Epidemics 8 (2014) 1–8

Contents lists available at ScienceDirect

Epidemics

journal homepage: www.elsevier.com/locate/epidemics

Surveillance of gastrointestinal disease in France using drug sales data

Mathilde Pivette[a,b,c,*], Judith E. Mueller[a,d], Pascal Crépey[a,e], Avner Bar-Hen[b]

[a] EHESP French School of Public Health, Sorbonne Paris Cité, Rennes, France
[b] Université Paris Descartes, MAP5, Paris, France
[c] Celtipharm, Vannes, France
[d] Institut Pasteur, Emerging Diseases Epidemiology Unit, Paris, France
[e] 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.

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 (M. Pivette).

http://dx.doi.org/10.1016/j.epidem.2014.05.001

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 stratificationon 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 antiinfectives antidiarrheals (A07A), intestinal adsorbents antidiarrheals (A07B), antidiarrheal microorganisms (A07F), motility inhibitors (A07H), other antidiarrheals (A07X), other antiemetics and antinauseants (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