over-the-counter pharmaceutical sales as a possible early warning indicator of human disease. *John Hopkins Univ. Appl. Phys. Lab. Tech. Dig.* 24, 349–353.
- Magruder, S.F., Lewis, S.H., Najmi, A., Florio, E., 2004. Progress in understanding and using over-the-counter pharmaceuticals for syndromic surveillance. *MMWR – Morb. Mortal. Wkly. Rep.* 53 (September (Suppl.)), 117–122.
- Najmi, A.-H., Magruder, S.F., 2005. An adaptive prediction and detection algorithm for multistream syndromic surveillance. *BMC Med. Inform. Decis. Mak.* 5 (January), 33.
- Ohkusa, Y., Shigematsu, M., Taniguchi, K., Okabe, N., 2005. Experimental surveillance using data on sales of over-the-counter medications – Japan, November 2003–April 2004. *MMWR – Morb. Mortal. Wkly. Rep.* 54 (August (Suppl.)), 47–52.
- Olson, D.R., Konty, K.J., Paladini, M., Viboud, C., Simonsen, L., 2013. Reassessing Google flu trends data for detection of seasonal and pandemic influenza: a comparative epidemiological study at three geographic scales. *PLoS Comput. Biol.* 9 (10).
- Pelat, C., Boëlle, P.-Y., Cowling, B.J., Carrat, F., Flahault, A., Ansart, S., et al., 2007. Online detection and quantification of epidemics. *BMC Med. Inform. Decis. Mak.* 7 (January), 29.
- Pelat, C., Turbelin, C., Bar-Hen, A., Flahault, A., Valleron, A.-J., 2009. More diseases tracked by using Google trends. *Emerg. Infect. Dis.* 15 (8), 1327–1328.
- Pelat, C., Boëlle, P.-Y., Turbelin, C., Lambert, B., Valleron, A.-J., 2010. A method for selecting and monitoring medication sales for surveillance of gastroenteritis. *Pharmacoepidemiol. Drug Saf.* 19 (October (10)), 1009–1018.
- Proctor, M.E., Blair, K.A., Davis, J.P., 1998. Surveillance data for waterborne illness detection: an assessment following a massive waterborne outbreak of *Cryptosporidium* infection. *Epidemiol. Infect.* 120 (February (1)), 43–54.
- Rodman, J.S., Frost, F., Davis-Burchat, L., 1997. Pharmaceutical sales; a method for disease surveillance? *J. Environ. Health*, 8–14.
- Serfling, R.E., 1963. Methods for current statistical analysis of excess pneumonia–influenza deaths. *Public Health Rep.* 78 (6), 494–506.
- Stirling, R., Aramini, J., Ellis, A., Lim, G., Meyers, R., Fleury, M., et al., 2001. Waterborne cryptosporidiosis outbreak. *Can. Commun. Dis. Rep.* 27 (22), 185–192.
- Valleron, A.J., Bouvet, E., Garnerin, P., Ménarès, J., Heard, I., Letrait, S., et al., 1986. A computer network for the surveillance of communicable diseases: the French experiment. *Am. J. Public Health* 76 (November (11)), 1289–1292.
- Van Cauteren, D., De Valk, H., Vaux, S., Le Strat, Y., Vaillant, V., 2012. Burden of acute gastroenteritis and healthcare-seeking behaviour in France: a population-based study. *Epidemiol. Infect.* 140 (April (4)), 697–705.
- Vergu, E., Grais, R.F., Sarter, H., Fagot, J.-P., Lambert, B., Valleron, A.-J., et al., 2006. Medication sales and syndromic surveillance, France. *Emerg. Infect. Dis.* 12 (March (3)), 416–421.