Review of syndromic surveillance: implications for waterborne disease detection

Magdalena Berger, Rita Shiau and June M Weintraub

J. Epidemiol. Community Health 2006;60;543-550
doi:10.1136/jech.2005.038539

EVIDENCE BASED POLICY AND PRACTICE

Review of syndromic surveillance: implications for waterborne disease detection

Magdalena Berger, Rita Shiau, June M Weintraub

J Epidemiol Community Health 2006;**60**:543–550. doi: 10.1136/jech.2005.038539

Syndromic surveillance is the gathering of data for public health purposes before laboratory or clinically confirmed information is available. Interest in syndromic surveillance has increased because of concerns about bioterrorism. In addition to bioterrorism detection, syndromic surveillance may be suited to detecting waterborne disease outbreaks. Theoretical benefits of syndromic surveillance include potential timeliness, increased response capacity, ability to establish baseline disease burdens, and ability to delineate the geographical reach of an outbreak. This review summarises the evidence gathered from retrospective, prospective, and simulation studies to assess the efficacy of syndromic surveillance for waterborne disease detection. There is little evidence that syndromic surveillance mitigates the effects of disease outbreaks through earlier detection and response. Syndromic surveillance should not be implemented at the expense of traditional disease surveillance, and should not be relied upon as a principal outbreak detection tool. The utility of syndromic surveillance is dependent on alarm thresholds that can be evaluated in practice. Syndromic data sources such as over the counter drug sales for detection of waterborne outbreaks should be further evaluated.

See end of article for authors' affiliations

Correspondence to: Ms M Berger, San Francisco Department of Public Health, Environmental Health Section, 1390 Market Street, Suite 910, San Francisco, CA 94102, USA; mberger@health.nyc.gov

Accepted for publication 19 December 2005

Syndromic surveillance is a tool for outbreak detection that has been used by public health departments since the mid-1990s.[1] The CDC has defined syndromic surveillance as "…surveillance using health-related data that precede diagnosis and signal a sufficient probability of a case or an outbreak to warrant further public health response."[2] In the USA, interest in syndromic surveillance increased after September 11, 2001 because of concern about the possibility of bioterrorist attacks.

Theoretically, syndromic surveillance systems have the potential to supplement traditional infectious disease surveillance systems by providing information about the extent of an outbreak or seasonal increases in disease incidence. They may also provide reassurance that an outbreak is not happening. As syndromic data are gathered before diagnostic or laboratory information is available, health departments may be able to recognise and respond to increases in disease incidence before formal diagnoses are made, and to respond to outbreaks early in their course. In this way, syndromic surveillance has the potential to effectively mitigate the extent of morbidity, mortality, and social and financial unrest resulting from natural or manmade outbreaks.

The effectiveness of syndromic surveillance in terms of informing timely and successful public health interventions has not been demonstrated in public health practice. Although the impetus for its development has been its potential use as a bioterrorism preparedness tool, it may be well suited to monitoring naturally occurring infectious and chronic disease. The utility of syndromic surveillance for monitoring waterborne or environmentally mediated disease incidence has not been assessed to date. However, any decision about the implementation of a syndromic surveillance system cannot focus on only one aspect of disease surveillance and response. This decision has to be made within the context of the existing public health system and its ability to respond to public health threats of all types. Additionally the design of an effective syndromic surveillance system must take into account the data needs and response capabilities of many different public health responders.

This review summarises the evidence gathered from retrospective, prospective, and simulation studies to assess the efficacy of syndromic surveillance for waterborne diseases detection. The original aim of this review was to determine the utility of syndromic surveillance systems for waterborne disease detection in the San Francisco Bay Area. All articles were selected from peer reviewed articles found on Medline under the search term "syndromic surveillance".

DATA SOURCES

Three levels of data may be used for disease surveillance: pre-clinical data, clinical pre-diagnostic data, and diagnostic data.[3] Syndromic surveillance usually uses two types of data sources: pre-clinical and clinical pre-diagnostic data. Traditional surveillance generally focuses on diagnostic data.

Pre-clinical data sources

Public behaviour may be an early indicator of an increase in disease incidence in the population. Information about school and work absenteeism,[1 4] calls to nurse help lines,[5] poison control centres,[6] water utility complaint lines,[7] and sales information of over the counter and prescription drugs[1 4 7–14] may give information about the level of disease in the population. Other environmental data, such as drinking water turbidity levels, may also be useful indicators of disease incidence. While pre-clinical data sources tend to be

more timely than clinical data sources, they are much less specific and therefore provide a less solid basis for public health decision making.

Clinical pre-diagnostic data sources

Most clinical data sources currently used by syndromic surveillance systems consist of electronic data gathered for an independent purpose and adapted to the needs of the specific surveillance system. Sources of clinical data used for syndromic surveillance include health care utilisation records such as registers of emergency department chief complaints,[1 4 15–24] discharge data,[20] ambulance dispatch data,[25] and ICD-9 or other procedure codes used in inpatient and outpatient settings.[4 20 26] Electronically captured orders for diagnostic tests may also be used as a data source.

Different groupings of symptoms, ICD-9 codes, or drugs, as well as the most effective free text search algorithms are being researched to determine how these data are best analysed to accurately reflect disease activity in the population being studied.[17 21 27 28]

DATA TIMELINESS AND ACCURACY

Usually, data sources that are chronologically distant from diagnostic information are timely but not as specific as data sources that are chronologically closer to diagnosis.[3] Potential syndromic surveillance data sources along with each source's capacity for being timely and specific are described in figure 1. Comparing two data sources from this figure illustrates the trade off between data timeliness (the amount of lag time from an event to data about the event being available for interpretation by a public health department) and specificity/accuracy (the degree to which information with relevant patient characteristics is available). Over the counter (OTC) drug data have the benefit of timeliness; for example, people who are beginning to have symptoms of an intestinal illness may purchase OTC drugs a day or two before presenting at the emergency department (ED) or other health care facility. However, OTC sales data are not very specific; while a rise in sales of OTC antidiarrhoeal remedies may be the result of an increase in incidence of diarrhoeal illness, it may be unrelated to disease incidence. An increase in OTC sales may be explained by store specials, hoarding behaviours, or duplicate data transfers. It would be difficult to make a public health decision based solely on increased OTC sales because it does not contain any patient specific data. Involving pharmacists or pharmacy managers in initial models and categorisation of OTC data may increase the correlation of OTC sales data and possible human disease. Timeliness of data acquisition is only valuable if the signal is accurate and specific enough to inform public health decision making.

The timeliness and specificity of OTC data may be compared with ED chief complaint data. An increase in chief complaints of diarrhoea is an unambiguous signal that the incidence of intestinal illness has increased. It is less timely than the OTC data but yields much more detailed information. None the less this information does not distinguish between a natural and expected increase, a natural outbreak, a result of a bioterrorism event, or a coincidental, simultaneous increase in intestinal illnesses of varying aetiologies. It would still be difficult to make a public health decision based on these data alone without further investigation.

An increase in sales of OTC antidiarrhoeal drugs occurring concurrently with spikes in other data such as increases in worker absenteeism, and increases in ED chief complaints of diarrhoea would be a more reliable indicator of a true increase in disease incidence than any of these signals on their own. Developing a mechanism for evaluating data from disparate sources and of varying relevance to public health response is one of the challenges in finding practical applications for syndromic surveillance information.[3]

The accuracy of aberration detection signals generated by a syndromic surveillance system is also greatly affected by the amount of baseline data that are available. Studies reporting the use of short term, or ''drop-in'' syndromic surveillance systems set up specifically for outbreak detection during a high profile event consistently cite the lack of baseline data as a factor that hampered the determination of an appropriate signal threshold for these systems, often leading to systems that were overly sensitive and did not account for seasonal or other temporal factors.[15 29 30] Before a syndromic surveillance system is instituted for use as a regular surveillance tool, individual jurisdictions should ensure that at least one year of historical data are available for use as a reference for signal threshold determination and signal investigation.

SYNDROMIC SURVEILLANCE EVALUATIONS

Retrospective evaluations

Retrospective analyses show that syndromic surveillance systems may provide information that would allow public health departments to predict outbreaks earlier than by using traditional surveillance. For example, in the Milwaukee, Wisconsin cryptosporidiosis outbreak of 1993,[31] newspaper reports of over the counter antidiarrhoeal drugs being sold out at pharmacies were one of the first clues that an unusual event had occurred.[7] Calls to nurse hotlines showed a fourfold increase in the standard deviation of diarrhoea rates one day before observation of unusual activity by local pharmacists and five days before the local public health department was notified of the possible outbreak.[5] Similarly, a retrospective review of pharmacy records showed that sales of over the counter drugs increased fivefold during an outbreak of cryptosporidiosis in North Battleford, Saskatchewan.[7 32] During a 1983 outbreak of waterborne campylobacteriosis in Florida, pharmacy sales of antidiarrhoeals increased eightfold when compared with the same period in the previous year.[16] Retrospective analysis of Canadian outbreaks of *Cryptosporidium*, *E coli* O157:H7, and *Campylobacter* also all showed increases in OTC sales that corresponded to increases in disease incidence.[33] In New York City, retrospective analyses of ambulance dispatch data and ED data showed that they were effective in predicting city wide respiratory, gastrointestinal, and influenza outbreaks.[25 34]

Simulations

Simulated outbreaks can examine the ability of syndromic surveillance systems to accurately predict outbreaks while minimising false alarms.[34 35] According to most simulations, detection algorithms used in syndromic surveillance can best identify large, geographically widespread increases in disease. Syndromic surveillance detection algorithms are less successful at identifying small counts or small increases in disease, showing that syndromic surveillance may augment traditional disease surveillance but may not improve timeliness or sensitivity. Simulations of syndromic surveillance suggest that it is best suited for detecting diseases that have a narrow incubation period distribution, a steep epidemic curve, a long prodromal phase, are not included on routine diagnostic tests, and do not have a specific disease identifying clinical feature.[36]

To date no simulations of waterborne disease outbreaks have been conducted. Stoto *et al* tested a hypothetical syndromic surveillance system by simulating a ''fast'' outbreak (a disease that would have a steep epidemic curve) and a ''slow'' outbreak (a disease with a more gradual epidemic curve) of influenza-like illness (ILI) using three years of data from an ED.[37] Results of this exercise show that an episode of

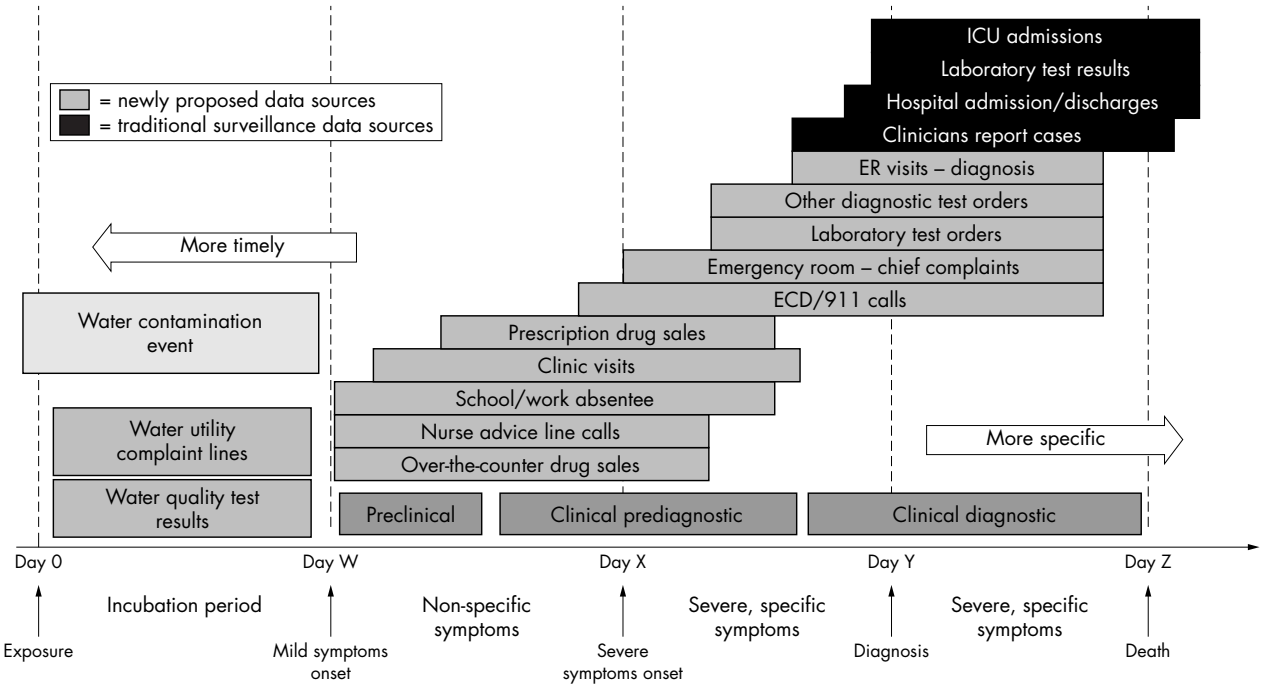


Figure 1 Surveillance data sources and health seeking behaviour.

either ''fast'' or ''slow'' ILI during the annual influenza season was undetectable by syndromic surveillance when set to a 1% false positive rate. During the non-flu (RSI) season syndromic surveillance algorithms were successful at detecting a ''fast'' outbreak by day three. However, the ''slow'' outbreaks were much harder to detect. Only one algorithm had a 50% probability of detecting this outbreak by day nine. By day nine, such outbreaks would usually be detected through traditional surveillance systems. As Stoto *et al* point out, ''The sudden appearance of large number of ILI cases…five times the daily average…especially in the summer- is clearly exceptional and ER physicians do not require any sophisticated statistical algorithm to tell them so''.[37]

Prospective reports

Several prospective evaluations of long term surveillance systems[18 23 34 38–43] have been published. Syndromic surveillance has been in use in New York City since 1995 to detect outbreaks of waterborne diarrhoeal disease. The waterborne disease surveillance system uses data from OTC antidiarrhoeal sales, ED chief complaint logs, reports from sentinel nursing homes, and reports of positive stool samples from clinical laboratories. Prospective studies evaluating this system between 2001 and 2004[34 40] reviewed citywide signals and smaller spatial clusters to find out if signals represented real outbreaks, and whether outbreaks detected by traditional surveillance were also detected by the syndromic system. These studies found that, 75% of the citywide outbreak signals correlated with true citywide viral gastroenteritis outbreaks. However, there was no correlation between smaller spatial clusters detected by the syndromic surveillance system and gastrointestinal (GI) outbreaks detected by traditional surveillance. Additionally, there were 36 GI outbreaks detected by traditional surveillance, none of which were detected by the syndromic surveillance system. Despite the extensive use of staff resources, the availability of fully electronic and timely data, and the ability to quickly initiate responses to various alerts, the syndromic system did not appreciably improve timeliness or sensitivity to overall disease surveillance.[40] A system in Connecticut that tracks hospital admissions shows similar discrepancies between GI outbreaks detected by traditional and syndromic surveillance.[42] This system generated 35 GI illness alarms in the course of one year, only one of which turned out to be a true GI outbreak. Additionally, none of 15 GI illness clusters detected by laboratory surveillance were detected by the syndromic surveillance system.

The University of Maryland implemented a syndromic surveillance system in its university hospital and evaluated its system's ability to detect actual clusters of patients with GI or respiratory symptoms presenting to the hospital from 2001 to 2002.[39] This system incorporated admission, discharge and transfer data as well as laboratory information system data. The evaluation showed that a peak in ED visits and an increase in stool test orders corresponded to a cluster of patients who later tested positive for infections with *Shigella spp*. The authors found the system to be timely and sensitive for their needs. A similar system is operated by the Department of Defense (DoD).[23] The DoD system detected three outbreaks of diarrhoeal diseases in three different locations, one of which was laboratory confirmed to be caused by rotavirus. A system based on ED chief complaint data administered by the Westchester County Department of Health was less specific. Over the course of 277 days, 59 signals were detected, none of which corresponded to an outbreak of a communicable disease.[43] A multi-jurisdictional system implemented for the 2002 Olympic games in Utah tracked encounters at urgent care centres and EDs using the real-time outbreak and disease surveillance system.[18] During the two month Olympic event the system's detection algorithms were exceeded twice; neither of these alarms corresponded to a real outbreak. The system was successfully maintained after the end of the Olympic event.

A study by Henry *et al* quantified the sensitivity, specificity, and positive predictive value (PPV) of the Kaiser Mid-Atlantic Region nurse hotline syndromic system by comparing the level of concordance between syndrome assignments based on the nurse hotline algorithm and the diagnosis made based on a subsequent Kaiser office visit. The authors found

Table 1 Characteristics of potentially waterborne disease agents

Agent	Incubation period[50] mean (range)	Disease identifying feature?[50]	Included on routine test?	Prodrome[50]
Parasitic				
*Cryptosporidium**	7 days (1–12 days)	No symptoms often include watery diarrhoea and abdominal pain, and may include general malaise, fever anorexia and vomiting	No. Not included on routine O and P stool examination.	None in adults; children may experience a prodrome of vomiting and anorexia.
*Cyclospora**	1 week	No. Symptoms include watery diarrhoea, and abdominal pain, nausea, anorexia, fatigue weight loss.	No. Not included on routine O and P stool examination	None.
Giardia	7–10 days (3–25 days)	No. Symptoms include watery diarrhoea, and abdominal pain, bloating, flatulence, fatigue weight loss, fatty bad smelling stool.	Yes. Included on standard O and P stool examination.	None.
Entamoeba	2–4 weeks (few days-years)	Sometimes. Symptoms range from mild diarrhoea to dysentery. Amebic granulomata, colitis, and liver lung, or brain abscess may occur.	Yes. Included on standard O and P stool examination.	None
Bacterial				
Campylobacter	2–5 days (1–10 days)	No. Symptoms include diarrhoea, abdominal pain, general malaise, fever, and vomiting.	Yes. Included on standard stool culture.	None
Enterohemorrhagic *E coli**	3–4 days (2–10 days)	Sometimes. Symptoms range from mild diarrhoea to dysentery. Fever is rare. Haemolytic uraemic syndrome and thrombotic thrombocytopenic purpura can develop.	No. Enterohemorrhagic strains require specific stool culture.	None
Enteropathogenic *E coli**	9–12 hours	Usually affects infants. Symptoms include watery diarrhoea and fever.	No. Requires specific stool culture.	None
Legionella[*51 52]	Legionnaires' disease: 5–6 days (2–10 days); Pontiac fever: 24–48 hours (5–66 hours)	No. Initial clinical presentation does not differ from other pneumonias. Symptoms include anorexia, malaise, myalgia, headache, fever, cough, abdominal pain, diarrhoea. In Legionnaires' disease these can proceed to respiratory failure and death.	No.	None.
Shigella	1–3 days (12 hours–1 week)	No. Symptoms depend on strain and range from mild to watery diarrhoea to dysentery and can include fever, nausea, and sometimes vomiting, and cramps.	Yes. Included on standard stool culture.	None.
Vibrio cholerae	2–3 days (Few hours-5 days)	Yes. Symptoms include sudden and profuse watery diarrhoea, nausea, vomiting, and rapid dehydration in some cases.	No. Special media required.	None.
*Vibrio parahaemolyticus**	12–24 hours (4–30 hours)	No. Symptoms include watery diarrhoea, abdominal pain, nausea, fever, and vomiting. Dysentery sometimes develops.	No. Special media required.	None.
*Vibrio vulnificus**	12–72 hours	Yes. Septicaemia, shock, hypotension, bullous skin lesions.	No. Special media required.	None.
Viral				
Hepatitis A*	28–30 days (15–50 days)	Yes. Symptoms include fever, anorexia, nausea, abdominal pain, jaundice.	Routine once hepatitis is suspected.	Yes. Prodrome before onset of jaundice usually lasts a few days.
Norovirus (Norwalk and Norwalk-like viruses)*	24–48 hours (10–50 hours)	No. Symptoms include nausea, vomiting, fever, diarrhoea, abdominal pain, fever.	No.	None.
Rotavirus*	24–72 hours	No. Symptoms include vomiting fever and watery diarrhoea, and most often affect children under 5 years old.	No.	None.

*May be good candidates for detection by syndromic surveillance. Criteria include: (1) narrow incubation period; (2) steep epidemic curve, long prodromal phase; (3) no clinical disease specific identifying feature; (4) disease not included in routine diagnostic tests.

Table 2 Potential syndromic surveillance data sources

	Lag time between exposure and information receipt	Validity Sensitivity Specificity PPV Other	Data quality and sustainability Accuracy Representativeness *Completeness*	System sustainability Costs Personnel Equipment	Other (pros and cons)
Automated and electronically available					
OTC surveillance available through NRDM (part of RODS)	Very timely, data are available within ~24 hours of sale. Cases tend to self treat before visiting clinic.	No formal tests of validity have been done.	Data coverage in CA is 25%–33% of major pharmacies, this varies by county. Data source is not very detailed. Recent experience suggests that outbreaks picked up by traditional PH are missed (San Luis Obispo and Fresno counties).	May not be sustainable, data are gathered for an independent purpose, but the commitment of pharmacies may be variable, especially as time from last BT or other terrorist event increases.	Staff time used is minimal. Probably cannot be used for outbreak detection because the data are not detailed and personal identifiers are not available. Allows coverage of multi-jurisdictional area. Could potentially be used to measure extent of an outbreak once identified.
Not automated and electronically available					
Complaint calls to water utilities	This could be a timely data source.	Unknown	Most probably not very detailed. Could be sustainable, the information is gathered for an independent purpose. Data are often not available in electronic format. Personal identifiers/addresses are sometimes available.	Large amount of time would be needed to set up and maintain. Would place a burden on utilities many of which have limited staff.	This could strengthen the relationship between water utilities and public health departments may lead to cooperation on other issues.
Sentinel nursing home surveillance	This could be a timely data source.	Unknown	Unclear if this population accurately represents surrounding population. Data based on clinical lab results would be accurate.	Difficult to implement across jurisdictions. Resource intensive in terms of DPH and facility staff time and stool testing materials, lab time.	Presents an added burden to nursing home residents.
Nurse call lines	Data can be available within 24 hours for some systems.[41]	According to Henry *et al*, Sensitivity = 72% Specificity = 96% PPV = 37%[38]	Data could potentially be linked to personal identifiers, and could be quite accurate. Data would probably be available in electronic format. Could be sustainable, the information is gathered for an independent purpose.	Large amount of time would be needed to set up and maintain.	This could strengthen the relationship between hospitals, insurance, and public health departments and may lead to cooperation on other issues.
School/work absenteeism	This could be a timely data source.	Unknown	Most probably not very detailed, without names or reasons for absenteeism. If reasons are provided accuracy is very questionable.	Large amount of time would be needed to set up and maintain.	It is not clear which populations to target and how representative they would be. Many people commute to work or school.

that the nurse hotline system achieved the highest sensitivity and PPV for respiratory and GI illness syndromes.[38]

Staff and resources used

Performance evaluations often do not provide information about time and resources that are used in maintaining the system and investigating the alarms. In New York City, staff maintain the system, spending several hours a day, seven days per week to download and analyse data.[1] Investigation of false signals is a significant burden on staff resources.[37] An often mentioned solution is to investigate only those alarms that are maintained for two or more days and in two or more geographical areas. However, if the main goal of syndromic surveillance is to increase timeliness of outbreak detection and response then waiting two or more days to start an investigation would negate the timeliness benefit of the system.[44] Similarly, if, because of the low specificity of syndromic signals, local health departments require that syndromic signals are substantiated by specimen collection and laboratory confirmation before public health action may be taken, any potential advantages of timeliness are lost.[45] Decision makers must take into account the ability of the local public health system to respond to signals when considering the implementation of any syndromic surveillance system.

SYNDROMIC SURVEILLANCE USES FOR WATERBORNE DISEASE DETECTION

Criteria derived from simulations can be used to assess which waterborne diseases may be appropriate for identification by syndromic surveillance.[36]

(1) The disease should have a narrow incubation period distribution.

Incubation periods of waterborne disease agents vary widely depending on dose, host susceptibility, and other factors.

(2) The disease should have a steep epidemic curve and a long prodromal phase.

Most waterborne diseases do not have a prodromal phase and an outbreak will not necessarily have a steep epidemic curve.

(3) The disease should not have a specific disease identifying clinical feature.

Initial symptoms of most potentially waterborne diseases are non-specific and generally include diarrhoea and other GI distress. Most of these illnesses do not have a disease identifying clinical or historical feature that allows clinicians to pinpoint the cause before performing laboratory tests.[46]

(4) The disease should not be included in routine diagnostic tests.

Laboratory tests are not routinely ordered for many waterborne diseases,[46] and acute GI illnesses are generally under diagnosed and underreported.[47–49]

Table 1 lists characteristics of potentially waterborne pathogens. Because many waterborne diseases lack clinically identifying features and are not part of routine testing, they may be good candidates for detection by syndromic surveillance, based on the four criteria listed above. However, many of the diseases that fit these criteria, while good candidates for syndromic surveillance, are often commonly occurring and nor necessarily of high public health importance.

Waterborne disease surveillance in the San Francisco Bay Area

In the San Francisco Bay Area four counties receive water from a common water utility. With the exception of a multi-county cryptosporidiosis surveillance project, surveillance for potentially waterborne disease in the San Francisco Bay Area is conducted by each county separately. There is no formal, timely coordination of waterborne disease surveillance across county lines. A system with multi-jurisdictional disease monitoring capabilities could provide public health benefit by permitting early detection of a multi-county waterborne outbreak. Surveillance data captured from multiple jurisdictions and interpreted centrally may lead to earlier outbreak detection than data gathered and interpreted by staff in separate counties who may not be aware of disease incidence in neighbouring jurisdictions.

Syndromic surveillance data sources may potentially provide cross-jurisdictional data, information about the geographical scope of an outbreak once one was identified, and additional reassurance that an outbreak was not occurring. However, the practical utility of syndromic data sources is uncertain; their outbreak detection benefits are currently theoretical and will remain so until accurate electronic capture of data and signal detection algorithms are refined or data sensitivity increases. The benefits of a regional waterborne disease surveillance system must be weighed against the resources required to set up such a system and the true risk of a waterborne disease outbreak in the San Francisco Bay Area. Based on the absence of any known waterborne outbreaks in the history of the water utility, extensive watershed protection measures, and a protected water source located in a national park it would seem that the risk of a waterborne disease outbreak occurring in the San Francisco Bay Area is quite small.

Potential waterborne disease syndromic surveillance data sources

Table 2 lists the potential waterborne disease syndromic surveillance data sources along with indicators of data quality and utility. Potential data sources in the table are divided into data that are currently available and accessible in electronic format in the San Francisco Bay Area and data that are not currently automated or electronically available but could be useful once they became automated and electronic.

Compared with other options, OTC surveillance for waterborne disease is currently the most feasible source of syndromic surveillance data available in the San Francisco metropolitan area because of the relative ease of implementing an existing, nationally funded system. Nursing home surveillance entails large inputs in terms of health department and on-site staff and fiscal resources for specimen testing, data monitoring, and signal investigation. Water utility complaint, nurse call line, and school and worker absenteeism logs are currently not compiled electronically in a central location; setting up a system to electronically capture these data would entail considerable commitments of

What this paper adds

To date, little attention has been paid to the possibility of using syndromic surveillance for monitoring waterborne disease incidence. This review of the benefits and shortcomings of syndromic surveillance may prove useful to public health practitioners and planners who are considering approaches to improve traditional surveillance systems.

Policy implications

This review will help policy makers weigh the costs and benefits of implementing a syndromic surveillance system and will clarify the drawbacks and advantages of potential data sources.

will and resources from the department of public health, other city agencies, and private and public partnerships with water utilities, insurance providers, hospitals and clinics, and large employers in the San Francisco Bay Area. Mechanisms for data storage, sharing, and retrieval would need to be established for each partnership. Finally and most importantly, dedicated public health staff who would compile, manage, and analyse syndromic data on a regular basis, and respond to syndromic surveillance signals would be needed. Signal verification and response activities may include: (1) determining data import and aberration detection algorithm problems that may lead to erroneous signals (for example, duplicate data, batch transfers from certain institutions, miscoding of information at the point of entry, text-string search algorithms that are too specific or not specific enough, etc); (2) verifying the validity of the signal by looking for the presence of signals in other data sources; (3) if the signal is deemed to be composed of possible true cases, hospital logs and charts may need to be manually reviewed by hospital or public health department staff and a line list compiled for clinical and/or laboratory based case verification; (4) traditional outbreak investigation activities and application of interventions. Timeliness provided by a syndromic surveillance system can only be useful if all of the above functions are supported by and integrated into the activities of the local health department on a sustained basis.

It is possible that outbreaks of cryptosporidiosis, cyclosporiasis, legionellosis, hepatitis A, and others may be detected through monitoring of OTC sales or that a suspected outbreak may be confirmed or better characterised through the use of these data. If OTC surveillance is to be used primarily as a back up data source to traditional surveillance, staff time for checking the web based interface would not exceed 15 minutes per day. Signal investigation would require additional resources. While the usefulness of OTC monitoring for waterborne diseases is only theoretical, given the availability of multi-county data and the low amount of staff time and effort needed to monitor the data, utilisation and prospective evaluation of these data may be recommended for two purposes: (1) reassurance of the absence of a waterborne disease outbreak and (2) establishing baseline familiarity that may prove helpful in the event of a waterborne disease outbreak. OTC data need to be correlated with known outbreaks in the geographical area where surveillance is occurring to clarify the validity and representativeness of these data before they can be used for prospective outbreak detection.

CONCLUSION

Because the efficacy of syndromic surveillance is not proved, it remains unclear whether it is worth the investment of personnel and financial resources to implement. In addition to the issues discussed in this review, it is challenging to develop sensible response protocols for syndromic surveillance systems because the likelihood of false alarms is so high, and because information is currently not specific enough to enable more timely outbreak detection or disease control activities. Although most syndromic surveillance systems do not use personally identifiable information, there are also considerations of personal privacy and public comfort with health information storage and analysis. The data needs of a syndromic surveillance system must be weighed against the needs and public demands for patient privacy.

There are many theoretical benefits of syndromic surveillance, including potential timeliness, increased response capacity, ability to establish baseline disease burdens, and ability to delineate the geographical reach of an outbreak. However, in the absence of empirical evidence of its efficacy to mitigate the effects of natural or intentional waterborne disease outbreaks through earlier detection and response, syndromic surveillance should not be implemented at the expense of traditional public health activities, and it should not be relied upon as a principal waterborne disease outbreak detection tool at this time. For waterborne diseases, syndromic surveillance systems should continue to be assessed and implemented as a supplemental system to help develop better statistical methods and sensible alarm thresholds that can be applied and evaluated in practice.

Authors' affiliations

M Berger, J M Weintraub, San Francisco Department of Public Health, Environmental Health Section, USA
R Shiau, San Francisco Department of Public Health, Communicable Disease Control and Prevention Section

Funding: none.

Competing interests: none.

REFERENCES

1 **Heffernan R**, Mostashari F, Das D, *et al*. New York City syndromic surveillance system. Syndromic surveillance: reports from a national conference, 2003. *MMWR* 2004;**53**(suppl):25–7.
2 **Centers for Disease Control**. http://www.cdc.gov/epo/dphsi/syndromic.htm (accessed on 3 Oct 2005).
3 **Buckeridge D**, Graham J, O'Connor M, *et al*. Knowledge-based bioterrorism surveillance. *Proc AMIA Symp* 2002:76–80.
4 **Lombardo J**, Burkom H, Elbert E, *et al*. Systems overview of the electronic surveillance system for the early notification of community based epidemics (ESSENCE II). *J Urban Health* 2003;**80**(suppl):i32–41.
5 **Rodman J**, Frost F, Jakubowski W. Using nurse hotline calls for disease surveillance. *Emerg Infect Dis* 1998;**4**:329–32.
6 **Derby M**, McNally J, Ranger-Moore J, *et al*. A comparison of foodborne illness cases reported to a poison control center and public health department. Syndromic surveillance: reports from a national conference, 2004. *MMWR* 2005;**54**(suppl):35–40.
7 **Proctor ME**, Blair KA, Davis JP. Surveillance data for waterborne illness detection: an assessment following a massive waterborne outbreak of cryptosporidium infection. *Epidemiol Infect* 1998;**120**:43–54.
8 **Magruder S**, Lewis S, Najmi A, *et al*. Progress in understanding and using over-the-counter pharmaceuticals for syndromic surveillance. Syndromic surveillance: reports from a national conference, 2003. *MMWR* 2004;**53**(suppl):117–22.
9 **Foldy S**. Linking better surveillance to better outcomes. Syndromic surveillance: reports from a national conference, 2003, *MMWR* 2004;**53**(suppl):12–17.
10 **Beaudeau P**, Payment P, Bourderont D, *et al*. A time series study of anti-diarrheal drug sales and tap water quality. *Int J Environ Health Res* 1999;**9**:293–311.
11 **Hogan W**, Tsui F, Ivanov O, *et al*. Detection of pediatric respiratory and diarrheal outbreaks from sales of over-the-counter electrolyte products. *J Am Med Inform Assoc* 2003;**10**:555–62.
12 **Davies G**, Finch R. Sales of over-the-counter remedies as an early warning system for winter bed crises. *Clin Microbiol Infect* 2003;**9**:858–63.
13 **Sacks J**, Lieb S, Baldy L, *et al*. Epidemic campylobacteriosis associated with community water supply. *Am J Public Health* 1986;**76**:424–8.
14 **Wagner M**, Tsui F, Espino J, *et al*. National retail data monitor for public health surveillance. syndromic surveillance: reports from a national conference, 2003. *MMWR* 2004;**53**(suppl):40–2.
15 **Foldy S**, Biedrzycki PA, Barthell EN, *et al*. Syndromic surveillance using regional emergency medicine internet. *Ann Emerg Med* 2004;**44**:242–6.
16 **Beitel AJ**, Olson KL, Reis BY, *et al*. Use of emergency department chief complaint and diagnostic codes for identifying respiratory illness in a pediatric population. *Pediatr Emerg Care* 2004;**20**:355–60.
17 **Begier EM**, Sockwell D, Branch LM, *et al*. The National Capitol Region's emergency department syndromic surveillance system: Do chief complaint and discharge diagnosis yield different results? *Emerg Infect Dis* 2003;**9**:393–6.
18 **Gesteland PH**, Gardner RM, Tsui FC, *et al*. Automated syndromic surveillance for the 2002 winter Olympics. *J Am Med Inform Assoc* 2003;**10**:547–54.
19 **Silka PA**, Geiderman JM, Goldberg JB, *et al*. Demand on ED resources during periods of widespread influenza activity. *Am J Emerg Med* 2003;**21**:534–9.
20 **Lober WB**, Trigg LJ, Karras BT, *et al*. Syndromic surveillance using automated collection of computerized discharge diagnoses. *J Urban Health* 2003;**80**(suppl 1):i97–106.
21 **Flaherty J**, Gillam M. Correlation of West Nile virus infection with emergency department chief complaints by using a passive syndromic surveillance model. (Abstract). *MMWR* 2004;**53**(suppl):237.
22 **Paladini M**. Daily emergency department surveillance system-Bergen County, New Jersey. Syndromic surveillance: reports from a national conference, 2003. *MMWR* 2004;**53**(suppl):47–9.
23 **Lewis M**, Pavlin J, Mansfield J, *et al*. Disease outbreak detection system using syndromic data in the greater Washington DC area. *Am J Prev Med* 2002;**23**:180–6.

24 **Irvin C**, Nouhan P, Rice K. Syndromic analysis of computerized emergency department patients' chief complaints: an opportunity for bioterrorism and influenza surveillance. *Ann Emerg Med* 2003;**41**:447–52.
25 **Mostashari F**, Fine A, Das D, *et al*. Use of ambulance dispatch data as an early warning system for communitywide influenzalike illness, New York City. *J Urban Health* 2003;**80**(suppl):i43–9.
26 **Lazarus R**, Kleinman K, Dashevsky I, *et al*. Use of automated ambulatory-care encounter records for detection of acute illness clusters, including potential bioterrorism events. *Emerg Infect Dis* 2002;**8**:753–60.
27 **Reis BY**, Mandl KD. Syndromic surveillance: the effects of syndrome grouping on model accuracy and outbreak detection. *Ann Emerg Med* 2004;**44**:235–41.
28 **Day FC**, Schriger DL, La M. Automated linking of free-text complaints to reason-for-visit categories and International Classification of Diseases diagnoses in emergency department patient record databases. *Ann Emerg Med* 2004;**43**:401–9.
29 **Osaka K**, Takahashi H, Ohyama T. Testing a symptom-based surveillance system at high-profile gatherings as a preparatory measure for bioterrorism. *Epidemiol Infect* 2002;**129**:429–34.
30 **Cho J**, Kim J, Yoo I, *et al*. Syndromic surveillance based on the emergency department in Korea. *J Urban Health* 2003;**80**:i124.
31 **Mac Kenzie WR**, Hoxie NJ, Proctor ME, *et al*. A massive outbreak in Milwaukee of cryptosporidium infection transmitted through the public water supply. *N Engl J Med* 1994;**331**:161–7.
32 **Stirling R**, Aramini J, Ellis A, *et al*. Waterborne cryptosporidiosis outbreak, North Battleford, Saskatchewan, Spring 2001. *Can Commun Dis Rep* 2001;**27**:185–92.
33 **Edge V**, Pollari F, Lim G, *et al*. 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* 2004;**95**:446–50.
34 **Heffernan R**, Mostashari F, Das D, *et al*. Syndromic surveillance in public health practice, New York City. *Emerg Infect Dis* 2004;**10**:858–64.
35 **Kulldorff M**, Zhang Z, Hartman J, *et al*. Benchmark data and power calculations for evaluating disease outbreak detection methods. Syndromic surveillance: reports from a national conference, 2003. *MMWR* 2004;**53**(suppl):144–51.
36 **Buehler JW**, Berkelman RL, Hartley DM, *et al*. Syndromic surveillance and bioterrorism-related epidemics. *Emerg Infect Dis* 2003;**9**:1197–04.
37 **Stoto M**, Schonlau M, Mariano L. Syndromic surveillance: Is it worth the effort? *Chance* 2004;**17**:20–5.
38 **Henry J**, Magruder S, Snyder M. Comparison of office visit and nurse advice hotline data for syndromic surveillance—Baltimore-Washington, DC, Metropolitan Area, 2002. Syndromic surveillance: reports from a national conference, 2003. *MMWR* 2004;**53**(suppl):112–16.
39 **Peterson D**, Perencevich E, Harris A, *et al*. Using existing electronic hospital data for syndromic surveillance. *J Urban Health* 2003;**80**:i122–3.
40 **Balter S**, Weiss D, Hanson H, *et al*. Three years of emergency department gastrointestinal syndromic surveillance in New York City: What have we found? Syndromic surveillance: reports from a national conference, 2004. *MMWR* 2004;**54**(suppl):175–80.
41 **Cooper D**, Smith G, Baker M, *et al*. National symptom surveillance using calls to a telephone health advice service—United Kingdom, December 2001–February 2003. Syndromic surveillance: reports from a national conference, 2003, *MMWR* 2004;**53**(suppl):179–83.
42 **Dembek Z**, Carley K, Siniscalchi A, *et al*. Hospital admissions syndromic surveillance- Connecticut, Septemeber 2001–November 2003. Syndromic surveillance: reports from a national conference, 2003. *MMWR* 2004;**53**(suppl):50–2.
43 **Terry W**, Ostrowsky B, Huang A. Should we be worried? Investigation of signals generated by an electronic surveillance system- Westchester County, New York. Syndromic surveillance: reports from a national conference, 2003. *MMWR* 2004;**53**(suppl):190–5.
44 **Reingold A**. If syndromic surveillance is the answer, what is the question? *Biosecur Bioterror* 2003;**1**:77–81.
45 **Steiner-Sichel L**, Greenko J, Heffernan R, *et al*. Field investigations of emergency department syndromic surveillance signals- New York City. Syndromic surveillance: reports from a national conference, 2003. *MMWR* 2004;**53**(suppl):184–9.
46 **Centers For Disease Control and Prevention**. Diagnosis and management of food borne illness: a primer for physicians and other health care professionals *MMWR* 2004;**53**:RR–4.
47 **Hellard M**, Sinclair M, Fairley C, *et al*. An outbreak of cryptosporidiosis in an urban swimming pool: why are such outbreaks difficult to detect? *Aust N Z J Public Health* 2000;**24**:272–5.
48 **Kasowski E**, Gackstetter G, Sharp T. Foodborne illness: new developments concerning an old problem. *Curr Gastroenterol Rep* 2002;**4**:308–18.
49 **Townes J**. Acute infectious gastroenteritis in adults: seven steps to management and prevention. *Postgrad Med* 2004;**115**:11–18.
50 **Chin J**. *Control of communicable diseases manual*. 17th ed. Washington, DC: American Public Health Association, 2000.
51 **Mulazimoglu L**, Yu V. Can Legionnaires disease be diagnosed by clinical criteria? *Chest* 2001;**120**:1049–53.
52 **Fiore A**, Butler J. Detecting nosocomial Legionnaires' disease. *Infect Med* 1998;**15**:625–35.

Kirian and Weintraub *BMC Medical Informatics and Decision Making* 2010, **10**:39
http://www.biomedcentral.com/1472-6947/10/39

BMC
Medical Informatics & Decision Making

RESEARCH ARTICLE Open Access

Prediction of gastrointestinal disease with over-the-counter diarrheal remedy sales records in the San Francisco Bay Area

Michelle L Kirian*, June M Weintraub

Abstract

Background: Water utilities continue to be interested in implementing syndromic surveillance for the enhanced detection of waterborne disease outbreaks. The authors evaluated the ability of sales of over-the-counter diarrheal remedies available from the National Retail Data Monitor to predict endemic and epidemic gastrointestinal disease in the San Francisco Bay Area.

Methods: Time series models were fit to weekly diarrheal remedy sales and diarrheal illness case counts. Cross-correlations between the pre-whitened residual series were calculated. Diarrheal remedy sales model residuals were regressed on the number of weekly outbreaks and outbreak-associated cases. Diarrheal remedy sales models were used to auto-forecast one week-ahead sales. The sensitivity and specificity of signals, generated by observed diarrheal remedy sales exceeding the upper 95% forecast confidence interval, in predicting weekly outbreaks were calculated.

Results: No significant correlations were identified between weekly diarrheal remedy sales and diarrhea illness case counts, outbreak counts, or the number of outbreak-associated cases. Signals generated by forecasting with the diarrheal remedy sales model did not coincide with outbreak weeks more reliably than signals chosen randomly.

Conclusions: This work does not support the implementation of syndromic surveillance for gastrointestinal disease with data available though the National Retail Data Monitor.

Background

Syndromic surveillance has received much attention as a method for health departments to accelerate the detection of, the reaction to, or the confirmation of disease outbreaks [1,2]. After the publication of reports suggesting that monitoring over-the-counter drug sales might have given advance notice of the 1993 outbreak of cryptosporidiosis in Milwaukee [3-5], federal agencies began to make explicit recommendations that water utilities and health departments consider implementing over-the-counter syndromic surveillance for enhanced waterborne outbreak detection [6-8]. However, the ability of over-the-counter syndromic surveillance to enhance the detection of waterborne disease outbreaks has not been adequately demonstrated [9].

* Correspondence: michelle.kirian@sfdph.org
City and County of San Francisco. Department of Public Health, Environmental Health. 1390 Market Street suite 910, San Francisco, California 94102, USA

In the San Francisco Bay Area, drinking water is provided by the San Francisco Public Utilities Commission (SFPUC) to 2.4 million customers in four counties. Supported by the SFPUC, the San Francisco Department of Public Health's Water Epidemiology Program maintains regional, distribution system-wide cryptosporidiosis surveillance. To clarify the validity and representativeness of sales of over-the-counter diarrheal remedies available through the National Retail Data Monitor (NRDM) for prospective outbreak detection, we sought to determine if these data are related to known outbreaks of infectious gastrointestinal illness in the drinking water service area [10].

Methods

County and state agencies receive reports of individual gastrointestinal cases as well as infectious disease outbreaks. Title 17 of the California Code of Regulations mandates case reporting of specified diagnosed diseases

as well as outbreaks of any disease to local health departments by health care providers [11]. Health departments may also become aware of outbreaks through follow-up with individual reported cases, citizen complaints and other modes. The definition of an outbreak differs by disease but typically entails a group of related cases for which a common source is identified or suspected; outbreaks may include as few as two cases.

Reports of cases of gastrointestinal disease from 2001-2007 among residents were requested from each of the county health departments in the drinking water service area. Data were transmitted in electronic formats from three adjacent counties. Reports for each case included etiology, date of report to the health department, gender, age, city and county.

Electronic records of outbreak data for all three participating counties were provided by the California Department of Public Health which receives outbreak reports following county and state health department outbreak investigations. These data were combined and reconciled with electronic records and records which were manually extracted from paper files from two of the participating county health departments. For each outbreak, information on etiology, number of cases, date of symptoms onset for the first and last cases, affected counties, and whether the outbreak occurred in an institutional setting such as a nursing home was provided. Outbreaks of reportable diseases as well as outbreaks of diseases that are not reportable as listed in Title 17 were included. Individual cases reportable under Title 17 associated with any outbreak may be included in the diarrhea case dataset; however, sufficient information was not available to link the outbreak and case datasets. The Committee on Human Research at the University of California, San Francisco approved the study protocol.

Over-the-counter drug sales records were purchased from the NRDM [10]. Records for the years 2005-2007 were provided as an electronic file. Records for years 2003-2004 were downloaded using the NRDM web interface. NRDM over-the-counter drug sales records are divided into 18 categories based on common use, form and whether intended for adult or pediatric populations. NRDM drug categories are: diarrhea remedies, anti-fever adult, anti-fever pediatric, bronchial remedies, baby/child electrolytes, chest rubs, cold relief adult liquid, cold relief adult tablet, cold relief pediatric liquid, cold relief pediatric tablet, cough syrup adult liquid, cough adult tablet, cough syrup pediatric liquid, cough/cold, hydrocortisones, nasal product internal, throat lozenges, and thermometers. Sales are based on the number of units sold regardless of the package size. Daily total sales are available for both all units sold by category and units sold by category excluding units for which discounts or other promotions were offered during the reporting period. NRDM provides information on the number of stores enrolled and reporting; from 2005 through 2007 approximately 47% of the stores enrolled to report anti-diarrhea drug sales actually reported (number of stores enrolled per week: 1389 -1706; number of stores reporting: 592-836).

Our analysis variable was the proportion of non-promotional diarrhea remedy sales to sales of non-promotional drugs for all categories combined (Diarrheal Remedy Sales). Diarrheal remedies are products taken for the relief of diarrhea and include bismuth, attapulgite, subsalicylate, and loperaminde hydrochloride products. Sales records of diarrheal remedies were available for the entire study area from July 2003 through 2007. Proportion sales were used instead of counts to control for unknown confounders such as changes in store hours.

Diarrheal Remedy Sales, and gastrointestinal case and outbreak data were aggregated by week for analysis. Diarrheal Remedy Sales were aggregated by week of sale, cases by week of report to the health department and outbreaks by week of onset of first outbreak-associated case. Data were divided into three parts for model building, model validation, and forecasting.

We used methods developed by Box and Jenkins to build autoregressive integrated moving average (ARIMA) models [12]. Estimates of model parameters were obtained through the method of least squares. All analyses were performed using SAS version 9.1 (SAS Institute Inc., Cary, NC, USA). Using Proc ARIMA, following either pre-whitening or double pre-whitening, Diarrheal Remedy Sales were cross correlated with the number of diarrhea cases in the same week and with weekly counts lagged one to 19 weeks before and after.

The relationship between Diarrheal Remedy Sales and gastrointestinal outbreaks was examined graphically and through regression. Because a 2006 report by Edge and colleagues [13] suggested that over-the-counter drug sales are sensitive to viral infection, specifically Norovirus, Diarrheal Remedy Sales were compared to outbreaks of all etiologies combined and to outbreaks of Norovirus alone. Furthermore, as institutionalized populations, such as those in a nursing home, may not purchase drugs from over-the-counter drug vendors in the same way as the non-institutionalized population, analyses were repeated excluding outbreaks that occurred in an institutional setting. Diarrheal Remedy Sales univariate model residuals were regressed on the number of outbreaks and on outbreak-associated cases per week.

The univariate Diarrheal Remedy Sales ARIMA model was used to auto-forecast sales for 105 weeks with weekly model updating (one week ahead forecasting). Signals were generated when actual observations

exceeded the upper 95% confidence limit. An outbreak week was any week when one or more outbreaks started that week or prior to that week but ended that week or later. Model sensitivity was calculated as the number of outbreak weeks with a signal divided by the total number of outbreak weeks. Specificity was calculated as the total number of weeks without a signal and no detected outbreaks divided by the total number of weeks without an outbreak. Calculations were done with all outbreaks and repeated in subsets of only larger outbreaks with 50 or more or 100 or more cases. To evaluate if model derived alerts identified outbreak weeks more reliably than randomly chosen alerts, sensitivity and specificity calculations were repeated for three sets of randomly chosen dates.

Results

Diarrheal case data were fit with a first order autoregressive model and Diarrheal Remedy Sales with an integrated first order moving average model (Case ARIMA (1,0,0): parameter estimate 0.33, T-ratio 3.09; Diarrheal Remedy Sales ARIMA(0,1,1): 0.4, 4.42). Figure 1 presents time series plots of the outbreak -associated gastrointestinal cases, individual gastrointestinal cases, Diarrheal Remedy Sales and differenced Diarrheal Remedy Sales.

From July 2003 through December 2007, there were 233 gastrointestinal outbreaks (Table 1). Most reported outbreaks were caused by Norovirus or by an unknown etiology of which many were suspected of being Norovirus. More Norovirus outbreaks were reported in each of 2006 and 2007 than previous years. Norovirus outbreaks were also larger than outbreaks of other diseases with a mean number of outbreak-associated cases of 30. The largest outbreak was of Norovirus at 153 cases. Thirty percent of outbreaks occurred in an institutional setting.

In the forecasting period, January 1, 2006 to January 1, 2008, there were 154 outbreaks; 20 with 50 or more, three of these with 100 or more cases. Table 2 lists details for outbreaks with 50 or more cases. Table 1 provides the number and size of outbreaks by study period.

From 2004 through 2007 there were 11,536 reported gastrointestinal cases. The majority of cases were of campylobacteriosis, cryptosporidiosis, salmonellosis, giardiasis, shigellosis and amoebiasis (Table 1). More cases were reported among children under 5 than for any other age group; incidence of gastrointestinal illness was similar across other ages. Sixty-one percent of cases were male. Although we collected case data from 2001 through 2007, there were abrupt changes in reporting at the start of 2004; review of the average number of cases reported by day of the week and year showed consistent lower overall reporting by day for 2004 through 2007 as compared to earlier data potentially indicating a change in surveillance protocols. As these county level changes persisted in aggregated regional data, the data were restricted to 2004 and later for analysis.

From July 2003 through December 2007, the proportion of diarrheal remedy sales to total drug sales ranged from 0.016 to 0.083 with an average of 0.044 and standard deviation of 0.014. Sales of diarrhea remedies ranged from 1216 to 3512 unit sales per week with an average of 2435 and standard deviation 441.

No significant correlation at any lag was found between Diarrheal Remedy Sales and diarrheal cases (Figure 2). Furthermore, regression analysis of the Diarrheal Remedy Sales univariate model residuals did not reveal an association between the weekly number of outbreaks or outbreak-associated cases and Diarrheal Remedy Sales when all outbreaks data were included or when restricted to Norovirus and/or non-institutional outbreaks.

Four signals were generated by the Diarrheal Remedy Sales model (on the weeks of 6/11/06, 1/29/06, 10/15/06 and 6/10/07). Four of the 20 outbreaks with 50 or more and one of three with 100 or more cases started during or lasted through a week with a signal. The two outbreaks with 100 or more cases without signals were both non-institutional Norovirus outbreaks with steep epidemic curves. Sensitivity for all outbreaks and for outbreaks with 50 or more or 100 or more cases was low and specificity high (Table 3). The sensitivity and specificity of the model was identical to a random selection of three sets of four signals, further supporting the conclusion that any relationship between Diarrheal Remedy Sales and gastrointestinal illness is spurious.

Discussion

NRDM Diarrheal Remedy Sales did not predict outbreaks of gastrointestinal disease or correlate with individual cases of diarrheal illness. Signals generated by the Diarrheal Remedy Sales model did not coincide with outbreak weeks more reliably than signals chosen randomly.

To generate Diarrheal Remedy Sales signals we employed ARIMA modeling and forecasting. Time series modeling, including ARIMA, has a long history of use in econometrics and statistical quality control [14,15]. More recently it has been adopted by public health practitioners to model subjects such as influenza and hospital admissions, weather and suicides, and gun bans and homicides [16-18]. Time series modeling accounts for autocorrelation, trend and seasonality which when present in data can cause ordinary regression techniques to present spurious variance estimates and incorrect inference.

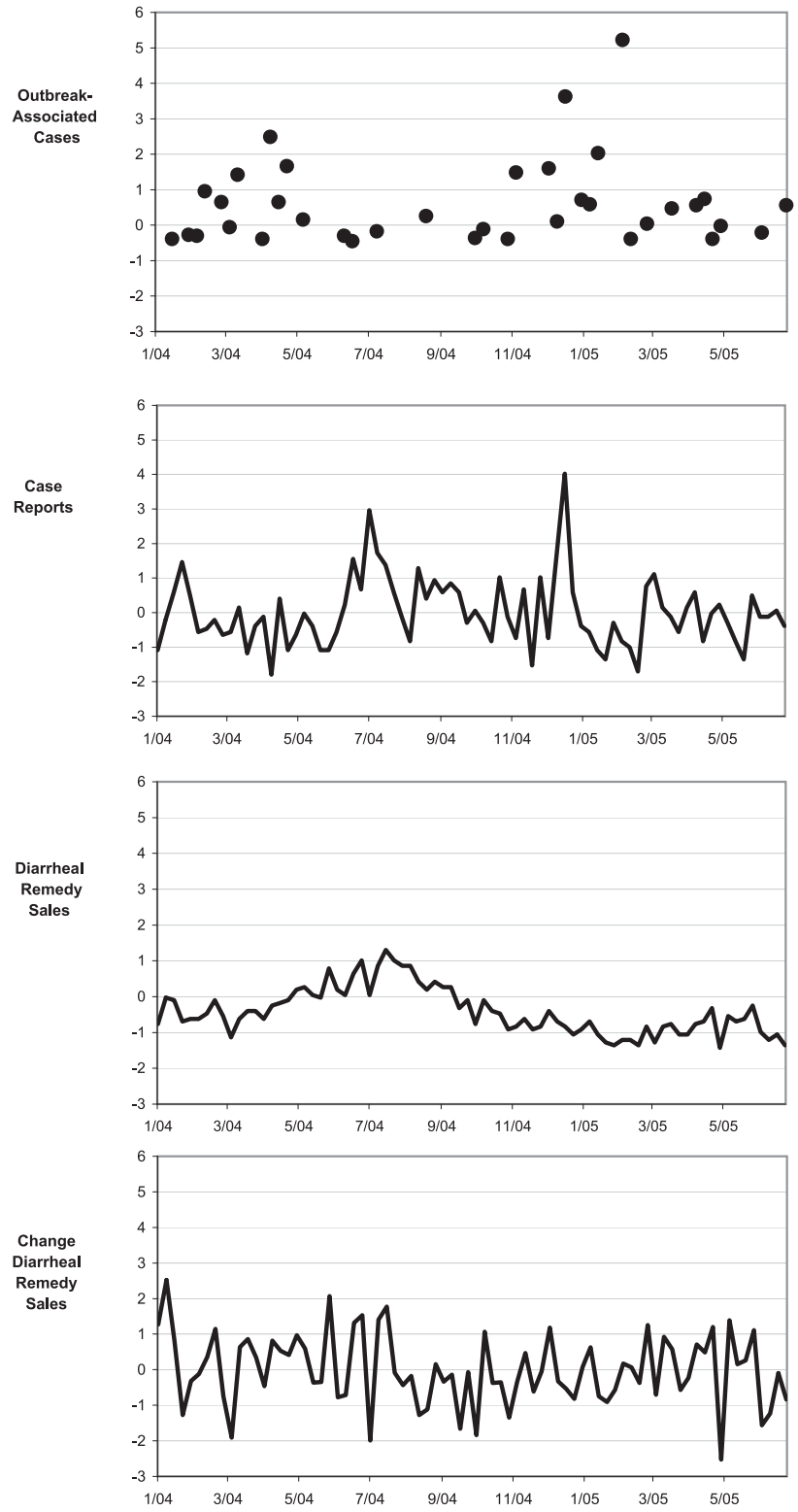


Figure 1 Plots of Outbreak-Associated Gastrointestinal Cases, Individual Gastrointestinal Cases, Diarrheal Remedy Sales, and Differenced Diarrheal Remedy Sales. Standardized weekly counts of gastrointestinal outbreak-associated cases, diarrheal illness case reports, Diarrheal Remedy Sales and differenced Diarrheal Remedy Sales in three San Francisco Bay Area Counties from January 2004 to July 2005. All data aggregated to the first Sunday of week. Diarrheal Remedy Sales are aggregated by week of sale, cases by week of report to the health department and outbreak cases by week of onset of the first associated case. Vertical axes are measured in standard deviations.

Table 1 Gastrointestinal Outbreak and Case Characteristics in the San Francisco Bay Area, July 2003 through December 2007

	N	Outbreak-Associated Cases*				N
		Maximum	Median	Mean		
All Disease Outbreaks	233	153	19	25	Case Etiology	
Outbreaks of Reportable Diseases	31	65	12	16	campylobacteriosis	3316
Outbreaks of Not-Reportable Diseases	202	153	21	27	cryptosporidiosis	2210
Study Period					salmonellosis	2187
Model (6/29/03 to 7/2/05)†	71	110	15	23	giardiasis	1739
Validation (7/3/05 to 12/31/05)	11	26	14	13	shigellosis	1002
Forecasting (1/1/06 to 12/30/07)	154	153	21	27	amoebiasis	512
Outbreak Etiology					hepatitis A	152
norovirus	144	153	24	30	Escherichia *coli* infection	151
unknown	41	80	18	21	vibriosis	118
salmonellosis	17	65	13	22	typhoid	62
Bacillus *cereus* /Clostridium *perfringens* infection	8	38	8	12	listeriosis	39
Escherichia *coli* infection	4	18	11	11	yersiniosis	26
scombroid poisoning	3	7	5	5	legionellosis	17
bacterial toxin poisoning	3	22	4	10	ciguatera poisoning	5
chemical toxin poisoning	2	4	3	3		
vibriosis	2	27	21	21		
ciguatera poisoning	2	3	3	3		
hepatitis A	1	2	2	2		
trichinosis	1	2	2	2		
cryptosporidiosis	1	16	16	16		
yersiniosis	1	1	1	1		
giardiasis	1	14	14	14		
rotavirus	1	6	6	6		
campylobacter	1	3	3	3		

*Number of cases not available for five outbreaks (3 salmonellosis, 2 norovirus) between January 1, 2006 and January 1, 2008.

† Modeling period was shorter –Jan 4, 2004 to Jul 2, 2005– for Univariate Case Modeling and Cross Correlation Analysis.

σ Outbreak and case data sets are not mutually exclusive or encompassing.

Over-The-Counter Anti-Diarrheal Drug Sales and Surveillance:

In the aftermath of the 1993 Milwaukee waterborne cryptosporidiosis outbreak in which thousands were sickened it was reported that sales of over-the-counter anti-diarrheal and anti-cramping drugs at one pharmacy increased by a factor of 17 to 20 as compared to the same period in the previous year [3]. This finding, supported by similar anecdotal reports stimulated the push for the implementation of waterborne disease surveillance with over-the-counter drug sales [19]. However, a later review of the feasibility and timeliness of surveillance data available during that outbreak–water treatment plant effluent turbidity logs, clinical laboratory diagnosis, nursing home diarrhea rates, hospital emergency room logs, random digit dialing telephone surveys, water utility complaint logs, school absentee logs and sales of anti-diarrhea drugs–revealed a poor response rate by pharmacies and a lack of timeliness [5].

A subsequent retrospective analysis of anti-nauseants and anti-diarrhea drug sales during waterborne outbreaks of cryptosporidiosis (Battlefords, Saskatchewan), and E. coli 0157:H7 infection and campylobacteriosis (Walkerton, Ontario), found that increased over-the-counter drug sales coincided with or lagged shortly behind illness onset [4]. The authors concluded that over-the-counter drug sales trends would provide a more timely and sensitive tool than monitoring hospital emergency department visits or traditional passive laboratory based surveillance. Nonetheless, over-the-counter drug sales data limitations were noted: data from only one of three pharmacies in Battlefords and one of six in Walkerton were available and formatted appropriately for analysis.

Studies of the seasonality of over-the-counter drug sales and diarrhea illness have also contributed evidence supporting over-the-counter drug sales for enhanced gastrointestinal surveillance. In an unidentified Canadian providence, sales of anti-nauseant and anti-diarrhea

Table 2 Gastrointestinal Outbreaks with 50 or More Cases in the San Francisco Bay Area, January 2006 through December 2007

Etiology	Cases	First Onset	Last Onset	Institutional	Over-the-Counter Drug ARIMA (0,1,1) Signal
Norovirus	101	1/25/06		Yes	Yes
Unknown/Norovirus	60	4/18/06		No	Yes
Unknown/Norovirus	62	4/24/06		Yes	Yes
Norovirus	107	4/25/06	5/2/2006	No	No
Unknown/Norovirus	55	4/26/06		No/Unknown	Yes
Norovirus	50	5/8/06		Yes	Yes
Norovirus	81	10/26/06		Yes	No
Norovirus	86	11/23/06		Yes	No
Norovirus	72	11/30/06		Yes	No
Norovirus	63	11/30/06		Yes	No
Unknown	80	12/7/06		Yes	No
Unknown	61	12/7/06		No/Unknown	No
Norovirus	76	1/3/07		Yes	No
Norovirus	60	1/8/07		Yes	No
Norovirus	92	7/13/07	7/17/07	No	No
Norovirus	153	8/3/07	8/17/07	No/Unknown	No
Norovirus	51	9/15/07	9/19/07	No/Unknown	No
Norovirus	52	12/20/07	1/1/08	Yes	No
Norovirus	52	12/22/07		Yes	No
Norovirus	76	12/22/07	1/15/08	No/Unknown	No

Abbreviations: ARIMA(0,1,1) First Order Integrated Moving Average

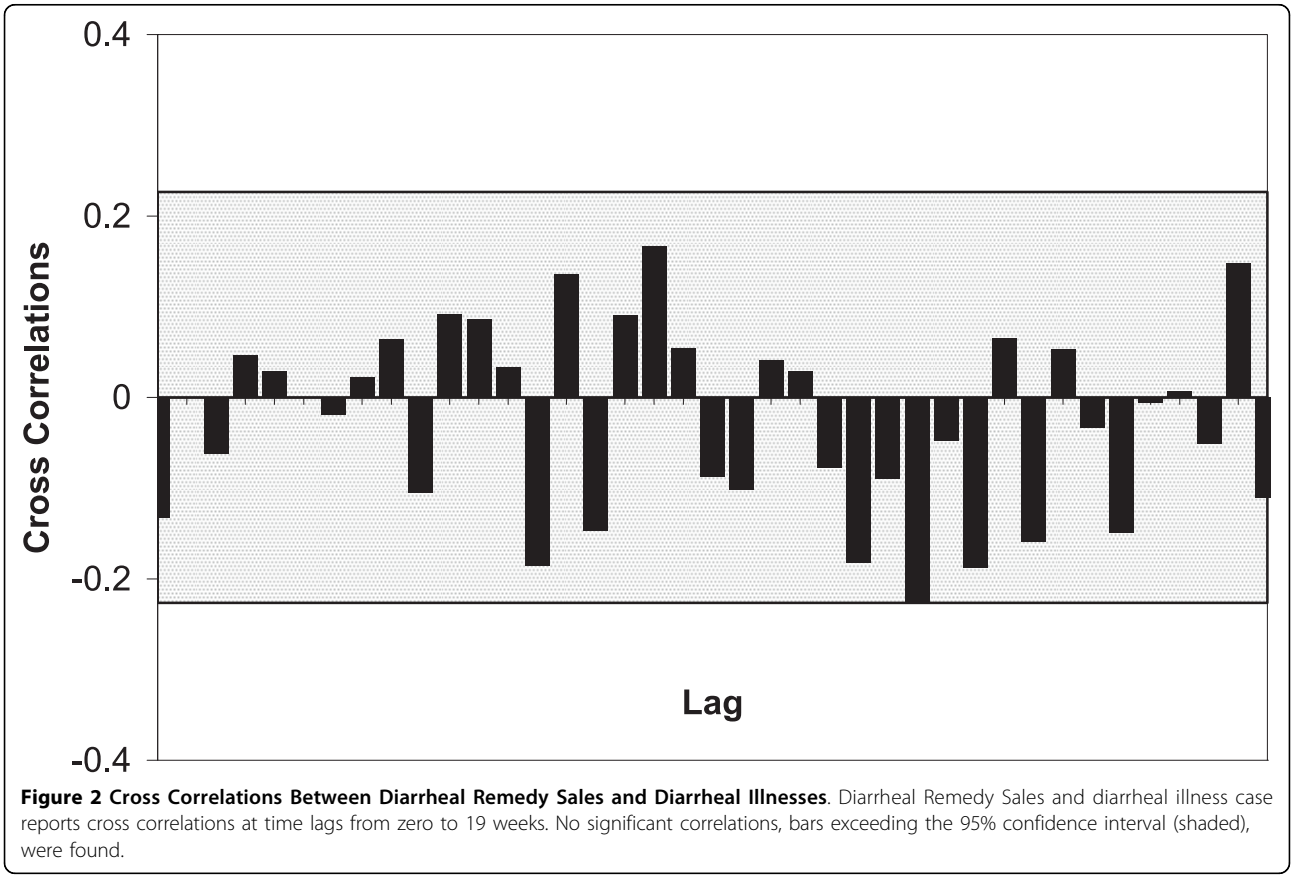


Figure 2 Cross Correlations Between Diarrheal Remedy Sales and Diarrheal Illnesses. Diarrheal Remedy Sales and diarrheal illness case reports cross correlations at time lags from zero to 19 weeks. No significant correlations, bars exceeding the 95% confidence interval (shaded), were found.

Table 3 Sensitivity and Specificity of Diarrheal Remedy Sales Model-Generated and Randomly Selected Signals

	All Outbreaks		Outbreaks with ≥ 50 Cases		Outbreaks with ≥ 100 Cases	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
Over-the-counter drug IMA(1) Signal (6/11/06, 1/29/06, 10/15/06, 6/10/07)*	4% (4/94)	100% (11/11)	4% (1/26)	97% (76/79)	14% (1/7)	97% (95/98)
Random Signals 1 (1/29/06, 6/11/06, 9/9/07, 2/18/07)*	4% (4/94)	100% (11/11)	8% (2/26)	97% (77/79)	14% (1/7)	97% (95/98)
Random Signals 2 (2/12/06, 7/9/06, 12/3/06, 9/23/07)*	3% (3/94)	91% (10/11)	4% (1/26)	97% (76/79)	0% (0/7)	96% (94/98)
Random Signals 3 (5/21/06, 1/3/07, 9/16/07, 12/9/07)*	3% (3/94)	100%(11/11)	4% (1/26)	97% (77/79)	0% (0/7)	97% (95/98)

* First day of week for each model generated and randomly generated signal.

over-the-counter drugs from one major retailer with 19 locations, accounting for only 12% of all pharmacies in the region, had similar seasonal temporality with reported Norovirus infections [13]. However, over-the-counter drug sales did not coincide with diarrhea due to other etiologies specifically bacterial or parasitic which are more prevalent during summer months. Similarly, electrolyte sales followed the same seasonal pattern as hospitalizations for selected pediatric diarrheal illness (Rotavirus and intestinal infections due to organisms not elsewhere classified (ICD9 008.61)) when combined with pediatric respiratory illnesses (Pneumonia, bronchopneumonia, influenza, bronchiolitis, respiratory syncytial virus) [20]. This study included very few diarrhea illness etiologies and the number of cases of each illness are not presented; the incidence of respiratory illnesses, especially seasonal influenza, is likely to greatly exceed that of diarrheal illnesses therefore obscuring the relationship between over-the-counter drug sales and diarrhea illness. The authors acknowledged that it is not possible to rule out a coincidental relationship which is driven by other phenomena.

Local and state health departments have implemented syndromic surveillance systems with over-the-counter anti-diarrhea drug sales monitoring components but few retrospective studies and no successful reports from ongoing surveillance projects are published [3,21,22]. Only one report, now antiquated, presents the progress of a functioning over-the-counter anti-diarrhea drug sales monitoring program. Similar to our results, Das et al (2005) reported that they had found no consistent relationship between over-the-counter anti-diarrhea drug sales and emergency department visits for gastrointestinal illness in New York City [22]. And, despite its availability nationwide for more than six years, no publications evaluate surveillance with NRDM over--the-counter diarrheal remedy drug sales in practice. One retrospective study presented graphs demonstrating the similar temporality of analgesic, anti-fever, anti-diarrhea and cough, and cold drugs combined and calls to the poison control center in 2003 [23]. Although our literature review did identify a number of reports suggesting that syndromic surveillance with over-the-counter anti-diarrheal drug sales could enhance traditional disease control activities, the widespread adoption of syndromic surveillance systems and the paucity of published reports on over-the-counter drug sales monitoring systems, and NRDM specifically, suggest publication bias may be present.

Limitations

Our study had several limitations. First, there were no large regional outbreaks in our dataset and the high data variability of diarrhea remedy sales may make it difficult to discern changes resultant from relatively small increases in illness. Although we do not believe that individual early health seeking behavior such as over-the-counter drug purchases would be different when an individual's illness is part of an undetected larger outbreak, in a large outbreak the number of people pursuing over-the-counter remedies might produce a signal that is significantly above the noise in the baseline.

Over-the-counter drug sales records as provided by the NRDM have several limitations. The usability of these data could be improved if participation by enrolled stores was increased or if meta-information on participating stores such as market coverage and on the drugs included in each category were made available. While we did not find any association between gastrointestinal disease and purchases of diarrheal remedies in general, it is possible that one product or a subset of products included in this category might have coincided with known disease. Furthermore, our study was not able to assess whether improvements in over-the-counter drug sales reporting systems might enhance the performance of this type of syndromic surveillance. The use of over-the-counter drugs sales for surveillance may be prohibitive due to the cost and logistics of data collection, or the proprietary and secret nature of the data [3].

County-by-county differences in disease reporting, and aggregations of diseases with varying severities may have masked finding a true association. These aggregations

could also have covered up localized Diarrheal Remedy Sales fluctuations resultant from isolated outbreaks. We therefore cannot rule out that county specific syndromic surveillance may be more sensitive than the region-wide surveillance examined in this analysis.

While studies show that over-the-counter drugs are the first option for many, health seeking behavior varies by factors including age, gender and culture [24-32]. One study examined healthcare-seeking behavior in response to diarrheal illness specifically. This survey of 351 adults reporting acute gastroenteritis (diarrhea, vomiting or both) found significant differences between those who use over-the-counter drugs and those who do not [24]. Although care should be exercised in applying these findings from Canada to the US as each have distinct health care systems, the lack of correlation that we found in our study between Diarrheal Remedy Sales and diarrheal cases could indicate that these data sources measure the occurrence of diarrhea in different populations. Similarly, high population mobility may increase the chances that Diarrheal Remedy Sales and cases are not both included in the region of study and that dispersed outbreaks may not be detected [33].

Conclusions

This study did not support the implementation of syndromic surveillance with National Retail Data Monitor Diarrheal Remedy Sales for enhanced gastrointestinal outbreak detection of waterborne or other origins. However, we cannot exclude the possibility that NRDM data maybe useful for detecting larger outbreaks.

A secondary finding of the study was of the increasing role of Norovirus in disease outbreaks in our region. From 2004 through 2006 approximately 56% of all outbreaks were due to Norovirus infection, 15% of these occurred in institutional settings. In 2007 the proportion attributable to Norovirus rose to 73%; 65% of outbreaks in 2007 were institutional. The increased incidence of outbreaks due to Norovirus may be attributable to enhanced detection or reporting; however, similar increases were noted in North Carolina, New York and Wisconsin [34]. Especially given the proven effectiveness of existing programs [35], public health departments must carefully evaluate the efficacy and added worth of surveillance systems to avoid the possibility that increased funding for programs such as syndromic surveillance are not accompanied by cutbacks in funding for programs such as institutional Norovirus prevention, resulting in a net increase in overall morbidity [36].

Abbreviations

The following abbreviations are used: Autoregressive Integrated Moving Average (ARIMA), National Retail Data Monitor (NRDM), and San Francisco Public Utilities Commission (SFPUC).

Acknowledgements

The authors thank Steve Allgeier and Chrissy Dangel of the U.S. Environmental Protection Agency, Water Security Division for their review and comments on this manuscript. Case and outbreak data were made available by the Santa Clara County Public Health Department, the San Mateo County Health Services Agency, the San Francisco Department of Pubic Health Communicable Disease Control Unit, and the California Department of Public Health.
This work was supported by a grant from the U.S. Environmental Protection Agency to the San Francisco Public Utilities Commission (H1-83380601-0).

Authors' contributions

JMW guided the study design, directed the implementation of the study, drafted portions of the manuscript, and critically reviewed the text. MLK designed and implemented the study, conducted the literature review, and was principal author of the text. JMW and MLK have read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Received: 29 March 2010 Accepted: 20 July 2010
Published: 20 July 2010

References

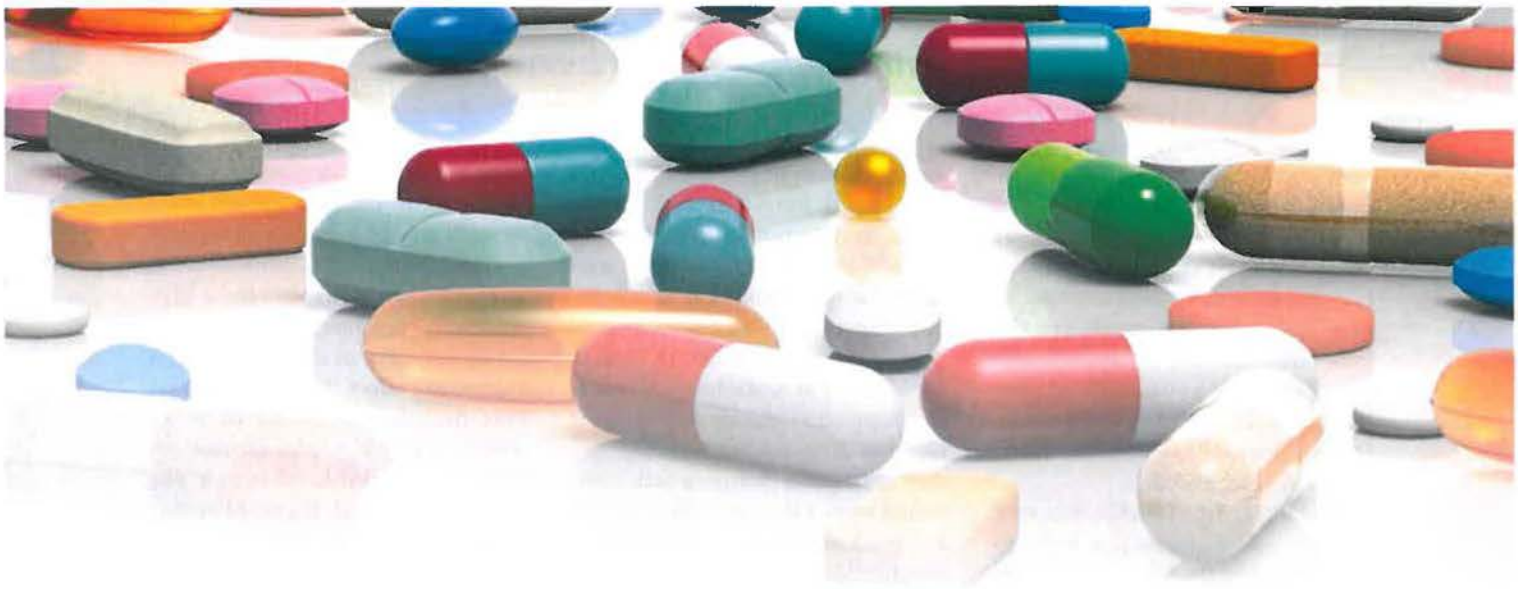
1. Buehler JW, Whitney EA, Smith D, Prietula MJ, Stanton SH, Isakov AP: **Situational uses of syndromic surveillance.** *Biosecur Bioterror* 2009, **7**:165-177.
2. Chretien JP, Tomich NE, Gaydos JC, Kelley PW: **Real-time public health surveillance for emergency preparedness.** *Am J Public Health* 2009, **99**:1360-1363.
3. Rodman JS, Frost F: **Pharmaceutical Sales–A Method of Disease Surveillance.** *J Environ Health* 1997, **60.**
4. Edge VL, Pollari F, Lim G, Aramini J, Sockett P, Martin SW, Wilson J, Ellis A: **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* 2004, **95**:446-450.
5. Proctor ME, Blair KA, Davis JP: **Surveillance data for waterborne illness detection: an assessment following a massive waterborne outbreak of Cryptosporidium infection.** *Epidemiol Infect* 1998, **120**:43-54.
6. **Recommendations Resulting From a Meeting Held November 2-4, 1999.** Balitmore, MD: National Drinking Water Advisory Council 1999.
7. ***Cryptosporidium* and Water: A Public Health Handbook.** Altlanta, GA: Centers for Disease Control and Prevention 1997.
8. **WaterSentinel System Architecture Draft Version 1.0.** Pennsylvania, PA: U. S. Environemental Protection Agency 2005, (Office of Ground Water and Drinking Water WSD ed., Draft, Version 1.0 edition.
9. Berger M, Shiau R, Weintraub JM: **Review of syndromic surveillance: implications for waterborne disease detection.** *J Epidemiol Community Health* 2006, **60**:543-550.
10. **National Retail Data Monitor, a public health surveillance tool.** [http://rods.health.pitt.edu/NRDM.htm].
11. **California Code of Regulations. Title 17 §2500.** State of California Office of Administrative Law 2010, **17.**
12. Box GEP, Jenkins GM: **Time Series Analysis: Forecasting and Control.** San Francisco, CA: Holden Day, 2 1976.
13. Edge VL, Pollari F, Ng LK, Michel P, McEwen SA, Wilson JB, Jerrett M, Sockett PN, Martin SW: **Syndromic Surveillance of Norovirus using Over-the-counter Sales of Medications Related to Gastrointestinal Illness.** *Can J Infect Dis Med Microbiol* 2006, **17**:235-241.
14. Hanssens D: **Bivariate time series analysis of the relationship between advertising and sales.** *Applied Econometrics* 1980, **12**:329-339.
15. Montgomery DC: **Introduction to Statistical Quality Control.** United States: John Wiley & Sons, 6 2009.
16. Ajdacic-Gross V, Lauber C, Sansossio R, Bopp M, Eich D, Gostynski M, Gutzwiller F, Rossler W: **Seasonal associations between weather conditions and suicide–evidence against a classic hypothesis.** *Am J Epidemiol* 2007, **165**:561-569.
17. Webster DW, Vernick JS, Hepburn LM: **Effects of Maryland's law banning "Saturday night special" handguns on homicides.** *Am J Epidemiol* 2002, **155**:406-412.

18. Upshur RE, Knight K, Goel V: **Time-series analysis of the relation between influenza virus and hospital admissions of the elderly in Ontario, Canada, for pneumonia, chronic lung disease, and congestive heart failure.** *Am J Epidemiol* 1999, **149**:85-92.
19. Angulo FJ, Tippen S, Sharp DJ, Payne BJ, Collier C, Hill JE, Barrett TJ, Clark RM, Geldreich EE, Donnell HD Jr, Swerdlow DL: **A community waterborne outbreak of salmonellosis and the effectiveness of a boil water order.** *Am J Public Health* 1997, **87**:580-584.
20. Hogan WR, Tsui FC, Ivanov O, Gesteland PH, Grannis S, Overhage JM, Robinson JM, Wagner MM: **Detection of pediatric respiratory and diarrheal outbreaks from sales of over-the-counter electrolyte products.** *J Am Med Inform Assoc* 2003, **10**:555-562.
21. Leach MJH, Evelyn T, McDonnell Dianna D, Morrow Scott: **Daily Situational Awareness Tool (DSAT): Utilizing real time infectious disease surveillance systems.** *APHA 135th Annual Meeting and Expo* Washington D.C 2007.
22. Das D, Metzger K, Heffernan R, Balter S, Weiss D, Mostashari F: **Monitoring over-the-counter medication sales for early detection of disease outbreaks–New York City.** *MMWR Morb Mortal Wkly Rep* 2005, **54**:41-46.
23. Krenzelok E, MacPherson E, Mrvos R: **Disease Surveillance and Nonprescription Can Predict Increases in Poison Exposure.** *J Med Toxicol* 2008, **4**:7-10.
24. Frosst GO, Majowicz SE, Edge VL: **Factors associated with the use of over-the-counter medications in cases of acute gastroenteritis in Hamilton, Ontario.** *Can J Public Health* 2006, **97**:489-493.
25. Espino DV, Lichtenstein MJ, Hazuda HP, Fabrizio D, Wood RC, Goodwin J, Stroup-Benham CA, Markides KS: **Correlates of prescription and over-the-counter medication usage among older Mexican Americans: the Hispanic EPESE study. Established Population for the Epidemiologic Study of the Elderly.** *J Am Geriatr Soc* 1998, **46**:1228-1234.
26. Stoehr GP, Ganguli M, Seaberg EC, Echement DA, Belle S: **Over-the-counter medication use in an older rural community: the MoVIES Project.** *J Am Geriatr Soc* 1997, **45**:158-165.
27. Amoako EP, Richardson-Campbell L, Kennedy-Malone L: **Self-medication with over-the-counter drugs among elderly adults.** *J Gerontol Nurs* 2003, **29**:10-15.
28. Farley D: **Label Literacy for OTC Drugs.** *FDA Consumer* U.S. Food and Drug Administration 1997.
29. Lam A, Bradley G: **Use of self-prescribed nonprescription medications and dietary supplements among assisted living facility residents.** *J Am Pharm Assoc (2003)* 2006, **46**:574-581.
30. Metzger KB, Hajat A, Crawford M, Mostashari F: **How many illnesses does one emergency department visit represent? Using a population-based telephone survey to estimate the syndromic multiplier.** *MMWR Morb Mortal Wkly Rep* 2004, **53**:106-111.
31. Qato DM, Alexander GC, Conti RM, Johnson M, Schumm P, Lindau ST: **Use of prescription and over-the-counter medications and dietary supplements among older adults in the United States.** *JAMA* 2008, **300**:2867-2878.
32. Vingilis E, Brown U, Hennen B: **Common colds. Reported patterns of self-care and health care use.** *Can Fam Physician* 1999, **45**:2644-2652.
33. **2000 Census Shows Big Gains in Public transportation Ridership to Bay Area Downtowns.** [http://www.mtc.ca.gov/news/press_releases/archive/rel263.htm].
34. **Norovirus activity–United States, 2006-2007.** *MMWR Morb Mortal Wkly Rep* Centers for Disease Control and Prevention 2007, **56**:842-846, 2007/08/25 edition. pp. 842-846.
35. Kirian ML, Meregillano G, Gennette D, Weintraub JM: **Multi-jurisdictional investigation of interactive fountain-associated cryptosporidiosis and salmonellosis outbreaks.** *Epidemiol Infect* 2008, **136**:1547-1551.
36. Reingold A: **If syndromic surveillance is the answer, what is the question?** *Biosecur Bioterror* 2003, **1**:77-81.

Pre-publication history

The pre-publication history for this paper can be accessed here:
http://www.biomedcentral.com/1472-6947/10/39/prepub

doi:10.1186/1472-6947-10-39
Cite this article as: Kirian and Weintraub: Prediction of gastrointestinal disease with over-the-counter diarrheal remedy sales records in the San Francisco Bay Area. *BMC Medical Informatics and Decision Making* 2010 **10**:39.



MICHELLE L. KIRIAN AND JUNE M. WEINTRAUB

Is syndromic surveillance of OTC drug sales effective in detecting outbreaks of waterborne GI disease?

AN EVALUATION OF THE ABILITY OF SALES OF OVER-THE-COUNTER DIARRHEAL REMEDIES TO PREDICT ENDEMIC AND EPIDEMIC GASTROINTESTINAL DISEASE FOUND NO SIGNIFICANT CORRELATIONS BETWEEN WEEKLY REMEDY SALES AND DIARRHEAL CASES OR OUTBREAK COUNTS.

Detection of waterborne disease is a basic concern of every provider of drinking water. Although routine water quality testing provides one means of verifying water quality, public health surveillance serves another potentially important role in ensuring safe drinking water. The water quality investigation in the 1993 cryptosporidiosis outbreak in Milwaukee, Wis., was initially prompted by reports of widespread work and school absenteeism (MacKenzie et al, 1995). That same year an investigation into the cause of several confirmed cases of *Salmonella typhimurium* led to the discovery of a contaminated municipal water supply in Gideon, Mo. (Angulo et al, 1997).

Traditional public health surveillance relies on case reporting by health professionals, doctors, and clinical laboratories. Each state mandates that cases of particular diseases be reported to local and/or state health departments so that disease trends can be followed and outbreaks identified early. In California, Title 17 of the California Code of Regulations lists more than 80 diseases and conditions that must be reported to the local health department where the case resides (CCR, 2009). Local health departments investigate the causes of disease and implement control measures.

Days, weeks, or even months may pass from the time an individual begins experiencing symptoms of disease to the time he or she goes to a clinic or doctor's office, and even more time may pass before a clinician may order diagnostic tests and report any laboratory-confirmed diagnosis to the health department. Because traditional disease surveillance is passive, relying on medical professionals to report cases as they occur, case reporting can be

slow and incomplete. Active surveillance, which includes automated laboratory reporting and routine contact with health care providers, can increase the completeness of case reporting and minimize the time lapse between a patient's visit to a doctor and notification of public health authorities of the confirmed illness (Buehler et al, 2009; Chretien et al, 2009).

Syndromic surveillance is a tool for outbreak detection that seeks to further reduce the time between the start of an individual's symptoms to the identification of a potential disease outbreak and initiation of an investigation into causes. Syndromic surveillance requires monitoring of behavioral patterns, symptoms, signs, or laboratory findings that precede diagnosis and correlate with illness in the population to a sufficient degree as to warrant action to protect the public health (Mandl et al, 2004). For waterborne disease, syndromic surveillance may use metrics that illustrate behavior changes that may occur before a sick individual visits a doctor, such as school and work absenteeism, water quality complaints, or sales of over-the-counter (OTC) drugs. Syndromic surveillance can also include measures of illness that may occur before an individual is actually diagnosed with a particular disease by a clinician, such as emergency room visits or calls to nurse or poison control hotlines (Berger et al, 2006).

The ability to implement a syndromic surveillance system depends on the existence or development of a means to collect, format, and analyze useful data (Rodman et al, 1997). OTC drug sales records are already available through the National Retail Data Monitor (NRDM), and for many jurisdictions, they represent an easily implemented source of syndromic surveillance data (NRDM, 2007; Berger et al, 2006). Although the use of syndromic surveillance is well-justified

Public Health Surveillance Data Sources

For many years, efforts to get an early handle on gastrointestinal (GI) disease outbreaks have relied on traditional surveillance methods to track outbreaks and cases and determine epidemiology. Advances in computer technologies and the growth of the Internet have led to the development of syndromic surveillance, which uses different types of data and various metrics for early detection of disease outbreaks. The following sections describe the types of reporting and data used by the two public health surveillance methods; this information has been adapted from a report by Berger and colleagues (2006) that provided a comparison of traditional surveillance and syndromic surveillance.

TRADITIONAL SURVEILLANCE RELIES ON REPORTING BY HEALTH PROFESSIONALS

Health professionals provide confirmed case reports. In traditional surveillance, doctors, laboratories, and other health professionals send reports on individual cases with clinical or laboratory diagnosis to the local health department. Reports include patient contact information and diagnosis details. Traditional surveillance does have its limitations, however. Generally, if no definitive diagnosis is made, a report is not sent, which may obscure the true number of cases. Furthermore, significant delays may occur between diagnosis and reporting. Traditional passive disease reporting may be enhanced with routine contact with medical professionals, active surveillance, or automated laboratory reporting.

SYNDROMIC SURVEILLANCE OBTAINS DATA FROM A VARIETY OF SOURCES

Over-the-counter (OTC) drug sales offer data that are timely but of indeterminate quality. OTC drug sales are available for purchase from the National Retail Data Monitor; data are available electronically within 24 hours of sales (NRDM, 2007). However, data are limited to total counts for predesigned categories of medications, and the market coverage of participating pharmacies is unknown and inconsistent. A moderate time commitment is required by health departments to develop models and follow up on signals. Anecdotal evidence supports the use of syndromic surveillance with drug sales, but low data quality and irregular patterns of self-medication may limit its true applicability.

Utility complaint call records may require data formatting but can have unexpected benefits. Water utilities are required to maintain customer complaint logs; however, use of these call records for surveillance may necessitate additional work to format the data. On the plus side, complaint logs may include personal identifiers, which are of use in signal investigations. An added advantage of implementation is that it may lead to enhanced relationships between utilities and health departments.

theoretically, there is a paucity of empirical evidence as to its efficacy for detection and early intervention for waterborne disease outbreaks.

METHODOLOGY EXPLAINED

Study focused on outbreaks in the San Francisco Bay area in California. The investigation described here explored the ability of OTC drug sales to predict gastrointestinal (GI) disease in the San Francisco Bay area. Because there were no known disease outbreaks in the region linked to drinking water contamination, this investigation included GI disease outbreaks from other and unknown causes. If a relationship were found between sales of OTC drugs and occurrence of any GI disease in the region, this might suggest that these data could be of use in predicting waterborne disease outbreaks more specifically. A lack of relationship between sales and disease would add important evidence to the body of literature on the efficacy of using surveillance systems that include these data.

The San Francisco Public Utilities Commission (SFPUC) provides an average of 240 mgd (910 ML/d) of water to approximately 2.4 million residents in the San Francisco Bay area. Surface water makes up 100% of the water supply. Although two filtration treatment plants treat water from local reservoirs, more than 80% of the water comes from the Hetch Hetchy Reservoir, 160 miles east of the city in the Sierra mountains. The water from this protected source is unfiltered. To ensure public health protection and to comply with the filtration waiver from the US Environmental Protection Agency, SFPUC funds systemwide cryptosporidiosis surveillance coordinated by the San Francisco Department of Public Health.

Cases and outbreak data included etiology, gender and age, and date of onset of symptoms. A case is an individual instance of disease, whereas an outbreak is a collection of related

---

Calls to nurse triage hotlines have been proposed as a potential source of surveillance data. One study found that nurse triage notes predicted outpatient visits by syndrome category with 72% sensitivity, 96% specificity, and a positive predictive value of 37% (Henry et al, 2004). Data from nurse call lines provide a benefit because they include personal identifying information that can be used to acquire additional information during outbreak investigations. As of this writing, the setup and maintenance of a surveillance system using nurse call logs are likely to be difficult, and clear benefit has yet to be demonstrated. However, modern electronic medical record systems and federal programs promoting the use of electronic health data to protect the public's health may facilitate surveillance with data originating in hospitals and medical offices when such a system is fully implemented (USDHHS, 2010).

School/work absenteeism records have demonstrated only limited use. School absenteeism was used to identify outbreaks of influenza during the 2009 flu season in Florida (Mann et al, 2011). However, limitations of the system include an inability to capture the reasons for absenteeism and incomplete coverage because only public schools are included. Furthermore, significant lags in data reporting may result from manual entry of data by public school administrators. As for work absenteeism, collection of these data is likely to be difficult and time-consuming because of the number and types of workplaces and the lack of a central reporting system.

Emergency department chief-complaint data are likely the most implemented and studied of the syndromic data sources. New York City has one of the longest operating systems of emergency department chief complaints, but the usefulness of this data source remains unproven. According to a 2009 report, a simulation drill with 42 patients designed to test the city's chief-complaint surveillance system failed to signal an alert for febrile respiratory illness (Wallace et al, 2009). An earlier study evaluating the same system found surveillance was not useful for detecting localized GI outbreaks (Balter et al, 2005). Syndromic surveillance with emergency department complaints proved most effective in detecting and monitoring annual citywide outbreaks of norovirus, rotavirus, and influenza. As with calls to nurse triage hotlines, full implementation of electronic medical record systems and federal programs to promote use of these data to protect public health should increase their utility.

Use of poison control center data has not been fully explored. The applicability of poison control calls for public health surveillance has not been thoroughly tested although some capacities have been incorporated by public health systems (Sutter et al, 2010). One poison control center reported the identification of a foodborne botulism outbreak with three cases (Brown et al, 2010). However, a study found that calls reporting GI symptoms to poison control centers did not coincide with cases of GI illness reported to the health department, which suggests that these data sources represented different populations (Derby et al, 2005).

cases for which a common source is identified or suspected. An outbreak may comprise as few as two cases. Case and outbreak data are neither mutually exclusive nor completely overlapping; not all cases are part of an outbreak, and not all outbreaks include cases that are reported individually.

For this study, cases of GI disease that occurred among residents of three counties served by the SFPUC were obtained from each of the respective county health departments in the drinking water service area. Reports for each case included etiology, date of report to the health department, gender, age, city, and county. Outbreak data for all three participating counties were provided by the California Department of Public Health. For each outbreak, information was provided on etiology, number of cases, date of onset of symptoms for the first and last cases, affected counties, and whether the outbreak occurred in an institutional setting such as a nursing home. The Committee on Human Research at the University of California, San Francisco, approved the study protocol.

Records of OTC sales of diarrheal remedies were obtained. Records of OTC drug sales were purchased from the NRDM (2007). Records for the years 2003–04 were downloaded using the NRDM web interface; records for the years 2005–07 were provided as an electronic file. NRDM OTC drug sales records are divided into categories based on the drug's common use and form and whether the drug is intended for adult or pediatric populations. The 18 categories are antidiarrheal, antifever adult, antifever pediatric, bronchial remedies, chest rubs, cold-relief adult liquid, cold-relief adult tablet, cold-relief pediatric liquid, cold-relief pediatric tablet, cough-adult tablet, cough-syrup adult liquid, cough syrup pedi-

TABLE 1 Gastrointestinal outbreak and case characteristics in the San Francisco Bay area, July 2003 through December 2007

	Outbreak-associated Cases*					
Parameter	***n***	**Maximum**	**Median**	**Mean**	**Parameter**	***n***
All disease outbreaks	233	153	19	25	Case etiology	
Outbreaks of reportable diseases	31	65	12	16	Campylobacteriosis	3,316
Outbreaks of not-reportable diseases	202	153	21	27	Cryptosporidiosis	2,210
Study period					Salmonellosis	2,187
Model (06/29/03 to 07/02/05)†	71	110	15	23	Giardiasis	1,739
Validation (07/03/05 to 12/31/05)	11	26	14	13	Shigellosis	1,002
Forecasting (01/01/06 to 12/30/07)	154	153	21	27	Amoebiasis	512
Outbreak etiology					Hepatitis A	152
Norovirus	144	153	24	30	*Escherichia coli* infection	151
Unknown	41	80	18	21	Vibriosis	118
Salmonellosis	17	65	13	22	Typhoid	62
Bacillus cereus/Clostridium perfringens infection	8	38	8	12	Listeriosis	39
Escherichia coli infection	4	18	11	11	Yersiniosis	26
Scombroid poisoning	3	7	5	5	Legionellosis	17
Bacterial toxin poisoning	3	22	4	10	Ciguatera poisoning	5
Chemical toxin poisoning	2	4	3	3		
Vibriosis	2	27	21	21		
Ciguatera poisoning	2	3	3	3		
Hepatitis A	1	2	2	2		
Trichinosis	1	2	2	2		
Cryptosporidiosis	1	16	16	16		
Yersiniosis	1	1	1	1		
Giardiasis	1	14	14	14		
Rotavirus	1	6	6	6		
Campylobacteriosis	1	3	3			

n—number

*Number of cases not available for five outbreaks (three salmonellosis, two norovirus) between Jan. 1, 2006, and Jan. 1, 2008
†Modeling period was shorter (Jan. 4, 2004, to July 2, 2005) for univariate case-modeling and cross-correlation analysis.

Outbreak and case data sets are not mutually exclusive or encompassing.

atric liquid, cough/cold, electrolytes pediatric, hydrocortisones, nasal product internal, thermometers, and throat lozenges.

Antidiarrheal remedies are defined as products that are taken for the relief of diarrhea and include bismuth, attapulgite, subsalicylate, and loperamide hydrochloride products. Sales are based on the number of units sold, regardless of the package size. Daily total sales are available for all units sold by category as well as units sold by category excluding units for which discounts or other promotions were offered during the reporting period. NRDM also provides information on the number of stores enrolled and reporting. From 2005 through 2007, approximately 47% of the stores enrolled to report antidiarrheal drug sales actually reported sales data (number of stores enrolled per week = 1,389–1,706; number of stores reporting per week = 592–836).

Statistical analysis incorporated time-series modeling. The analysis variable was the proportion of nonpromotional diarrheal remedy sales to sales of nonpromotional drugs for all categories combined (diarrheal remedy sales). Analyzing the proportion of sales rather than counts provided some control for secondary factors—such as changes in store hours—that could have contributed to changes in consumer purchasing.

Diarrheal remedy sales were aggregated by week of sale, cases by week of report to the health department, and outbreaks by week of onset of first outbreak-associated case. Data were divided into three parts: model building, model validation, and forecasting. The relationship between diarrheal remedy sales and GI outbreaks was examined graphically, through regression and through additional techniques (Kirian & Weintraub, 2010). Methods developed by other researchers were used to build autoregressive integrated moving average models of sales and individual case counts (Box & Jenkins, 1976). This type of time-series modeling has a long history of use in econometrics and statistical quality control (Montgomery, 2009; Hanssens, 1980). More recently, it has been adopted by public health practitioners to model subjects such as influenza and hospital admissions, weather and suicides, and gun bans and homicides (Ajdacic-Gross et al, 2007; Webster et al, 2002; Upshur et al, 1999). Time-series modeling accounts for autocorrelation, trend, and seasonality, which, when present in data, can cause ordinary regression techniques to present spurious variance estimates and incorrect inference.

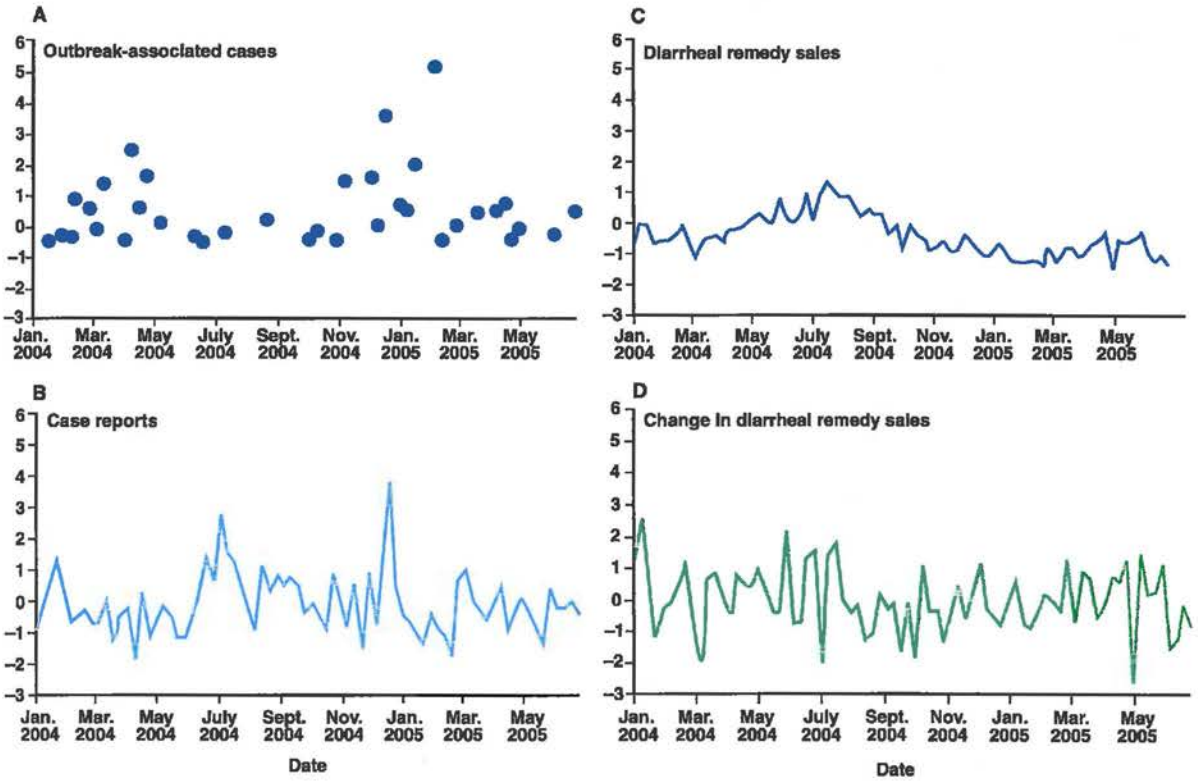
TABLE 2 Gastrointestinal outbreaks with 50 or more cases in the San Francisco Bay area in California, January 2006 through December 2007

Etiology	Cases *n*	Date of First Onset	Date of Last Onset	Institutional	OTC Drug ARIMA (0,1,1) Signal
Norovirus	101	01/25/06	NA	Yes	Yes
Unknown/norovirus*	60	04/18/06	NA	No	Yes
Unknown/norovirus	62	04/24/06	NA	Yes	Yes
Norovirus	107	04/25/06	05/02/2006	No	No
Unknown/norovirus	55	04/26/06	NA	Unknown	Yes
Norovirus	50	05/08/06	NA	Yes	Yes
Norovirus	81	10/26/06	NA	Yes	No
Norovirus	86	11/23/06	NA	Yes	No
Norovirus	72	11/30/06	NA	Yes	No
Norovirus	63	11/30/06	NA	Yes	No
Unknown	80	12/07/06	NA	Yes	No
Unknown	61	12/07/06	NA	Unknown	No
Norovirus	76	01/03/07	NA	Yes	No
Norovirus	60	01/08/07	NA	Yes	No
Norovirus	92	07/13/07	07/17/07	No	No
Norovirus	153	08/03/07	08/17/07	Unknown	No
Norovirus	51	09/15/07	09/19/07	Unknown	No
Norovirus	52	12/20/07	01/01/08	Yes	No
Norovirus	52	12/22/07	NA	Yes	No
Norovirus	76	12/22/07	01/15/08	Unknown	No

ARIMA (0,1,1)—first-order integrated moving average, *n*—number, NA—not available, OTC—over-the-counter

*Etiology was not conclusively determined but was suspected to be norovirus.

FIGURE 1 Plots of outbreak-associated gastrointestinal cases, individual gastrointestinal cases, diarrheal remedy sales, and differenced diarrheal remedy sales



Standardized weekly counts of gastrointestinal outbreak–associated cases, diarrheal illness case reports, diarrheal remedy sales, and differenced diarrheal remedy sales in three San Francisco Bay area counties of California from January 2004 to July 2005 were analyzed. All data were aggregated to the first Sunday of the week. Diarrheal remedy sales were aggregated by week of sale, cases by week of report to the health department, and outbreak cases by week of onset of the first associated case. Vertical axes are measured in standard deviations.

RESULTS HELPED ASCERTAIN PREDICTIVE ACCURACY

Most reported outbreaks were caused by norovirus. From July 2003 through December 2007, 233 GI outbreaks occurred in the study area; Table 1 provides the number and size of outbreaks by study period. Norovirus was identified as the cause of most reported outbreaks and was suspected to have caused most of the outbreaks with unknown etiology. The largest outbreak during this period was of norovirus at 153 cases. Of these outbreaks, 30% occurred in an institutional setting such as a nursing home. In the forecasting period from Jan. 1, 2006, to Jan. 1, 2008, in which sales data were used to predict outbreaks, there were 154 outbreaks. Of these, 20 outbreaks recorded 50 or more cases and 3 of the 20 outbreaks had 100 or more cases; Table 2 provides details for outbreaks with 50 or more cases.

During the study period of July 2003 through December 2007, 11,536 individual GI cases were reported. The majority of cases were of campylobacteriosis, cryptosporidiosis, salmonellosis, giardiasis, shigellosis, and amoebiasis (Table 1). More cases were reported among children younger than age five than for any other age group; incidence of GI illness was similar across other ages. Of the cases reported, 61% of individuals afflicted were male.

Diarrheal remedy sales analyzed in proportion to drug sales. The proportion of diarrheal remedy sales to total drug sales ranged from 0.016 to 0.083 with an average of 0.044 and standard deviation of 0.014. Sales of diarrheal remedies ranged from 1,216 to 3,512 unit sales per week with an average of 2,435 and standard deviation of 441.

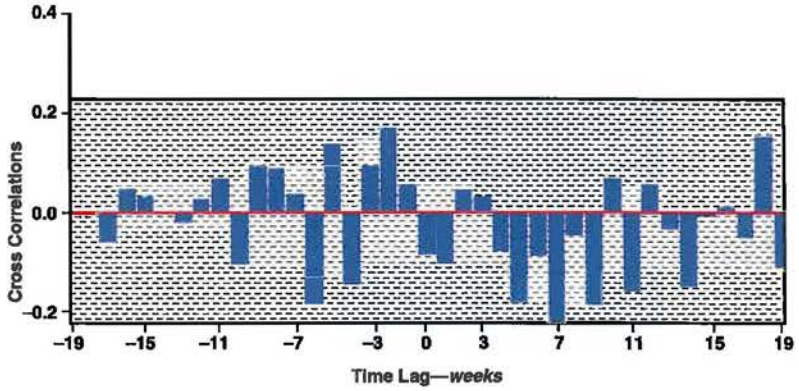
Analysis showed no significant correlation between sales and diarrheal cases. No significant correlation was found between sales in a particular week and occurrence of diarrheal cases in the same week; lags of up to 19 weeks between sales and cases were also investigated, and no significant correlations were found (Figures 1 and 2). Figure 2 shows time-series plots of the outbreak-associated GI cases, individual GI cases, diarrheal remedy sales, and differenced diarrheal remedy sales. Furthermore,

regression analysis of the diarrheal remedy sales univariate model residuals did not show an association between the weekly number of outbreaks or outbreak-associated cases and diarrheal remedy sales when all outbreaks data were included or when restricted to norovirus and/or noninstitutional outbreaks.

During the forecasting period, four signals were generated by the diarrheal remedy sales model, indicating a statistically significant increase in sales. The signals occurred during the weeks of Jan. 29, June 11, and Oct. 15, 2006, and June 10, 2007 (Figure 3). Although some of the outbreaks occurred during or lasted through these weeks, most did not. Of the 20 outbreaks with 50 or more cases, 4 occurred during a signal. Of the three outbreaks with 100 or more cases, one occurred during a signal week; the other two large outbreaks were both norovirus outbreaks that occurred outside of institutional settings.

Model did not confirm predictive abilities of sales. Sensitivity and specificity were calculated as a measure of predictive accuracy. Sensitivity was calculated as the number of outbreak weeks with a signal divided by the total number of outbreak weeks; specificity was calculated as the total number of weeks without a signal and no detected outbreaks divided by the total number of weeks without an outbreak. As shown in Table 3, sensitivity was low, and specificity was high. The sensitivity and specificity of the model were identical to a random selection of three sets of four signals, further supporting the conclusion that any relationship between diarrheal remedy sales and GI illness is spurious.

FIGURE 2 Cross correlations between diarrheal remedy sales and diarrheal illnesses



Diarrheal remedy sales and diarrheal illness case reports cross correlations are shown at time lags from zero to 19 weeks. No significant correlations (indicated by bars exceeding the shaded 95% confidence interval) were found.

IMPLEMENTATION OF SYNDROMIC SURVEILLANCE MAY BE UNSUPPORTED

In the current study, NRDM diarrheal remedy sales did not predict outbreaks of GI disease or correlate with individual cases of diarrheal illness. Signals generated by the diarrheal remedy sales model did not coincide with outbreak weeks more reliably than signals chosen randomly.

Review of the literature uncovers limitations. This authors' finding of no association was consistent with much previous work. Most studies that have examined relationships between outbreaks and OTC drug sales have been forthcoming about the limitations of the study designs and advised caution when using their conclusions

TABLE 3 Sensitivity and specificity of diarrheal remedy sales model-generated and randomly selected signals

	All Outbreaks		Outbreaks With ≥ 50 Cases		Outbreaks With ≥ 100 Cases	
Signal	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
OTC drug IMA (1) signal (06/11/06, 01/29/06, 10/15/06, 06/10/07)*	4% (4/94)	100% (11/11)	4% (1/26)	97% (76/79)	14% (1/7)	97% (95/98)
Random signals 1 (01/29/06, 06/11/06, 09/09/07, 02/18/07)*	4% (4/94)	100% (11/11)	8% (2/26)	97% (77/79)	14% (1/7)	97% (95/98)
Random signals 2 (02/12/06, 07/09/06, 12/03/06, 09/23/07)*	3% (3/94)	91% (10/11)	4% (1/26)	97% (76/79)	0% (0/7)	96% (94/98)
Random signals 3 (05/21/06, 01/03/07, 09/16/07, 12/09/07)*	3% (3/94)	100% (11/11)	4% (1/26)	97% (77/79)	0% (0/7)	97% (95/98)

IMA—integrated moving average, OTC—over-the-counter

*First day of week for each model-generated and randomly generated signal

to support policy. For example, in the aftermath of the 1993 Milwaukee waterborne cryptosporidiosis outbreak in which thousands were sickened, it was reported that sales of OTC antidiarrheal and anticramping drugs at one pharmacy increased by a factor of 17 to 20 compared with the same period in the previous year (Rodman et al, 1997). This finding, supported by similar anecdotal reports stimulated the push for the implementation of waterborne disease surveillance with OTC drug sales (USEPA, 2005; NDWAC, 1999; CDC, 1997). However, Proctor and co-workers (1998) reviewed the feasibility and timeliness of surveillance data available during that outbreak, including water treatment plant effluent turbidity logs, clinical laboratory diagnoses, nursing home diarrheal rates, hospital emergency room logs, random digit dialing telephone surveys, water utility complaint logs, school absentee logs, and sales of antidiarrheal drugs. They found that the surveillance data showed a poor response rate by pharmacies and a lack of timeliness. The study concluded that because "... no single set of recommended surveillances will be applicable to all communities, a combination of surveillance options should be developed locally drawing on existing disease surveillance methods ..." (Proctor et al, 1998).

A subsequent retrospective analysis of antinauseants and antidiarrheal drug sales during waterborne outbreaks of cryptosporidiosis (Battlefords, Sask.) and *Escherichia coli* 0157:H7 infection and campylobacteriosis (Walkerton, Ont.) found that increased OTC drug sales coincided with or lagged shortly behind illness onset (Edge et al, 2004). The authors concluded that OTC drug sales trends would provide a more timely and sensitive tool than monitoring of hospital emergency room visits or tradi-

FIGURE 3 Actual and forecast diarrheal remedy sales

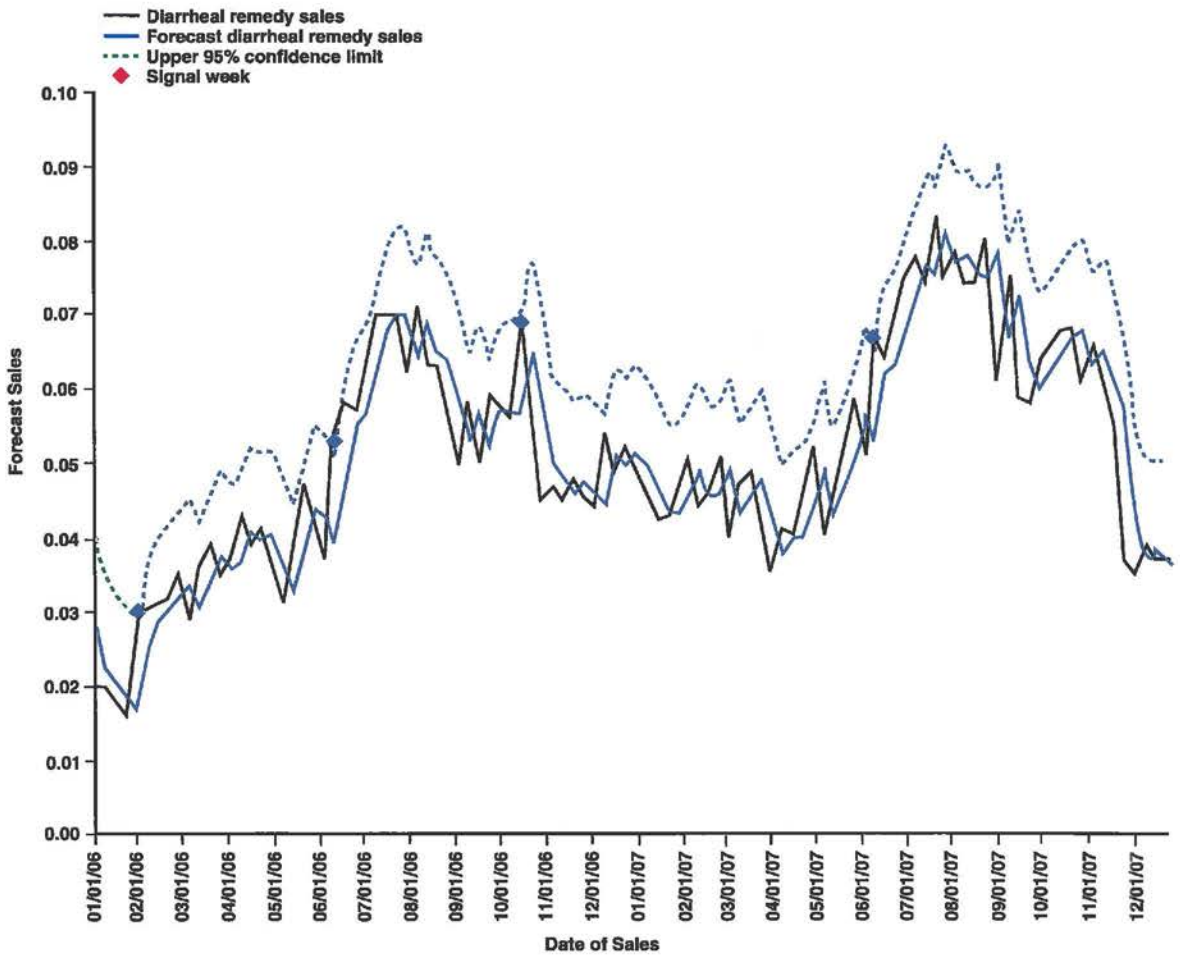


Figure shows weekly counts of diarrheal remedy sales and forecast diarrheal remedy sales from January 2006 through December 2007. Red diamonds indicate weeks when actual diarrheal remedy sales exceeded the forecast upper 95% confidence interval.

tional passive laboratory-based surveillance. Nonetheless, the limitations of OTC drug sales data were noted; data from only one of three pharmacies in Battlefords and one of six in Walkerton were available and formatted appropriately for analysis. Low participation was also noted in a later report comparing drug sales and norovirus temporality; only 12% of pharmacies participated in surveillance (Edge et al, 2006).

Local and state health departments have implemented syndromic surveillance systems with OTC antidiarrheal drug sales monitoring components, but few retrospective studies and no successful reports from ongoing surveillance projects have been published (Uscher-Pines et al, 2009; Leach, 2007; Das et al, 2005; Rodman et al, 1997). Only one report, now antiquated, presents the progress of a functioning OTC antidiarrheal drug sales monitoring program (Das et al, 2005). The researchers reported that from Aug. 1, 2002, through Mar. 31, 2005, they had found no consistent relationship between OTC antidiarrheal drug sales and emergency room visits for GI illness in New York City.

Furthermore, no publications have evaluated surveillance with NRDM OTC diarrheal remedy drug sales in practice, even though these records have been available nationwide for more than six years. One retrospective study presented graphs demonstrating the similar temporality of analgesic, antifever, antidiarrheal, and cough and cold drugs combined and calls to the poison control center in 2003 (Krenzelok et al, 2008). Although the literature review conducted in the current research did identify a number of reports suggesting that syndromic surveillance with OTC antidiarrheal drug sales could enhance traditional disease control activities, the widespread adoption of syndromic surveillance systems and the paucity of published reports on OTC drug sales monitoring systems—and NRDM specifically—suggest publication bias may be present. Publication bias would affect the number of published reports if these types of surveillance are not detecting verifiable outbreaks and operators are not motivated to document null results.

Current study also had limitations. One limitation of this study was that the data set did not include any large regional outbreaks. In a large outbreak, the number of people pursuing OTC remedies might produce a signal that is significantly above the noise in the baseline. The high data variability of diarrheal remedy sales may make it difficult to discern changes resulting from relatively small increases in illness. Although the current study lacked sufficient power to test for large and small outbreaks alike, the findings are unlikely to be biased with regard to outbreak size because there is no reason to believe that individual early health-seeking behavior (such as OTC drug purchases) would be different when an individual's illness is part of an undetected larger outbreak.

OTC drug sales records, as provided by the NRDM, have several limitations. The usability of these data could be improved if participation by enrolled stores was increased or if meta-information on participating stores (such as market coverage) and on the drugs included in each category was made available to users of the data. Because of the way that product data are reported, the authors were not able to assess whether a single product or a subset of products included in the larger diarrheal remedies category might have coincided with known disease, even though their analysis and findings were valid for the larger group of products represented by the category. Furthermore, the current study was not able to assess whether improvements in OTC drug sales reporting systems might enhance the performance of this type of syndromic surveillance. The use of OTC drug sales for surveillance may be prohibitive because of the cost and logistics of data collection or the proprietary and secret nature of the data (Rodman et al, 1997).

County-by-county differences in disease reporting and aggregations of diseases with varying severities may have masked finding a true association. These aggregations also could have covered up localized diarrheal remedy sales fluctuations resulting from isolated outbreaks. Therefore, the authors cannot rule out that county-specific syndromic surveillance may be more sensitive than the regionwide surveillance examined in this analysis.

Many studies have demonstrated that people who become ill purchase OTC drugs before they visit their health care provider; however, such health-seeking behavior varies by factors that include age, gender, culture, and access to health care (Qato et al, 2008; Frosst et al, 2006; Lam & Bradley, 2006; Metzger et al, 2004; Amoako et al, 2003; Vingilis et al, 1999; Espino et al, 1998; Farley, 1997; Stoehr et al, 1997). One study (Frosst et al, 2006) examined health-seeking behavior in response to diarrheal illness specifically. The survey of 351 adults reporting acute gastroenteritis (diarrhea, vomiting, or both) found significant differences between those who used OTC drugs and those who did not; OTC antinauseants, antidiarrheals, and/or rehydration therapies were more likely to be used by women, children 10–14 years of age, those with more severe symptoms, and those with a history of antacid use (Frosst et al, 2006). Although care should be exercised in applying findings from Canada to the United States because each has distinct health care systems, the lack of correlation that the authors of the current study found between diarrheal remedy sales and diarrheal cases could indicate that these data sources measure the occurrence of diarrhea in different populations.

CONCLUSION

Findings from this study did not support the implementation of syndromic surveillance with

NRDM diarrheal remedy sales for enhanced GI outbreak detection of waterborne or other origins. However, because of the lack of large outbreaks in the data set and the relative high variability in sales, the authors cannot exclude the possibility that NRDM sales records may be useful for detecting larger outbreaks. Nonetheless, public health departments and water utilities should carefully evaluate the efficacy and added worth of surveillance systems to avoid the possibility that increased funding for programs such as syndromic surveillance are not accompanied by cutbacks in funding for other programs, resulting in a net increase in overall morbidity (Balter et al, 2005; Reingold, 2003). This is especially true given the proven effectiveness of collaborations between water utilities and health departments in ensuring safe drinking water and recreational water (SFDPH, 2010; Kirian et al, 2008).

ACKNOWLEDGMENT

The authors thank Steve Allgeier and Chrissy Dangel of the US Environmental Protection Agency (USEPA) Water Security Division for their review and comments. Case and outbreak data were made available by the Santa Clara County (Calif.) Public Health Department, the San Mateo County (Calif.) Health Services Agency, the San Francisco (Calif.) Department of Pubic Health Communicable Disease Control Unit, and the California Department of Public Health. This work was supported by a grant from USEPA to the San Francisco Public Utilities Commission.

ABOUT THE AUTHORS

Michelle L. Kirian (to whom correspondence should be addressed) is an epidemiologist at the City and County of San Francisco, Department of Public Health (SFDPH), 1390 Market St., Ste. 910, San Francisco, CA 94102; michelle.kirian@sfdph.org. She has a bachelor's degree from Tulane University in New Orleans, La., and a master's of public health from the University of California at Berkeley. For the past five years, she has worked in the Water Epidemiology Program at SFDPH in such areas as cryptosporidiosis surveillance, drinking water emergency planning, and development of water health fact sheets. June M. Weintraub is senior epidemiologist in the Environmental Health Section at SFDPH.

Date of submission: 10/12/10
Date of acceptance: 03/26/11

JOURNAL AWWA welcomes comments and feedback at journal@awwa.org.

REFERENCES

Ajdacic-Gross, V.; Lauber, C.; Sansossio, R.; Bopp, M.; Eich, D.; Gostynski, M.; Gutzwiller, F.; & Rössler, W., 2007. Seasonal Associations Between Weather Conditions and Suicide—Evidence Against a Classic Hypothesis. *Amer. Jour. Epidemiol.*, 165:5:561.

Amoako, E.P.; Richardson-Campbell, L.; & Kennedy-Malone, L., 2003. Self-Medication With Over-the-Counter Drugs Among Elderly Adults. *Jour. Gerontol. Nursing*, 29:8:10.

Angulo, F.J.; Tippen, S.; Sharp, D.J.; Payne, B.J.; Collier, C.; Hill, J.E.; Barrett, T.J.; Clark, R.M.; Geldreich, E.E.; Donnell Jr., H.D.; & Swerdlow, D.L., 1997. A Community Waterborne Outbreak of Salmonellosis and the Effectiveness of a Boil Water Order. *Amer. Jour. Public Health*, 87:4:580.

Balter, S.; Weiss, D.; Hanson, N.; Reddy, V.; Das, D.; & Heffernan, R., 2005. Three Years of Emergency Department Gastrointestinal Syndromic Surveillance in New York City: What Have We Found? *Morbidity & Mortality Weekly Rept.*, 54 Supp.:54 Supp.:175.

Berger, M.; Shiau, R.; & Weintraub, J.M., 2006. Review of Syndromic Surveillance: Implications for Waterborne Disease Detection. *Jour. Epidemiol. Community Health*, 60:6:543.

Box, G.E.P. & Jenkins, G.M., 1976 (2nd ed.). *Time Series Analysis: Forecasting and Control.* Holden Day, San Francisco.

Brown, J.; Sutter, M.E.; Algren, D.A.; Thomas, J.D.; Ragone, S.; Schier, J.G.; & Geller. R.J., 2010. The Role of a Poison Control Center in Identifying and Limiting an Outbreak of Foodborne Botulism. *Amer. Jour. Preventive Medicine*, 38:6:675.

Buehler, J.W.; Whitney, E.A.; Smith, D.; Prietula, M.J.; Stanton, S.H.; & Isakov, A.P., 2009. Situational Uses of Syndromic Surveillance. *Biosecurity & Bioterrorism*, 7:2:165.

CCR (California Code of Regulations), 2009. Reportable Diseases and Conditions. Reporting to the Local Health Authority. Title 17, CCR §2500, amended Sept. 22, 2009, Sacramento, Calif.

CDC (Centers for Disease Control and Prevention), 1997. *Cryptosporidium* and Water: A Public Health Handbook. www.cdc.gov/ncidod/diseases/crypto/crypto.pdf.

Chretien, J.P.; Tomich, N.E.; Gaydos, J.C.; & Kelley, P.W., 2009. Real-Time Public Health Surveillance for Emergency Preparedness. *Amer. Jour. Public Health*, 99:8:1360.

Das, D.; Metzger, K.; Heffernan, R.; Blater, S.; Weiss, D.; & Mostashari, F., 2005. Monitoring Over-the-Counter Medication Sales for Early Detection of Disease Outbreaks—New York City. *Morbidity & Mortality Weekly Rept.*, 54 Supp.:41.

Derby, M.P.; McNally, J.; Ranger-Moore, J.; Hulette, L.; Villar, R.; Hysong, T.; MacNeill, E.; Lebowitz, M.; & Burgess, J., 2005. Poison Control Center-Based Syndromic Surveillance for Foodborne Illness. *Morbidity & Mortality Weekly Rept.*, 54 Supp.:35.

Edge, V.L.; Pollari, F.; Ng, L.K.; Michel, P.; McEwen, S.A.; Wilson, J.B.; Jerrett, M.; Sockett, P.N.; & Martin, S.W., 2006. Syndromic Surveillance of Norovirus Using Over-the-Counter Sales of Medications Related to Gastrointestinal Illness. *Canadian Jour. Infectious Diseases & Medical Microbiol.*, 17:4:235.

Edge, V.L.; Pollari, F.; Lim, G.; Aramini, J.; Sockett, P.; & Ellis, A., 2004. Syndromic Surveillance of Gastrointestinal Illness Using Pharmacy Over-the-Counter Sales.

A Retrospective Study of Waterborne Outbreaks in Saskatchewan and Ontario. *Canadian Jour. Public Health,* 95:6:446.

Espino, D.V.; Lichtenstein, M.J.; Hazuda, H.P.; Fabrizio, D.; Wood, R.C.; Goodwin, J.; Stroup-Benham, C.A.; & Markides, K.S., 1998. Correlates of Prescription and Over-the-Counter Medication Usage Among Older Mexican Americans: the Hispanic EPESE Study. Established Population for the Epidemiologic Study of the Elderly. *Jour. Amer. Geriatrics Soc.,* 46:10:1228.

Farley, D., 1997. Label Literacy for OTC Drugs. *FDA Consumer,* 31:6.

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. *Canadian Jour. Public Health,* 97:6:489.

Hanssens, D., 1980. Bivariate Time Series Analysis of the Relationship Between Advertising and Sales. *Appl. Econometrics,* 12:1:329.

Henry, J.V.; Magruder, S.; & Snyder, M., 2004. Comparison of Office Visit and Nurse Advice Hotline Data for Syndromic Surveillance—Baltimore–Washington, D.C., Metropolitan Area, 2002. *Morbidity & Mortality Weekly Rept.,* 53 Supp.:112.

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 Medical Informatics & Decision Making,* 10:1:39.

Kirian, M.L.; Meregillano, G.; Gennette, D.; & Weintraub, J.M., 2008. Multi-jurisdictional Investigation of Interactive Fountain-Associated Cryptosporidiosis and Salmonellosis Outbreaks. *Epidemiol. & Infection,* 136:11:1547.

Krenzelok, E.; MacPherson, E.; & Mrvos, R., 2008. Disease Surveillance and Nonprescription Medication Sales Can Predict Increases in Poison Exposure. *Jour. Medical Toxicol.,* 4:1:7.

Lam, A. & Bradley, G., 2006. Use of Self-Prescribed Nonprescription Medications and Dietary Supplements Among Assisted Living Facility Residents. *Jour. Amer. Pharmacists Assn.,* 46:5:574.

Leach, M.J.H.; Evelyn, T.; McDonnell, D.D.; & Morrow, S., 2007. Daily Situational Awareness Tool (DSAT): Utilizing Real Time Infectious Disease Surveillance Systems. Proc. APHA Ann. Meeting & Expo., Washington.

Mackenzie, W.R.; Schell, W.L.; Blair, K.A.; Addiss, D.G.; Peterson, D.E.; Hoxie, N.J.; Kazmierczak, J.J.; & Davis, J.P., 1995. Massive Outbreak of Waterborne *Cryptosporidium* Infection in Milwaukee, Wisconsin: Recurrence of Illness and Risk of Secondary Transmission. *Clinical Infectious Diseases,* 21:1:57.

Mandl, K.D.; Overhage, J.M.; Wagner, M.M.; Lober, W.B.; Sebastiani, P.; Mostashari, F.; Pavlin, J.A.; Gesteland, P.H.; Treadwell, T.; Koski, E.; Hutwagner, L.; Buckeridge, D.L.; Aller, R.D.; & Grannis, S., 2004. Implementing Syndromic Surveillance: A Practical Guide Informed by the Early Experience. *Jour. Amer. Medical Informatics Assn.,* 11:2:141.

Mann, P.; O'Connell, E.; Zhang, G.; Llau, A.; Rico, E.; & Leguen, F.C., 2011. Alert System to Detect Possible School-Based Outbreaks of Influenza-Like Illness. *Emerging Infectious Diseases,* 17:2:262.

Metzger, K.B.; Hajat, A.; Crawford, M.; & Mostashari, F., 2004. How Many Illnesses Does One Emergency Department Visit Represent? Using a Population-Based Telephone Survey to Estimate the Syndromic Multiplier. *Morbidity & Mortality Weekly Rept.,* 53 Supp.:106.

Montgomery, D.C., 2009 (6th ed.). *Introduction to Statistical Quality Control.* John Wiley & Sons, Hoboken, N.J.

NDWAC (National Drinking Water Advisory Committee), 1999. Recommendations Resulting From a Meeting Held November 2–4, 1999. www.epa.gov/ogwdw000/ndwac/pdfs/NDWAC%20Meetings/Nov%202%20and%203,%201999%20Minutes%20and%20Recommendations.pdf (accessed June 28, 2011).

NRDM (National Retail Data Monitor), 2007. National Retail Data Monitor: A Public Health Surveillance Tool. http://rods.health.pitt.edu/nrdm.htm (accessed December 2009).

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. & Infection,* 120:1:43.

Qato, D.M.; Alexander, G.C.; Conti, R.M.; Johnson, M.; Schumm, P.; & Lindau, S.T., 2008. Use of Prescription and Over-the-Counter Medications and Dietary Supplements Among Older Adults in the United States. *Jour. Amer. Medical Assn.,* 300:24:2867.

Reingold, A., 2003. If Syndromic Surveillance Is the Answer, What Is the Question? *Biosecurity & Bioterrorism,* 1:2:77.

Rodman, J.S.; Frost, F.; Davis-Burchat, L.; Fraser, D.; Langer, J.; & Jakubowski, W., 1997. Pharmaceutical Sales—A Method of Disease Surveillance. *Jour. Envir. Health,* 60:4.

SFDPH (San Francisco Department of Public Health), 2010. Water Epidemiology. www.sfphes.org/index_WaterEpi.htm (accessed May 31, 2011).

Stoehr, G.P.; Ganguli, M.; Seaberg, E.C.; Echement, D.A.; & Belle, S., 1997. Over-the-Counter Medication Use in an Older Rural Community: The MoVIES Project. *Jour. Amer. Geriatrics Soc.,* 45:2:158.

Sutter, M.E.; Bronstein, A.C.; Heard, S.E.; Barthold, C.L.; Lando, J.; Lewis, L.S.; & Schier, J.G., 2010. The Role of Clinical Toxicologists and Poison Control Centers in Public Health. *Amer. Jour. Preventive Medicine,* 38:6:658.

Upshur, R.E.; Knight, K.; & Goel, V., 1999. Time-Series Analysis of the Relation Between Influenza Virus and Hospital Admissions of the Elderly in Ontario, Canada, for Pneumonia, Chronic Lung Disease, and Congestive Heart Failure. *Amer. Jour. Epidemiol.,* 149:1:85.

Uscher-Pines, L.; Farrell, C.L.; Cattani, J.; Hsieh, Y.H.; Moskal, M.D.; Babin, S.M.; Gaydos, C.A.; & Rothman, R.E., 2009. A Survey of Usage Protocols of Syndromic Surveillance Systems by State Public Health Departments in the United States. *Jour. Public Health Mgmt. & Practice,* 15:5:432.

USDHHS (US Department of Health and Human Services), 2010. Medicare and Medicaid Programs. Electronic Health Record Incentive Program. Proposed Rule. CFR Parts 412, 413, 422, and 495. http://edocket.access.gpo.gov/2010/pdf/E9-31217.pdf (accessed June 28, 2011).

USEPA (US Environmental Protection Agency), 2005. WaterSentinel System Architecture Draft Version 1.0. www.epa.gov/safewater/watersecurity/pubs/watersentinel_wq_monitoring.pdf (accessed June 28, 2011).

Vingilis, E.; Brown, U.; & Hennen, B., 1999. Common Colds. Reported Patterns of Self-Care and Health Care Use. *Canadian Family Physician,* 45:1:2644.

Wallace, D.J.; Arquilla, B.; Heffernan, R.; Kramer, M.; Anderson, T.; Berstein, D.; & Augerbraun, M., 2009. A Test of Syndromic Surveillance Using a Severe Acute Respiratory Syndrome Model. *Amer. Jour. Emerging Medicine,* 27:4:419.

Webster, D.W.; Vernick, J.S.; & Hepburn, L.M., 2002. Effects of Maryland's Law Banning "Saturday Night Special" Handguns on Homicides. *Amer. Jour. Epidemiol.,* 155:5:406.

Contamination Warning System Demonstration Pilot Project

Syndromic Surveillance Alarm Threshold Analysis Summary

In 2009, the U.S. Environmental Protection Agency awarded the San Francisco Public Utilities Commission (SFPUC) a three-year grant to evaluate the effectiveness of contamination warning systems for community drinking water distribution systems. The public health surveillance component of the project included an assessment of the practical use of over-the-counter drug sales to enhance waterborne disease detection. This fact sheet provides a summary of the methods and findings from that analysis.

Methods: We compared weekly counts of cases and outbreaks of gastrointestinal disease to sales of diarrheal remedies, as measured by the proportion of over-the-counter diarrheal remedies sales to total over-the-counter drug sales, in three adjacent San Francisco Bay Area Counties who receive water from the San Francisco Public Utilities Commission. Sales records for diarrheal remedies, defined as products taken for the relief of diarrhea such as bismuth, attapulgite, subsalicylate, and loperaminde hydrochloride products, were purchased from the National Retail Data Monitor. Case and outbreak records were obtained from county and state health departments.

Methods developed by Box and Jenkins were used to build autoregressive integrated moving average (ARIMA) models for sales and non-outbreak-associated gastrointestinal case counts.

- To determine if the number of units sold or cases detected each week was able to self predict the number in subsequent weeks, sales and case counts were first examined independently.
- To assess whether sales were higher before or during periods of higher diagnosed individual cases, the cross correlations of the sales and case data were examined at several different time lags.
- To assess whether higher sales coincided with known outbreaks, simple regression analysis of the sales and the number of outbreaks or number of outbreak-associated cases per week were run.
- To assess the predictive ability of the model, actual weekly sales were used to predict the expected sales per week. The observed sales per week were then compared to the expected sales. Alerts were generated when the observed sales exceeded the upper 95% confidence limit for the estimated expected sales. The sensitivity and specificity of the model generated alerts and randomly chosen alerts in predicting weeks with outbreaks were calculated and compared.

Results and Conclusions: No relationship between sales and the number of outbreaks, or outbreak-associated or individual cases, was identified. Alerts generated with the sales model did not coincide with outbreaks or outbreak cases better than randomly selected alerts; the sensitivity and specificity of model generated alerts was identical to that of randomly generated alerts.

This study did not support the implementation of syndromic surveillance with National Retail Data Monitor Diarrheal Remedy Sales for enhanced gastrointestinal outbreak detection of waterborne or other origins. However, the study was limited because the largest outbreak in the forecasting period included 153 cases; the NRDM data may be useful for detecting larger outbreaks. The study was also limited because of incomplete participation by enrolled stores, and because the NRDM data do not include meta-information on participating stores such as market coverage and on the drugs included in each category; it is possible that syndromic surveillance with improved sales data might better predict disease. Nevertheless, the use of over-the-counter drugs sales for surveillance may be prohibitive due to the cost and logistics of data collection, or the proprietary and secret nature of the data.

For More Information:

- Kirian ML, Weintraub JM. (2010) Prediction of gastrointestinal disease with over-the-counter diarrheal remedy sales records in the San Francisco Bay Area. *BMC Medical Informatics and Decision Making*. 10(1): 39.
- Kirian ML, Weintraub JM. (2011) Syndromic Surveillance with Over-the-Counter Drug Sales for Waterborne Gastrointestinal Disease? Journal AWWA. (Accepted for Publication August 2011)
- The full analysis report: (available in 2012)

San Francisco Public Utilities Commission
Water Quality Division
1657 Rollins Road
Burlingame, CA 94010
(650) 652-3100

San Francisco Department of Public Health
Environmental Health
1390 Market Street, Suite 910
San Francisco, CA 94102
(415) 252-3800

Syndromic Surveillance for Enhanced Detection of Waterborne Disease Outbreaks?

Michelle Kirian, MPH
San Francisco Department of Public Health
Environmental Health Section

137th APHA Conference
Philadelphia, PA
November 9, 2009

Abstract

Syndromic surveillance has received much attention as a method for health departments to accelerate the detection of and the reaction to outbreaks. However, the ability of syndromic surveillance to enhance the detection of waterborne disease outbreaks and under what circumstances has not been demonstrated. In our current study our aims are to investigate whether monitoring of over-the-counter drug sales can be used to detect regional or local diarrhea waterborne outbreaks earlier than traditional surveillance, and if so, to determine appropriate alert levels. For the analysis we are employing time series and control chart techniques on retrospective over-the-counter drug sales data, provided by the National Retail Data Monitor, and reports of diarrhea disease cases and outbreaks to four county health departments. We will present the model development process and sensitivity, specificity and timeliness estimates for a range of models and model parameters. We will also discuss the problems and benefits in using this data or technique, and will develop recommendations on the implementation of syndromic surveillance with over-the-counter drug sales data in the early detection of waterborne disease outbreaks.

Learning Objectives:

Evaluate the benefits and disadvantages to syndromic surveillance in waterborne disease detection. Describe some time series and control chart methods that can be used to monitor syndromic and other time series data. Discuss the difficulties in applying syndromic surveillance.

Keywords: Surveillance, Infectious Diseases

http://apha.confex.com/apha/137am/webprogram/Paper193637.html

Presentation Outline

Background

Methods and Data

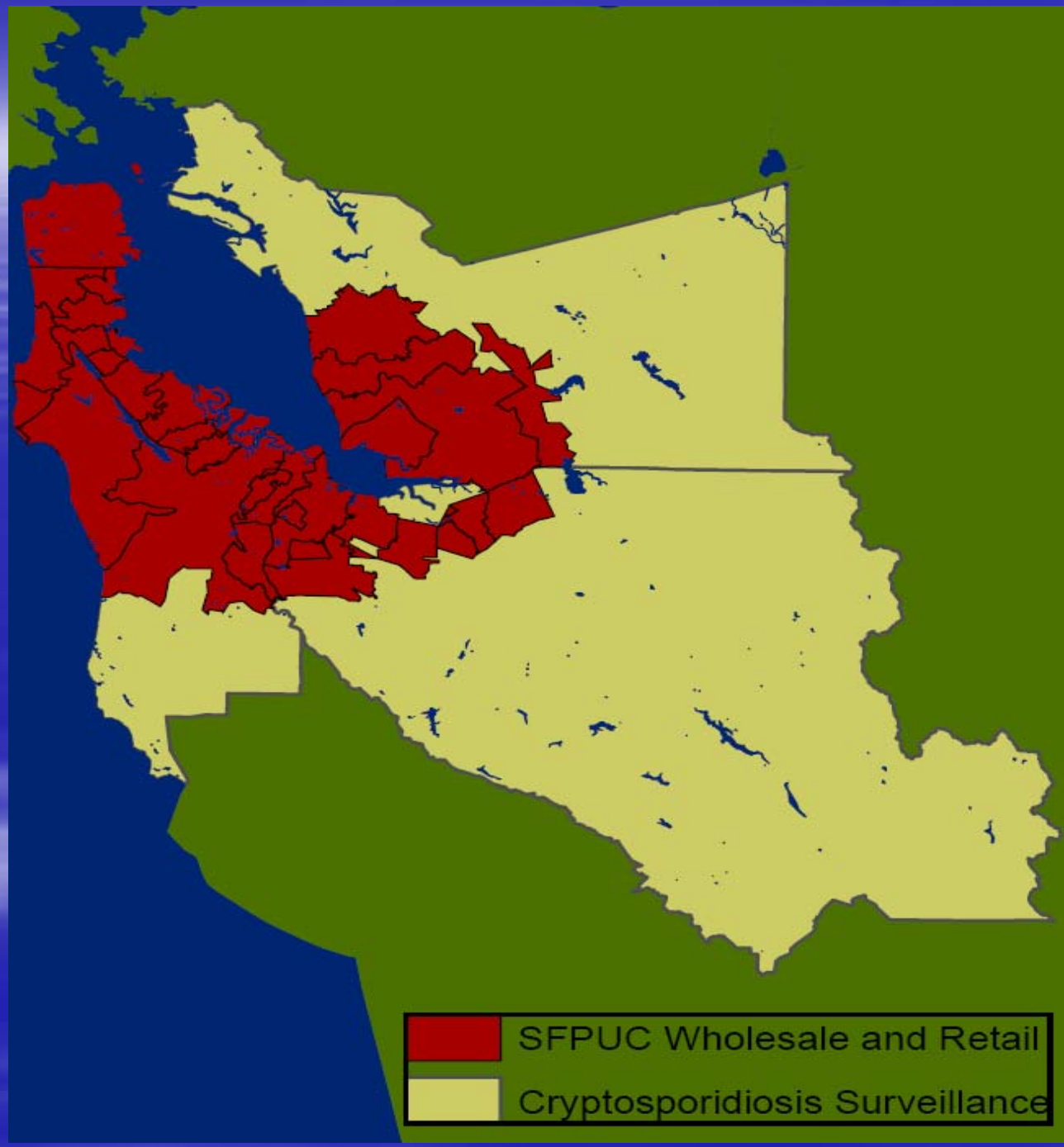
Analysis Results

Conclusions

San Francisco Bay Area

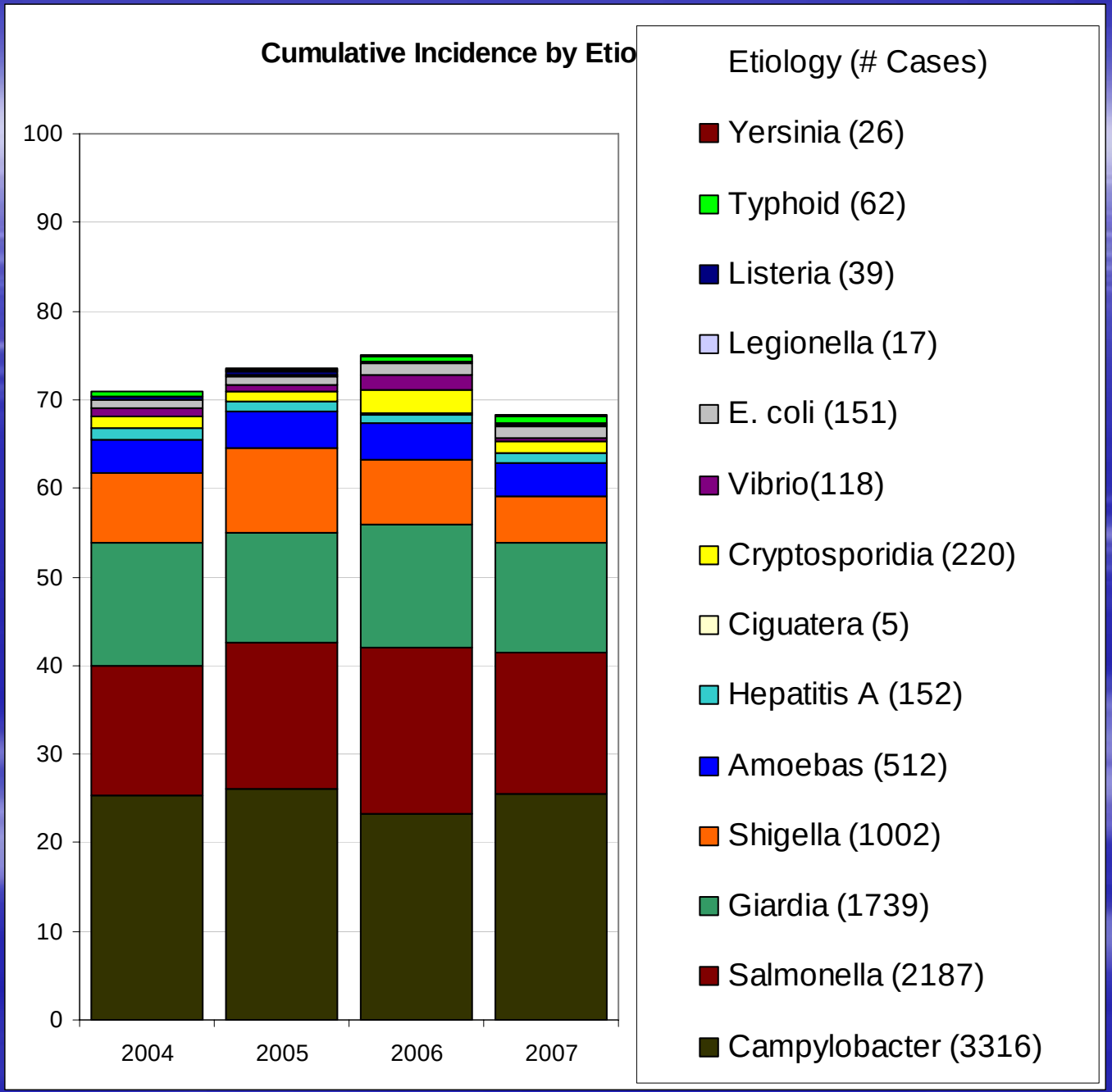
San Francisco Public Utilities Commission (SFPUC) Drinking Water

Cryptosporidiosis Surveillance

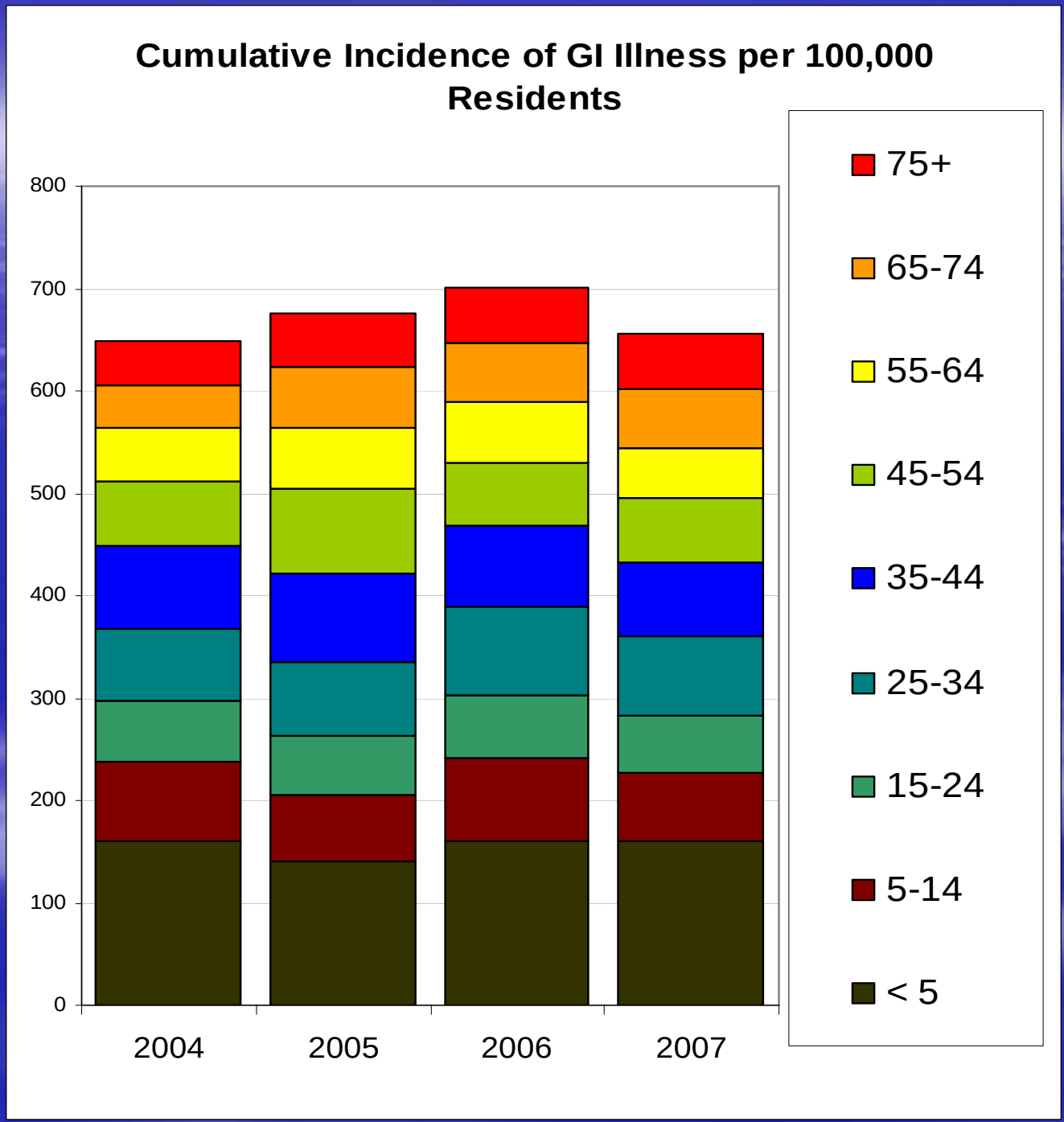


Methods and Data

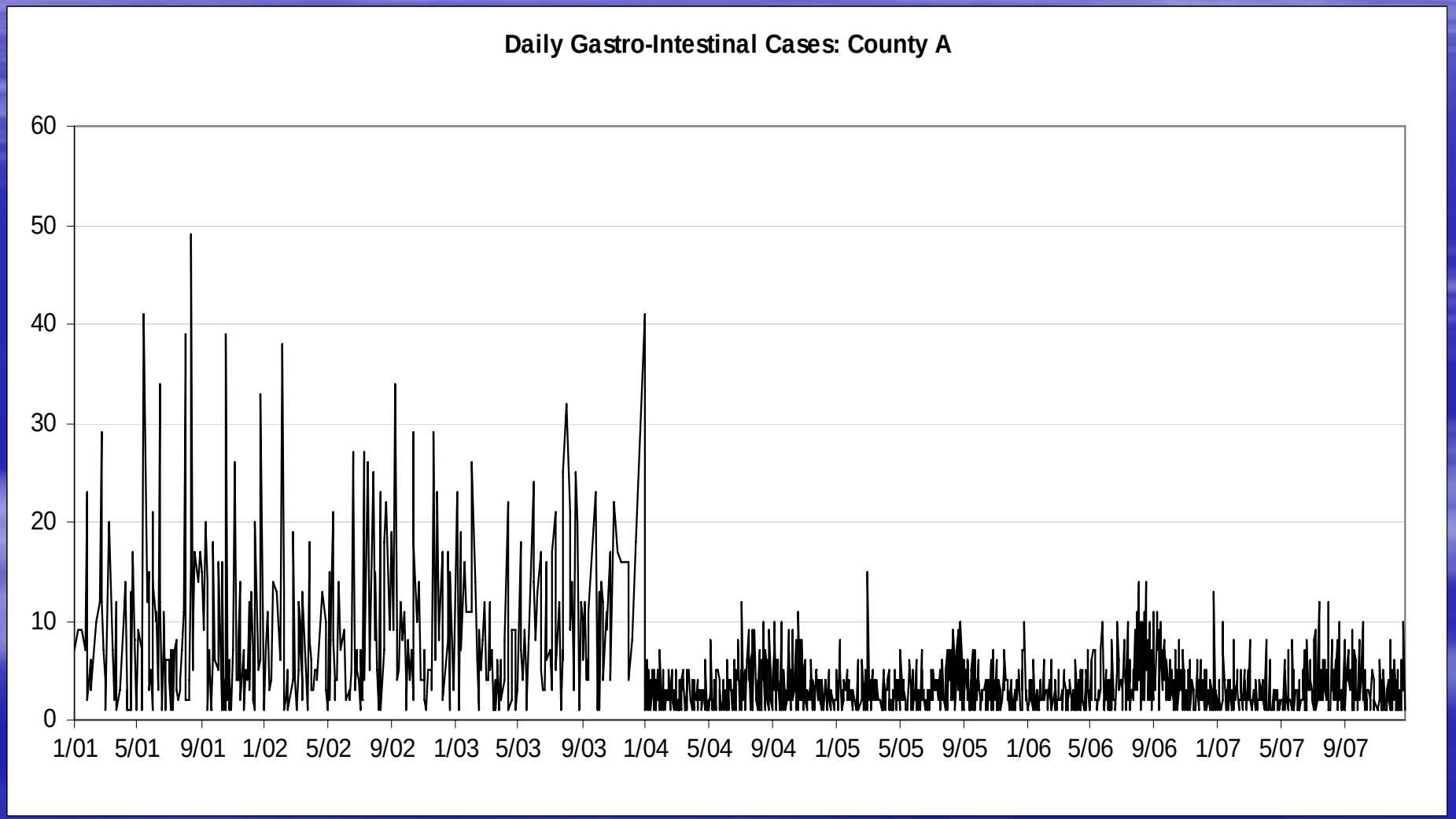
Case Data


Cumulative Incidence by Etio
100
90
80
70
60
50
40
30
20
10
0
2004
2005
2006
2007
Etiology (# Cases)
Yersinia (26)
Typhoid (62)
Listeria (39)
Legionella (17)
E. coli (151)
Vibrio(118)
Cryptosporidia (220)
Ciguatera (5)
Hepatitis A (152)
Amoebas (512)
Shigella (1002)
Giardia (1739)
Salmonella (2187)
Campylobacter (3316)

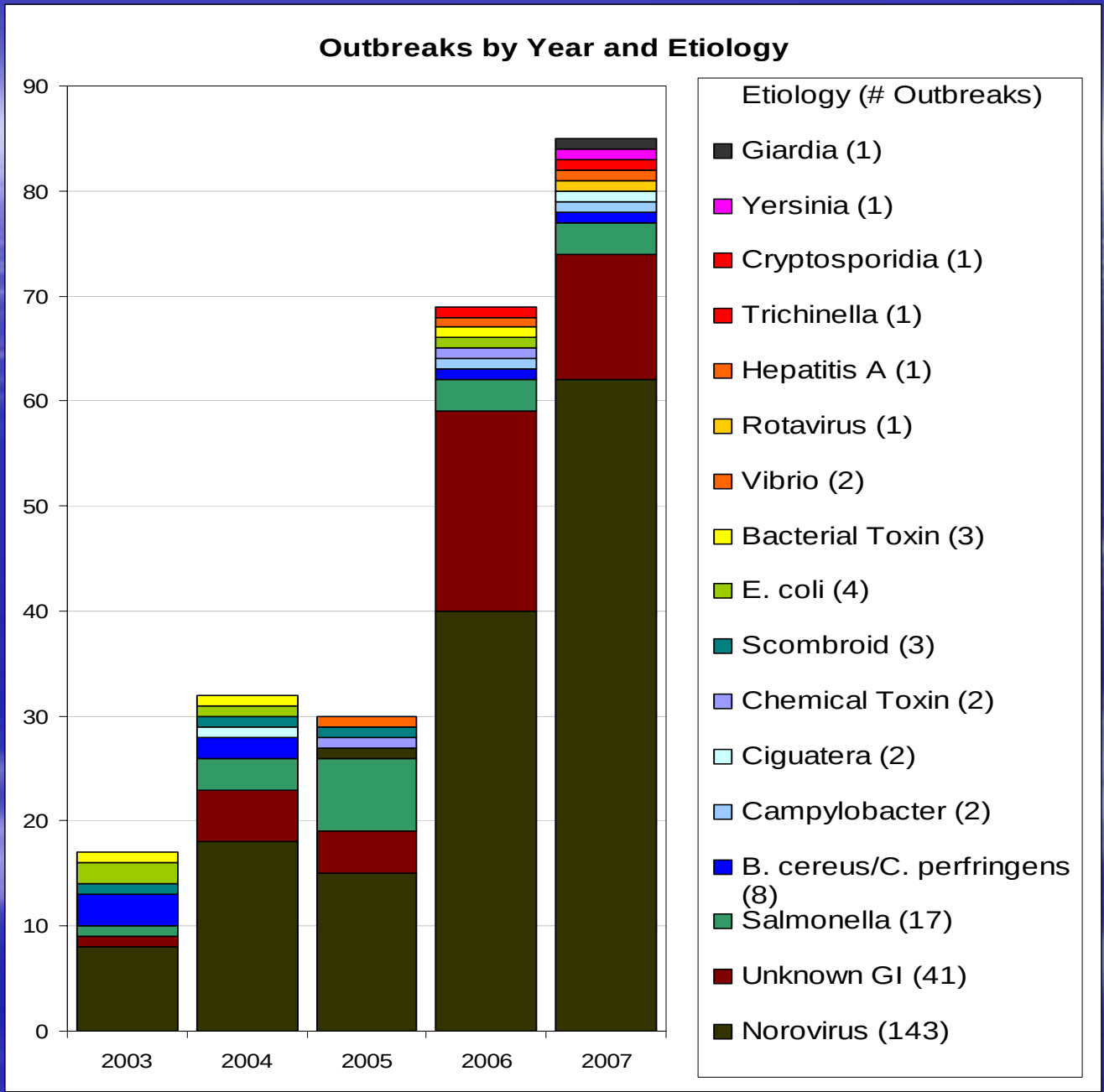
Case Data(2)


Cumulative Incidence of GI Illness per 100,000 Residents
800
700
600
500
400
300
200
100
0
2004
2005
2006
2007
75+
65-74
55-64
45-54
35-44
25-34
15-24
5-14
< 5

Case Data(3)


Daily Gastro-Intestinal Cases: County A
60
50
40
30
20
10
0
1/01
5/01
9/01
1/02
5/02
9/02
1/03
5/03
9/03
1/04
5/04
9/04
1/05
5/05
9/05
1/06
5/06
9/06
1/07
5/07
9/07

Outbreak Data



*2003 includes outbreaks for July –December.

Outbreak Data (2)

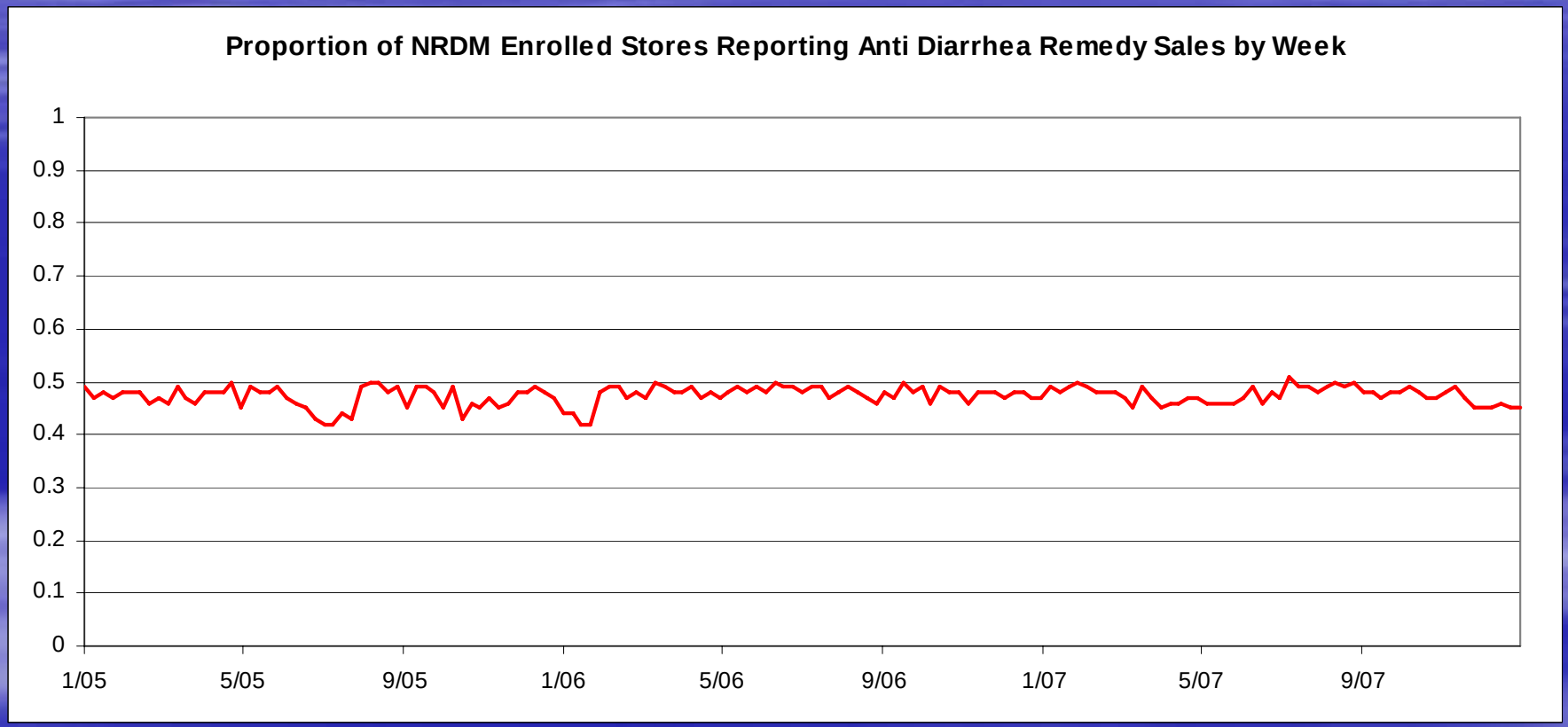
Outbreaks	N	Outbreak-Associated Case *		
		Max	Median	Mean
All Diseases	228	153	19	25
Reportable Diseases	27	65	12	16
Not-Reportable Diseases	201	153	21	27
Study Period				
Model (Jun 29, 2003 –Jul 2, 2005)†	71	110	15	23
Validation (Jul 3, 2005-Dec 31, 2005)	11	26	14	13
Forecasting (Jan 1, 2006-Dec 30, 2007)	149	153	21	27

*Number of cases not available for five outbreaks between January 1, 2006 and January 1, 2008.
† Modeling period was shorter --Jan 4, 2004 to Jul 2, 2005-- for Univariate Case Modeling and Cross Correlation Analysis.

OTC Data

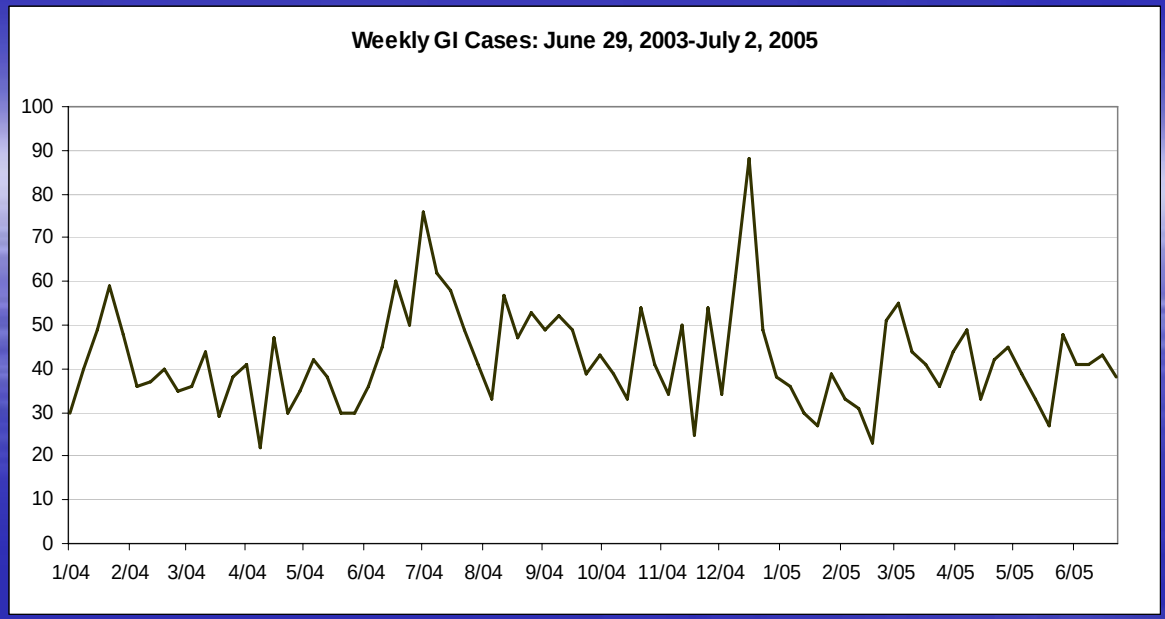
National Retail Data Monitor Over-the-Counter Drug Categories
Diarrhea Remedies*
Anti-Fever Adult
Anti-Fever Pediatric
Bronchial Remedies
Baby/Child Electrolytes
Chest Rubs
Cold Relief Adult Liquid
Cold Relief Adult Tablet
Cold Relief Pediatric Liquid
Cold Relief Pediatric Tablet
Cough Syrup Adult Liquid
Cough Adult Tablet
Cough Syrup Pediatric Liquid
Cough/Cold
Hydrocortisones
Nasal Product Internal
Throat Lozenges
Thermometers

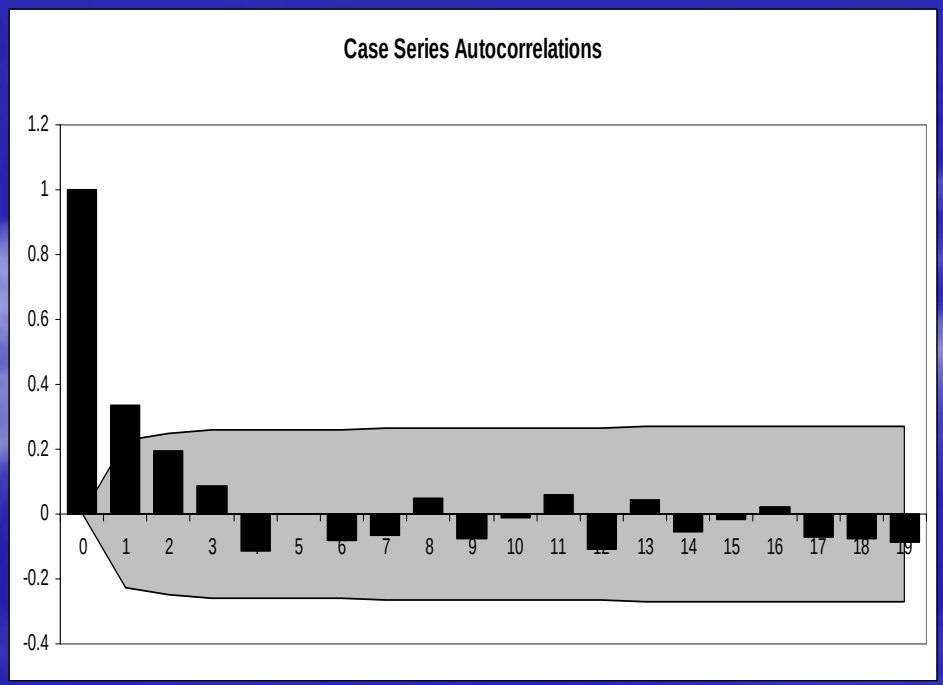
OTC Data (2)

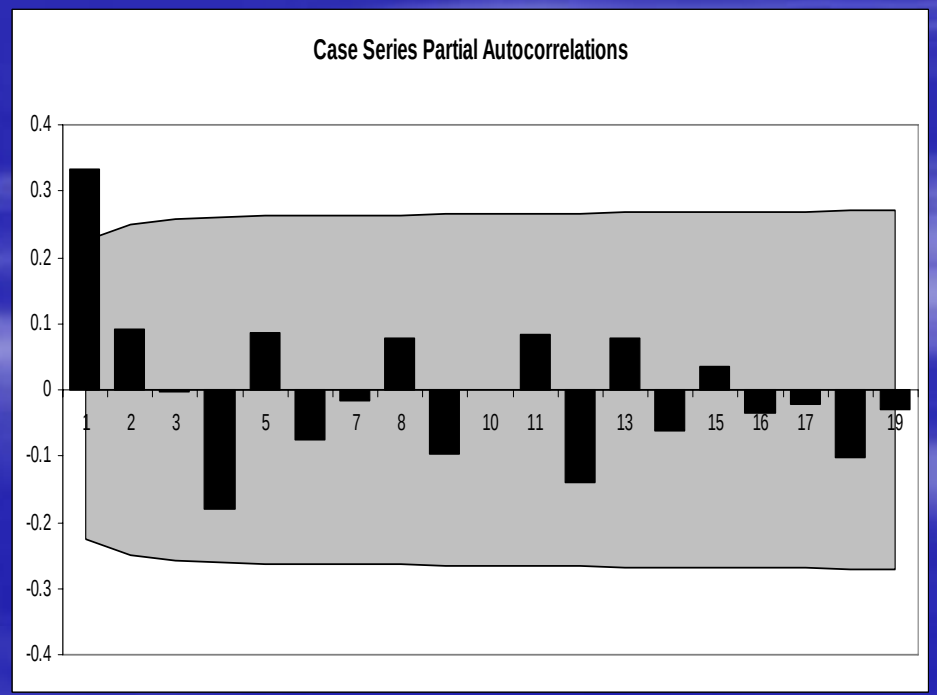

Proportion of NRDM Enrolled Stores Reporting Anti Diarrhea Remedy Sales by Week
1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
1/05
5/05
9/05
1/06
5/06
9/06
1/07
5/07
9/07

Analysis Results

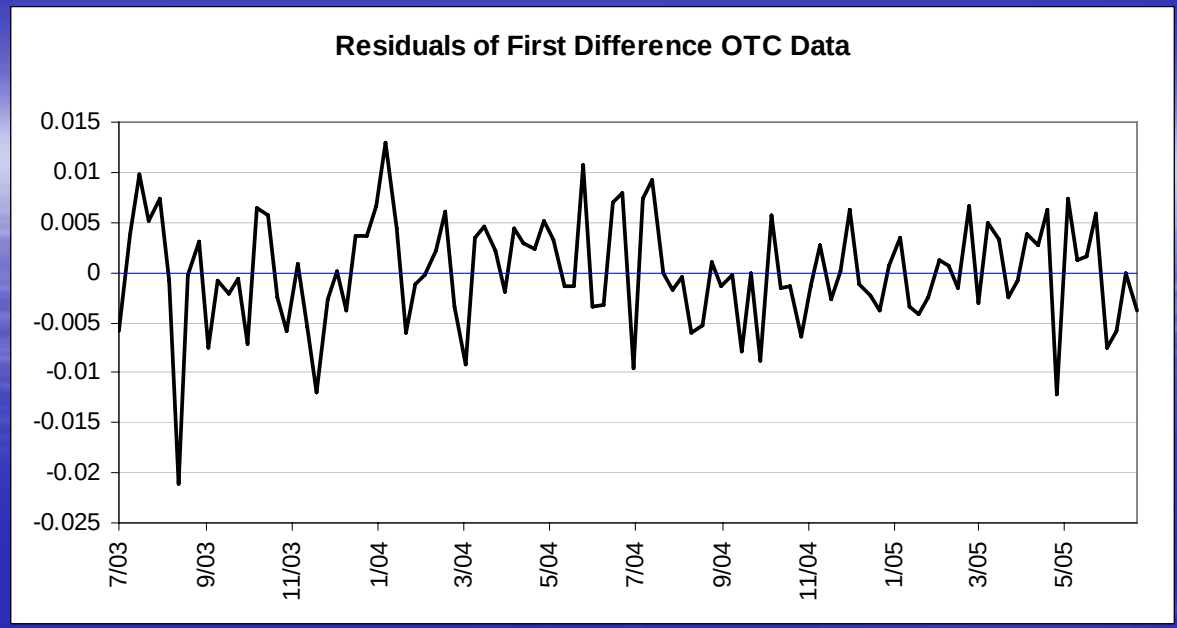
Case Univariate Modeling

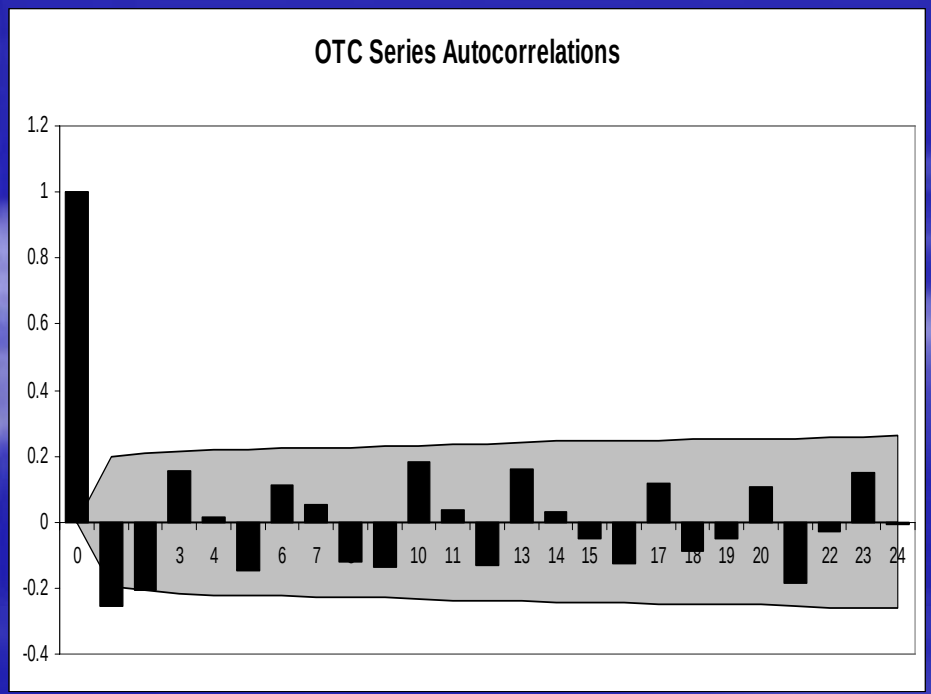


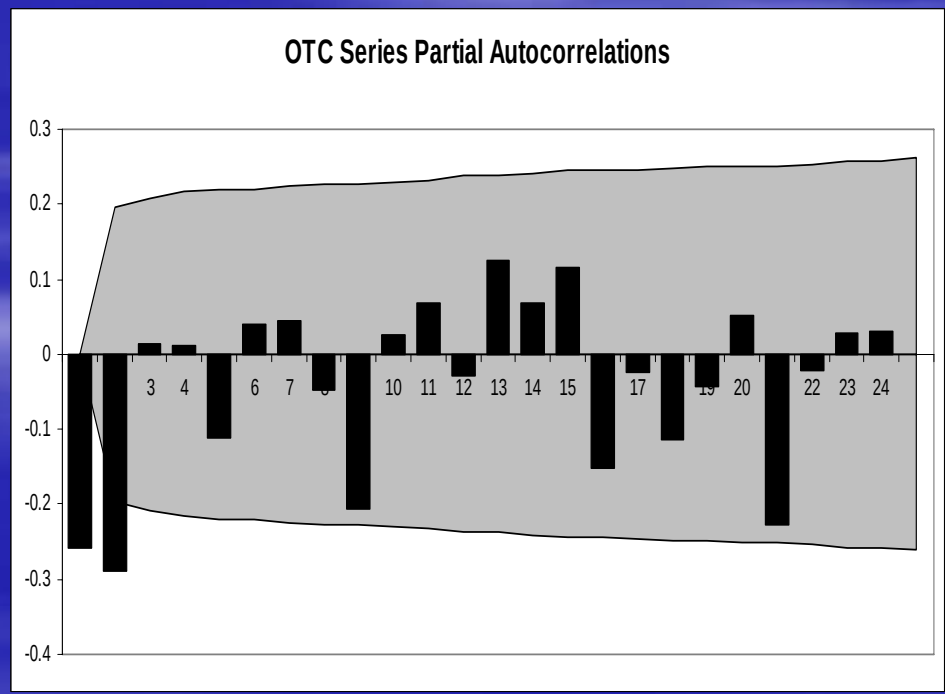




OTC Univariate Modeling







Univariate ARIMA Models

Case Data: ARIMA(1,0,0) = AR(1)

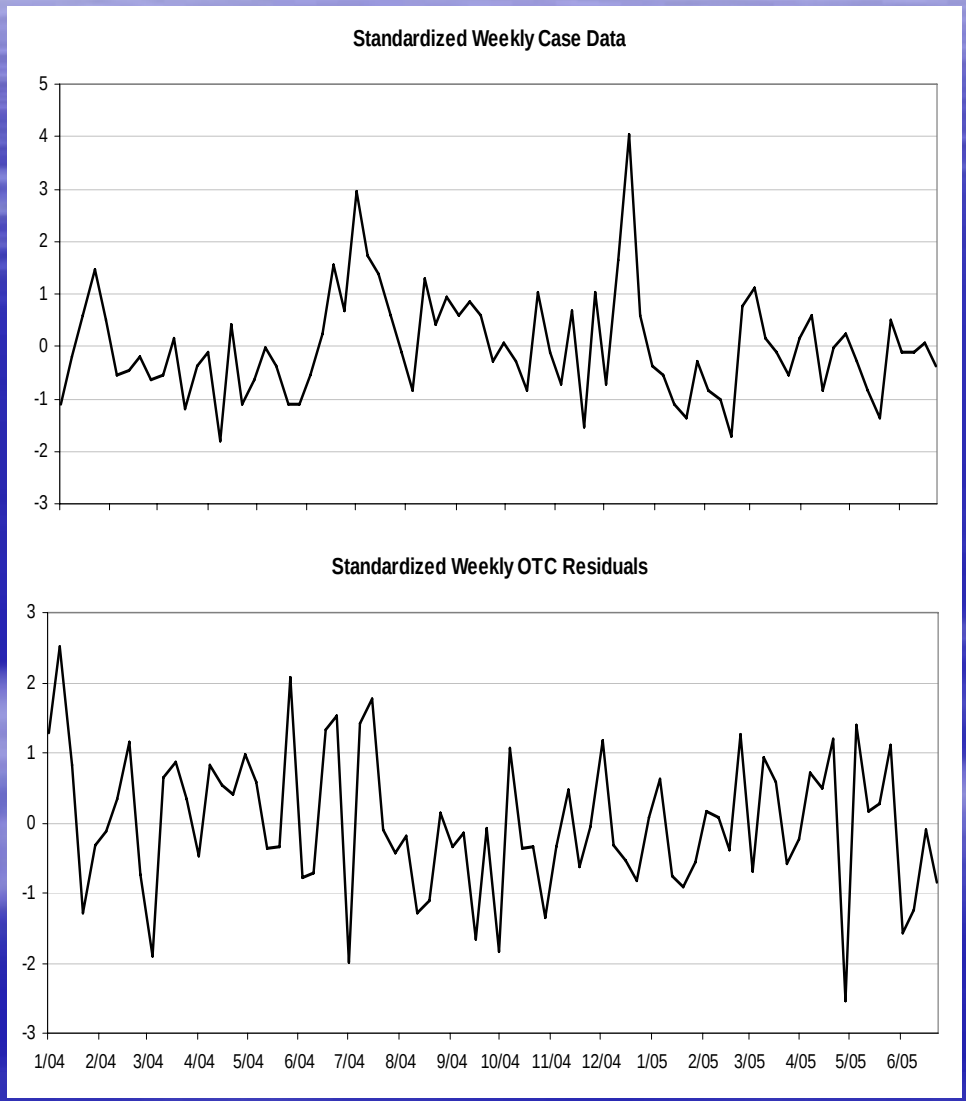
- Basic AR(1) model: $Y_t = \Phi_1 Y_{t-1} + e_t$

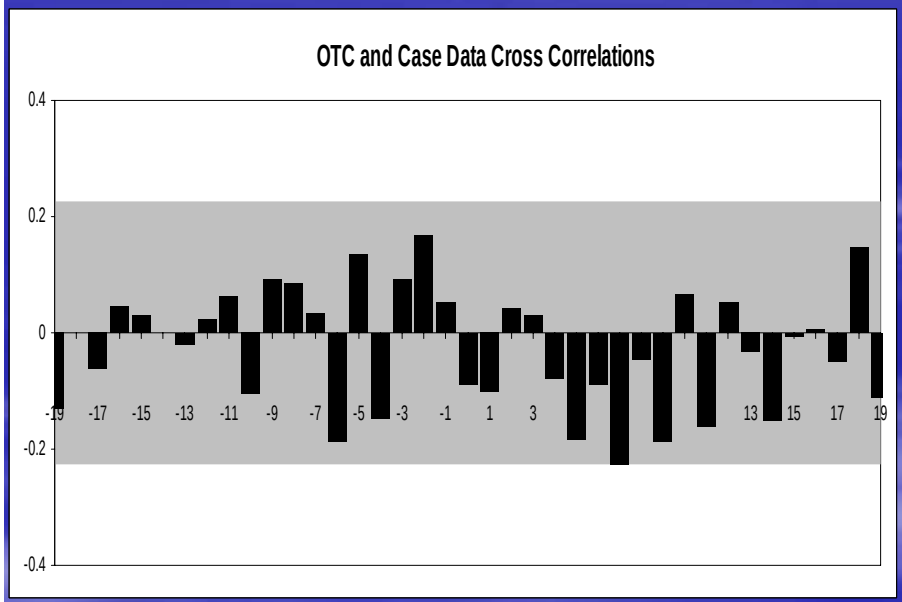
OTC Data: ARIMA(0,1,1) = IMA(1) =SES

- Basic IMA(1) model: $Y_t - Y_{t-1} = e_t - \theta_1 e_{t-1}$

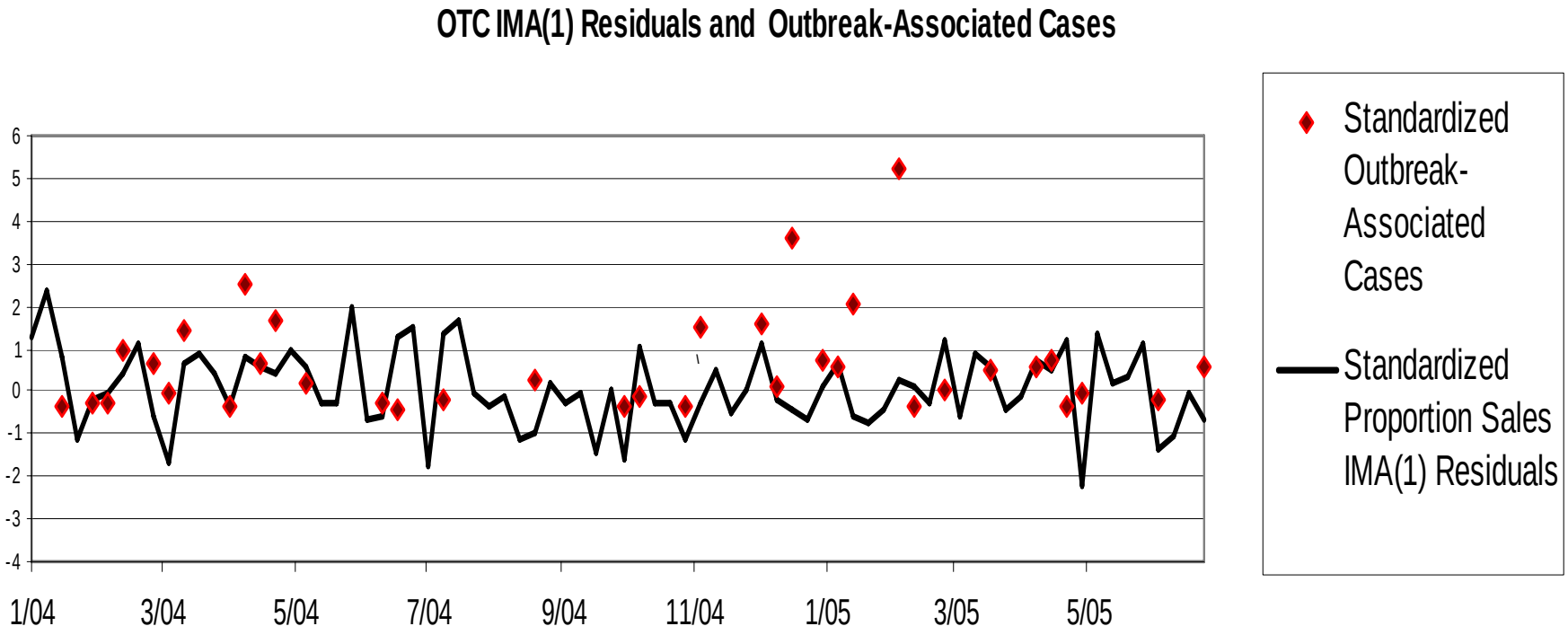
Data	Parameter	Estimate	Standard Error	T-Ratio
OTC	θ_1	0.40	0.09	4.42
Case	Φ_1	0.33	0.11	3.09

OTC and Case Data Cross Correlations





OTC and Outbreak Data



Univariate Forecasting

Details for Outbreaks with 50 or More Cases						
Report Date	First Onset	Last Onset	Etiology	Cases	Institutional	Signal
1/30/06	1/25/06	Missing	Norovirus	101	Yes	Yes
Missing	4/18/06	Missing	Unk/ Norovirus	60	No	Yes
Missing	4/24/06	Missing	Unk/ Norovirus	62	Yes	Yes
4/28/06	4/25/06	5/2/2006	Norovirus	107	No	No
5/01/06	4/26/06	Missing	Unk/ Norovirus	55	No/ Unk	Yes
5/12/06	5/8/06	Missing	Norovirus	50	Yes	Yes
10/31/06	10/26/06	Missing	Norovirus	81	Yes	No
12/05/2006	11/23/06	Missing	Norovirus	86	Yes	No
12/06/06	11/30/06	Missing	Norovirus	72	Yes	No
12/07/06	11/30/06	Missing	Norovirus	63	Yes	No
12/13/06	12/7/06	Missing	Unknown	80	Yes	No
12/11/06	12/7/06	Missing	Unknown	61	No/ Unk	No
1/22/07	1/3/07	Missing	Norovirus	76	Yes	No
1/8/07	1/8/07	Missing	Norovirus	60	Yes	No
7/17/07	7/13/07	7/17/07	Norovirus	92	No	No
Missing	8/3/07	8/17/07	Norovirus	153	No/ Unk	No
Missing	9/15/07	9/19/07	Norovirus	51	No/ Unk	No
12/20/07	12/20/07	1/1/08	Norovirus	52	Yes	No
12/22/07	12/22/07	Missing	Norovirus	52	Yes	No
Missing	12/22/07	1/15/08	Norovirus	76	No/ Unk	No

Univariate Forecasting (2)

Sensitivity and Specificity (OTC IMA(1) Signals)						
	All Outbreaks		Outbreaks with > 49 Cases		Outbreaks with > 99 Cases	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
Signal Week	4% (4/94)	100% (11/11)	4% (1/26)	97% (76/79)	14% (1/7)	97% (95/98)

Sensitivity and Specificity (Random Signals)						
	All Outbreaks		Outbreaks with > 49 Cases		Outbreaks with > 99 Cases	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
Signal Week	3% (3/94)	91% (10/11)	4% (1/26)	97% (76/79)	0% (0/7)	96% (94/98)

Sensitivity and Specificity (Random Signals 2)						
	All Outbreaks		Outbreaks with > 49 Cases		Outbreaks with > 99 Cases	
	Sensitivity	Specificity	Sensitivity	Specificity	Sensitivity	Specificity
Signal Week	4% (4/94)	100% (11/11)	8% (2/26)	97% (77/79)	14% (1/7)	97% (95/98)

Limitations and Future Work

No very large regional outbreaks during study period to test system

Aggregation of data into regional syndromic surveillance system may not be sensitive to isolated outbreaks.

Future work to look at de-aggregated data.

Conclusions

We did not find any relationship between OTC drug sales as reported by NRDM and known outbreaks or GI cases in the SF Bay Area

This analysis did not support the utility of OTC drug sales for predicting waterborne disease outbreaks

Norovirus Outbreaks remain an unaddressed public health problem

Thanks…

Santa Clara County Public Health Department

San Mateo County Health Services Agency

San Francisco Department of Public Health, Communicable Disease Control Unit

California Department of Public Health

U.S. Environmental Protection Agency

San Francisco Public Utilities Commission