Original **Investigations**

JAMIA

Research Paper ■

Detection of Pediatric Respiratory and Diarrheal Outbreaks from Sales of Over-the-counter Electrolyte Products

William R. Hogan, MD, Fu-Chiang Tsui, PhD, Oleg Ivanov, MD, MPH, Per H. Gesteland, MD, Shaun Grannis, MD, J. Marc Overhage, MD, PhD, J. Michael Robinson, MA, Michael M. Wagner, MD, PhD, for the Indiana–Pennsylvania–Utah Collaboration

A b s t r a c t   **Objective:** To determine whether sales of electrolyte products contain a signal of outbreaks of respiratory and diarrheal disease in children and, if so, how much earlier a signal relative to hospital diagnoses.

Design: Retrospective analysis was conducted of sales of electrolyte products and hospital diagnoses for six urban regions in three states for the period 1998 through 2001.

Measurements: Presence of signal was ascertained by measuring correlation between electrolyte sales and hospital diagnoses and the temporal relationship that maximized correlation. Earliness was the difference between the date that the exponentially weighted moving average (EWMA) method first detected an outbreak from sales and the date it first detected the outbreak from diagnoses. The coefficient of determination (r^2) measured how much variance in earliness resulted from differences in sales' and diagnoses' signal strengths.

Results: The correlation between electrolyte sales and hospital diagnoses was 0.90 (95% CI, 0.87–0.93) at a time offset of 1.7 weeks (95% CI, 0.50–2.9), meaning that sales preceded diagnoses by 1.7 weeks. EWMA with a nine-sigma threshold detected the 18 outbreaks on average 2.4 weeks (95% CI, 0.1–4.8 weeks) earlier from sales than from diagnoses. Twelve outbreaks were first detected from sales, four were first detected from diagnoses, and two were detected simultaneously. Only 26% of variance in earliness was explained by the relative strength of the sales and diagnoses signals ($r^2 = 0.26$).

Conclusion: Sales of electrolyte products contain a signal of outbreaks of respiratory and diarrheal diseases in children and usually are an earlier signal than hospital diagnoses.

■ **J Am Med Inform Assoc.** 2003;10:555–562. DOI 10.1197/jamia.M1377.

Subsequent to the anthrax bioterrorism attacks of 2001, research in public health surveillance has placed a new emphasis on the very early detection of outbreaks of disease.[1] The rationale is that further terrorism in the United States is a real possibility and that—other than prevention—the best approach to mitigation of an attack is to initiate an emergency response as soon as possible.

Identifying new types of surveillance data that are inherently earlier indicators of an outbreak is one research priority.[1] Data about sales of over-the-counter (OTC) health care products are of particular interest because symptomatic individuals often treat themselves with OTC products instead of, or prior to, having contact with the health care system.[2–5] Additionally, the retail industry already collects OTC sales

data electronically at the point of sale; thus, the technical effort to obtain these data for public health surveillance is low relative to other types of data. OTC health care products are, in fact, already being monitored routinely by surveillance systems developed by the Johns Hopkins Applied Physics Laboratory for the National Capital Area,[6] the New York City Department of Health,[7] and the RODS laboratory for national level monitoring (The National Retail Data Monitor is located in Pittsburgh but monitors nationwide data).[8]

There are, however, two fundamental questions about the relationship between sales of any OTC health care product (or any other nontraditional type of surveillance data) and disease, and these questions have been only partially answered for but a few OTC health care products. The first question is: *Do the sales data contain a signal of disease?* The second is: *Can the signal support earlier or more sensitive detection than current surveillance methods?* There is prior work that partially addresses these questions for sales of OTC antipyretics, cold remedies, and diarrhea remedies.[5,9–14] During an influenza B epidemic, sales of OTC cold remedies peaked three weeks before positive influenza cultures, but sales of OTC antipyretics (both pediatric and adult) did not change significantly from baseline.[9] Increases in sales of OTC "flu" and cough preparations precede increases in influenza activity and asthma activity (indicated by outpatient billing diagnoses).[10] Sales of OTC diarrhea remedies rose six-fold during a waterborne outbreak of *Salmonella typhimurium*.[11] Sales of OTC diarrhea remedies such as Imodium and Kaopectate increased five-fold over baseline levels during a *Cryptosporidium* outbreak.[12] In that outbreak, the increase in sales began approximately two weeks prior to public health first becoming aware of the outbreak. In two other *Cryptosporidium* outbreaks, sales of OTC diarrhea remedies increased by as much as 26-fold.[13] An even larger *Cryptosporidium* outbreak occurred in Milwaukee, Wisconsin, in 1993, affecting 403,000 individuals.[15] During that outbreak, sales of OTC diarrhea remedies from the one pharmacy that was studied retrospectively increased approximately three-fold from baseline for the month of March (public health became aware of the outbreak on April 5, 1993).[14] A survey by Corso et al.[5] of the behavior of individuals during the Milwaukee outbreak provides corroborative evidence that sales of OTC medications are an early indicator of diarrheal outbreaks: approximately 30% of the Milwaukee cryptosporidosis patients studied self-medicated with OTC diarrhea remedies prior to visiting the emergency room.

These studies provide important results about the potential of routine monitoring of sales of OTC health care products. They show that sales of OTC cold remedies and diarrhea remedies can signal a disease outbreak and that the *peak* of sales activity precedes the peak activity of disease. However, the methods used did not quantify how much earlier the outbreak could have been detected from OTC sales relative to alternative methods. The prior studies also examined very few outbreaks (one each), thus, repetition of these experiments for additional outbreaks would increase confidence in the result that OTC sales can provide an early warning.

The current research studies the correlation between sales of electrolyte products—a category of OTC products that has been rarely studied—and pediatric disease. The current evidence about this relationship is limited to a single study

that found that sales of electrolytes increased four-fold during a waterborne *Cryptosporidium* outbreak.[13] The previously discussed study by Stirling et al.[12] included one electrolyte product in the diarrhea remedy category (which showed a four-fold rise), but the degree to which sales of that electrolyte product contributed to the category increase was not studied.

The current study addresses the sample size limitations of prior studies by examining the sales of electrolyte products during seasonal outbreaks of diarrheal and respiratory disease in children in six urban regions over a four-year period. It also uses an outbreak detection algorithm to measure how much earlier the outbreaks could be detected from sales of electrolyte products relative to detection from hospital diagnoses. The hypotheses were that sales of electrolyte products correlate with (contain a signal of) pediatric outbreaks and that detection of outbreaks from electrolyte sales precedes detection from hospital diagnoses.

Methods

Urban Regions
The research used historical data about sales of electrolyte products and hospital discharges from six urban regions in three states for the years 1998 through 2001. The six urban regions were: Harrisburg, Pennsylvania; Indianapolis, Indiana; Philadelphia, Pennsylvania; Pittsburgh, Pennsylvania; Salt Lake City, Utah; and Scranton, Pennsylvania.

Sales of Electrolyte Products
Electrolyte products are aqueous solutions of sodium and glucose formulated to take advantage of the glucose-sodium coupled transport mechanism to promote uptake of sodium and, by subsequent osmosis, water.[16] They are sold by retail stores including grocery stores, pharmacies, and "mass" retailers (e.g., Kmart) without a prescription.

Information Resources, Inc. (IRI), a company that aggregates sales data from retailers, provided weekly counts of unit sales of electrolyte products broken down by store zip code for the entire study period for all regions. IRI estimates that the market share represented in the dataset is greater than 90% for all regions.

Hospital Diagnoses
The study used hospital discharge datasets to create an indicator of respiratory and diarrheal disease in children. Hospital discharge datasets are compilations of machine-readable abstracts of hospital inpatient stays. These datasets include discharge diagnoses coded using ICD-9-CM with dates of admission and discharge and patient age. The study included patients with hospital diagnoses of respiratory as well as diarrheal diseases because of evidence from the authors' unpublished research that, in addition to infectious diarrheal illness, infectious pediatric respiratory illness also correlates with sales of electrolyte products.

The Pennsylvania Health Care Cost Containment Council (PHC4) provided a hospital discharge dataset that comprised all discharges for all hospitals in Pennsylvania for the study period. The Utah Department of Health provided a discharge dataset for the Salt Lake region from the Utah Hospital Discharge Database, which includes all discharges for all hospitals in Utah except Intermountain Shriners Hospital, which is exempt from reporting requirements because it is

a charity hospital. Because that hospital specializes in orthopedics, its absence is unlikely to be of significance in this study. The Indianapolis Network for Patient Care provided a hospital discharge dataset for the Indianapolis region for the study period, covering 95% of hospital discharges in the region. All datasets included the pediatric hospitals in the regions.

Hospitalizations were included in this study only for patients aged 5 years or less at the time of admission. The American Academy of Pediatrics recommends oral rehydration therapy with electrolyte products for children of ages 1 month to 5 years.[17]

The authors had created four categories of ICD-9-CM codes that represent selected infectious respiratory and diarrheal pediatric diseases (Table 1, available in an online data supplement at www.jamia.org): pneumonia and influenza, bronchiolitis caused by respiratory syncytial virus (RSV) and other causes, rotavirus gastroenteritis, and pediatric gastroenteritis from all causes. For this study, all codes from Table 1 (available in an online data supplement at www.jamia.org) were merged into a single category representing respiratory and diarrheal diseases.

Information Resources, Inc., defines a week as Monday 12 AM local time through Sunday midnight; that definition was used to compile hospitalizations into weekly counts to match the weekly aggregation of the electrolyte sales data. We aggregated patients with these diagnoses using their dates of admission rather than their dates of discharge. As a proxy for the date of first diagnosis, the date of admission was better for this study than the date of discharge for several reasons. First, it is likely that the infectious disease was a cause of the hospitalization so the date of admission represents the first date that an astute clinician could possibly accurately classify the patient as having either an infectious respiratory or diarrheal disease. More importantly, using the date of admission does not introduce a bias in favor of detection from sales of electrolytes as would the use of discharge date. In the case of hospital-acquired (nosocomial) infections, the diagnosis may be established quite late in the admission, and patients with hospital-acquired infections obviously do not purchase electrolytes from retail stores to treat the infection (or at least not until after discharge). All of these factors made the date of admission a conservative research choice that biased the study against the second hypothesis and in favor of earlier detection from hospital diagnoses. A limitation of the available hospital discharge datasets was that they did not support the removal of hospital-acquired infections from the analysis.

Plotting the Data

To plot electrolyte sales and hospital diagnoses on the same graphs, even though the absolute counts of each typically differ by a factor of approximately 30, each time series was normalized by computing the mean and standard deviation of the data over all four years. The weekly counts then were plotted as the number of standard deviations that they deviated from the mean. This is a standard way of plotting time-series data in the field of signal analysis (for example, Lobanov used this method to compare two time series in the domain of speech analysis[18]).

Outbreaks

For purposes of this study, we considered the large seasonal increases in respiratory and diarrheal disease that occur in children in most years to be outbreaks. There were 18 such outbreaks, one per year per urban region. Even though the period of the study was four years, the datasets started and ended in the middle of outbreak periods resulting in only three full seasons for study.

Measurement of Correlation

This study used the cross correlation function to determine the maximum correlation of electrolyte sales with hospital diagnoses and the time latency at which the maximum correlation occurred. The cross-correlation function is a standard method from the field of signal analysis for measuring the similarity of two signals and the time latency of one signal relative to the other.[19] The time latency derived from correlation analysis is one measure that other researchers have used in published[20] and unpublished reports to characterize the earliness of signals for detection of outbreaks.

Detection Algorithm Method for Measuring Earliness

The *detection algorithm method* measures earliness as the time difference (in weeks) between the date that an outbreak-detection algorithm first signals an outbreak from hospital diagnosis data and the date that it first signals an outbreak from sales of electrolyte products. This method was first introduced by Tsui et al.[20] to evaluate the earliness of detection of influenza from ICD-9-CM coded chief complaints relative to detection from influenza surveillance data.

The exponentially weighted moving average (EWMA) is a control-chart method[21] that has been used to detect outbreaks.[22] In this study, EWMA served as an unbiased method for determining the dates of first detection of outbreaks from hospital diagnoses and sales of electrolytes. EWMA sets detection thresholds for a time series based on its variance and mean and at the same time smoothes the time series to reduce the likelihood of false-positive results caused by noise in the data. Specifically, a detection threshold is computed as the mean of the time series plus a constant multiple of *sigma*, where sigma is a function of the standard deviation. Varying the constant multiple varies the detection threshold to achieve more or less sensitivity and specificity of detection.

The *date of detection* was the date that EWMA first signaled an alarm. EWMA detection thresholds (also known as upper control limits) were computed from time intervals in the datasets during which no outbreaks could be discerned by one of the authors (WRH). Specifically, WRH marked *nonoutbreak periods* in each region's time series of hospital diagnosis data by selecting contiguous weeks of low, baseline counts in the summer when counts for the diseases in Table 1 (available in an online data supplement at www.jamia.org) are typically at their lowest levels. The time intervals between non-outbreak periods are referred to in this report as *suspicious periods*.

Because the reference (gold standard) date for the start of the outbreak using the EWMA detection algorithm was based on hospital diagnosis data, it was important that the detection algorithm be accurate (i.e., 100% sensitive and specific when run on that time series). For that reason, the detection thresholds in this research were limited to those that were sufficiently high to avoid false-positive results when run on the hospital diagnosis data.

Measurement of Sensitivity and Specificity

The purpose of measurement of specificity and sensitivity of outbreak detection in this research was to ensure that the earliness measurements were made at detection thresholds that exhibited 100% sensitivity and specificity. This limitation ensured that the earliness measurement was not made using a false-positive result in the electrolyte sales time series. A false-positive result was detection from electrolyte sales during a nonoutbreak period or detection from electrolyte sales during a suspicious period in which no corresponding detection from hospital diagnoses occurred. A false-negative result was the absence of detection from electrolyte sales during a suspicious period in which detection did occur from hospital diagnoses.

Sensitivity Analyses

To test whether the choice of detection threshold (that is, the choice of sigma multiple) affected the EWMA results, the analysis was repeated at all multiples that retained sensitivity and specificity of 100% in the electrolyte sales time series. There was a wide range of thresholds that had these characteristics because the outbreaks being studied were so distinct and strong.

To determine whether the measurement of earliness was sensitive to the choice of ICD-9-CM codes used to create the hospital diagnosis time series, the Pennsylvania HC4 data were analyzed for ICD-9-CM codes that (1) exhibited seasonal variation, (2) added at least five admissions to a suspicious period week, and (3) could reasonably be associated with electrolyte usage. This analysis identified ten codes (Table 3), from which additional sets of ICD-9-CM codes were created. Each new set included the codes from Table 1 (available in an online data supplement at www.jamia.org) plus a combination of the ten codes (Table 4, available in an online data supplement at www.jamia.org). These sets were used to create additional hospital diagnosis time series. Earliness using each of the four new time series was measured and the results compared with the original result for the four urban regions in Pennsylvania using Student's t-test.

To determine whether the results were sensitive to manual selection of the nonoutbreak periods (and thus potentially biased by that procedure), the analysis was performed also using an objective definition for the nonoutbreak period. The alternate definition was the period June through September, chosen because those four months had the lowest average weekly hospital admission counts in all regions.

Effect of SIGNAL Strength on Earliness

In this report, signal strength was measured as the number of standard deviations that the EWMA-smoothed curve at its peak exceeded the mean of the nonoutbreak period (signal strength has a different meaning in the field of electrical engineering). Because there was variability in signal strength between weekly electrolyte sales counts and hospital diagnoses as well as across years and regions, any variability in earliness measurements between regions and years could have been explained by differences in signal strengths. For example, if the signal strength of the electrolyte sales signal were weaker than that of hospital diagnoses, then its time of detection could be later. Thus, the earliness of detection may be a function of the relative strengths of the two signals. *Relative signal strength* in this study was the ratio of the signal strength of electrolyte sales to the signal strength of hospital diagnoses. A relative signal strength greater than 1 means that electrolyte sales were the stronger signal during an outbreak, a ratio less than 1 means that hospital diagnoses were the stronger signal, and a ratio equal to 1 means the two signals had the same strength. To determine the influence of relative signal strength on the measurement of earliness, a plot of relative signal strength versus earliness was made, a simple linear regression performed, and the coefficient of determination (r^2) computed.

IRB Approval

The institutional review board of each participating institution approved the conduct of this study.

Results

Table 2 shows the range of weekly counts of electrolyte sales and hospital diagnoses for the six urban regions. Average sales ranged from 85.0 sales per 10,000 children per week in Indianapolis to 128.9 in Harrisburg. Hospital diagnoses for respiratory and diarrheal diseases varied from an average of 1.9 per 10,000 children per week in Indianapolis to 4.0 in Philadelphia. The urban regions with the largest sales-to-diagnoses ratio were Indianapolis (because it has the lowest incidence of diagnoses) and Harrisburg (because it had the highest incidence of sales).

There were 18 outbreaks of pediatric respiratory and diarrheal disease in the six urban regions (Fig. 1). For the 18 outbreaks, there was strong correlation of electrolyte sales with hospital diagnoses with a correlation coefficient of 0.90 (95% CI, 0.87–0.93), which occurred at a time latency of 1.7

Table 2 ■ The Six Urban Regions and Rates of Sales and Admissions in Each

Urban Region	Counties Comprising Region	Electrolyte Sales*	Hospital Diagnoses*	Average Ratio	Population ≤5 Years Old	Total Population
Harrisburg	Cumberland, Dauphin	128.9 (54.1–447.1)	2.8 (0.0–14.0)	46.5	27,152	475,446
Indianapolis	Boone, Hamilton, Hancock, Hendricks, Madison, Marion, Morgan, Johnson, Shelby	85.0 (25.9–271.1)	1.9 (0.3–9.9)	45.5	119,751	1,557,997
Philadelphia	Bucks, Delaware, Lehigh, Montgomery, Philadelphia	115.1 (59.3–282.7)	4.0 (0.6–13.2)	28.5	236,895	3,495,537
Pittsburgh	Allegheny, Beaver, Butler, Fayette, Washington, Westmoreland	88.3 (47.1–224.4)	3.0 (0.3–12.5)	29.6	130,946	2,349,694
Salt Lake	Davis, Salt Lake, Utah, Weber	88.9 (44.0–173.3)	3.0 (0.4–14.6)	29.2	161,239	1,643,281
Scranton	Lackawanna, Luzerne	117.1 (53.3–315.3)	3.9 (0.0–15.9)	29.9	27,024	508,849

*Average (minimum–maximum) per 10,000 children at or less than age 5 per week.

Table 3 ■ List of Additional ICD-9-CM Codes for Sensitivity Analysis

Dehydration	Asthma	Other Respiratory and Diarrheal Illness	Otitis Media and Codes with Lower Admissions Counts
276.5—Volume depletion	493.01—Extrinsic asthma with status asthmaticus 493.91—Asthma, unspecified type, with status asthmaticus	465.9—Acute upper respiratory infections of unspecified site 464.4—Croup 558.9—Other and unspecified noninfectious gastroenteritis and colitis	382.9—Unspecified otitis media 493.00—Extrinsic asthma without mention of status asthmaticus 493.90—Asthma, unspecified type, without mention of status asthmaticus 786.09—Other dyspnea and respiratory abnormality

weeks (95% CI, 0.50–2.9). The average time latency of 1.7 weeks means that the electrolyte sales time series precedes that of hospital diagnoses by 1.7 weeks on average. The minimum correlation for an outbreak was 0.75, and the maximum was 0.99.

Using EWMA with a nine-sigma threshold (the lowest with 100% sensitivity and specificity), detection of the outbreaks from retail electrolyte sales preceded detection from hospital diagnoses by an average of 2.4 weeks (95% CI, 0.1–4.8). For 12 outbreaks, detection occurred from sales prior to detection from diagnoses, for two outbreaks, detection occurred the same week, and for four, detection from electrolyte sales occurred after detection from diagnoses. In general, there was relatively high variability across outbreaks.

Differences among regions and years in relative signal strength for the 18 outbreaks explained only part of the variability in earliness. Figure 2A is an example taken from the outbreak with the largest relative signal strength (ratio of 3.1). The strength of the electrolyte sales signal was nearly 30 standard deviations, and the strength of the hospital diagnoses signal was nearly 10 standard deviations. Figure 2A also shows the nine-sigma detection thresholds to illustrate the differential in the time of detection from electrolyte sales and hospital diagnoses.

The scatterplot of relative signal strength versus earliness shown in Figure 2B was fitted by simple linear regression. The coefficient of determination from the linear regression analysis was 0.26, meaning that differences in relative signal strength explains 26% of the variation in earliness found in this study.

The results were not sensitive to use of different detection thresholds. For the range 9.0 to 21.0 sigma, average earliness ranged from a minimum of 2.4 weeks to a maximum of 4.7 (Table 4, available in an online data supplement at www.jamia.org). The number of outbreaks in which detection from retail sales of electrolytes occurred prior to detection from hospital diagnoses varied with the choice of sigma from 12 to 16 (Table 4, available in an online data supplement at www.jamia.org).

For four alternative sets of ICD-9-CM codes, earliness of detection in the four Pennsylvania urban regions was not significantly different from the original result for those four urban regions (p-values 0.27 to 0.47).

The results also were not sensitive to choice of nonoutbreak period. Using a four-month nonoutbreak period of June through September, earliness was 3.2 weeks (95% CI, 1.2–5.2) at 9.5 sigma (lowest with 100% sensitivity and specificity). At ten sigma, earliness was 3.4 weeks compared with the original result of 2.9 weeks.

Discussion

The correlation of seasonal outbreaks of pediatric diarrheal and respiratory diseases and sales of electrolyte products was strong and repeatedly observed over many locations and

Table 5 ■ Mean Earliness of Detection for Various Sigma Detection Thresholds

Detection Threshold	Earliness (wk)*	Earliness 95% CI	Sensitivity (%)	Specificity (%)	Electrolytes Precede Admissions*	Electrolytes No Later than Admissions*	Admissions Precede Electrolytes*
8.0			100	89			
9.0	2.4	(0.1–4.8)	100	100	12	14	4
10.0	2.9	(0.8–5.1)	100	100	13	15	3
11.0	3.3	(1.3–5.3)	100	100	13	15	3
12.0	3.2	(1.1–5.2)	100	100	13	14	4
13.0	3.6	(1.7–5.5)	100	100	14	16	2
14.0	3.9	(2.1–5.7)	100	100	16	16	2
15.0	4.1	(2.2–6.0)	100	100	16	16	2
16.0	4.1	(2.1–6.0)	100	100	16	16	2
17.0	4.1	(2.4–5.8)	100	100	16	16	2
18.0	4.3	(2.6–6.0)	100	100	16	16	2
19.0	4.4	(2.7–6.1)	100	100	16	16	2
20.0	4.4	(2.7–6.1)	100	100	16	16	2
21.0†	4.7	(2.8–6.5)	100	100	16	16	2

*n = 18 outbreaks.
†Above a threshold of 21.0, EWMA failed to detect at least one outbreak from hospital diagnoses.
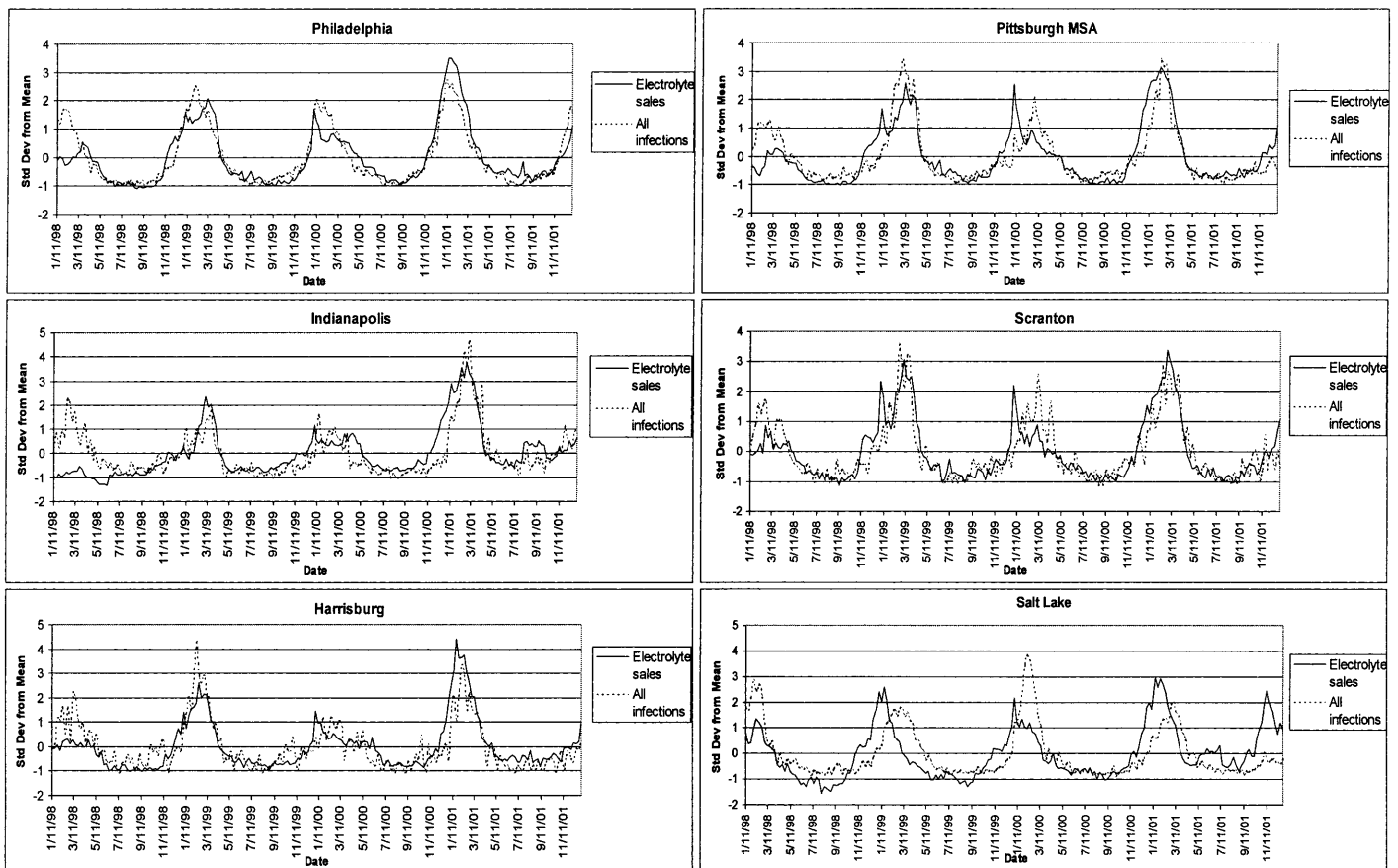
Figure 1. Sales of electrolyte products versus pediatric respiratory and diarrheal diseases. Each value is plotted as the number of standard deviations from the mean.

several years, supporting the first hypothesis that sales of electrolyte products contain a signal of outbreaks of respiratory and diarrheal infections in children.

The EWMA analysis of detection from electrolyte sales and hospital diagnosis suggests that monitoring of sales of electrolyte products may lead to earlier detection of seasonal outbreaks of respiratory and diarrheal infections in children; however, there was considerable variability in earliness from year to year and in different locations. This variability is partly a result of differences in signal strength, although that effect does not account for most of the variability. The variability may be a result of nosocomial outbreaks in some years and locations. It is known that RSV and rotavirus are a significant cause of pediatric nosocomial infections,[23,24] but there are no statistics about how often hospital-based outbreaks precede community outbreaks of these diseases, and we do not know for the study regions the degree of nosocomial activity for the study years. Future studies using hospital discharge datasets must remove nosocomial cases when possible. The variability may also be due to unobserved disease activity affecting sales of electrolyte products. Our sensitivity analysis of ICD-9-CM codes was designed to look for such a disease and found none, although the possibility remains that there is a disease in the outpatient setting partly explaining the variability. Further studies using outpatient diagnostic data could address this question. Another potential source of variability is which disease (among the ones included in the hospital diagnoses time series) was prominent

at the leading edge of the outbreak. For example, if RSV was prominent at the beginning of some outbreaks and rotavirus was prominent at the beginning of others, earliness may have varied because earliness may be different for these two diseases. Study of pure outbreaks is preferable (for example, the studies of *Cryptosporidium* outbreaks mentioned in the introduction), yet, achieving a large enough sample size to attain statistical confidence is difficult due to their rarity. The possibility also remains that the observed correlation between electrolyte sales and hospital diagnoses for 18 outbreaks in six urban regions is coincidental and that different phenomena that happen to have a seasonal timing are causing the two observed time series. Studies of observational data generally are unable to eliminate this possibility.

The confidence interval for the measurement of earliness of detection from sales of electrolytes relative to hospital diagnoses, nevertheless, did not include zero, suggesting that even with the variation introduced by the experimental conditions that detection occurs earlier from electrolyte sales than from hospital diagnoses. The fact that the study design was biased strongly in favor of hospital diagnoses only increases the confidence in this result. Although further study is needed, the results obtained are supportive of our second hypothesis that sales of electrolyte products can provide an early warning of pediatric outbreaks.

The sensitivity analyses show that the results were not sensitive to many of the experimental conditions. The results
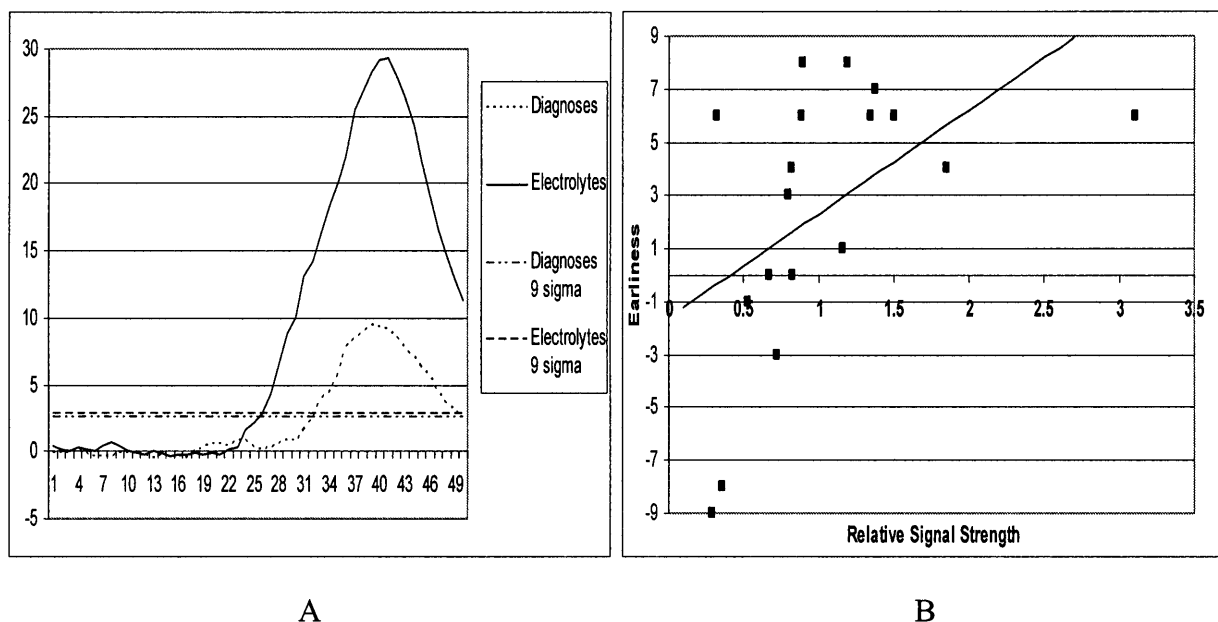
Figure 2. Analysis of effect of signal strength on earliness. (**A**) Electrolytes are a stronger signal than admissions and the earliness of detection appears to be largely due to this factor. (**B**) Scatterplot of relative signal strength versus earliness. The solid line is a line fitted to the points using simple linear regression.

were not sensitive to selection of detection threshold, the choice of sets of ICD-9-CM codes to compile diagnoses into category counts, or the definition of the nonoutbreak periods. The nine-sigma threshold is a reasonable level to operate EWMA on electrolyte sales and hospital diagnoses data as it represents 1.2 to 3.2 nonoutbreak standard deviations, and there is not a requirement that the outbreaks be as large as the ones we studied in which peak electrolyte sales and hospital diagnoses ranged from approximately five to 45 nonoutbreak standard deviations from the nonoutbreak mean. In practice, an even lower detection threshold might be used, but at the cost of false alarms.

The relevance of this study to the problem of early detection of bioterrorism lies in the finding that electrolyte sales correlate with pediatric diseases that are respiratory or diarrheal (as would be expected as dehydration due to such diseases is the indication for treatment with electrolyte products). Many "bioterrorist" diseases have respiratory or diarrheal presentations. For example, a number of CDC Category B Agents[25] (organisms that represent a bioterrorism threat) such as *Cryptosporidium parvum*, *Shigella* species, *Vibrio cholerae*, and *Salmonella* species resemble rotavirus in their ability to cause secretory diarrhea in children that is likely to result in dehydration.

This study adds pediatric electrolytes to the list of products for which there is evidence to support routine monitoring of sales for disease outbreaks. Several large cities, including Washington, DC, and New York City, monitor sales of OTC products, and the results of this study suggest that electrolyte products be included in such monitoring. This study also showed the importance of "large *N*" studies as opposed to studies of single outbreaks by revealing variability in earliness measurements and identifying possible causes of the variability and directions for future study. One such cause is differences in signal strength, which needs to be considered

in all future studies of the relative earliness of detection from one type of data relative to another.

The methods used in this study generally are applicable to other types of data to answer the two key questions of detection research: *Do the data contain signal of outbreaks of some types of disease?* and *how early can the outbreak be detected?* Other types of data that are of immediate interest are other OTC health care products (e.g., cough syrup), poison center calls, 911 calls, transit data, absenteeism, chief complaints from emergency room and doctors' office visits, diagnoses from emergency room and doctors' office visits, orders for (and sales of) prescription medications, volume of telephone calls to medical providers, and telephone triage data (that is, data about calls to nurse-staffed call centers run by managed care organizations).

Conclusion

Sales of electrolyte products correlate strongly with disease in children aged 5 years and younger. In 12 of 18 outbreaks studied, the initial rise in electrolyte sales was detectable weeks before the rise in hospital diagnoses of respiratory and diarrheal illnesses. The variability in earliness of detection from year to year and in different locations is an important subject of future research.

References ■

1. Wagner MM, Tsui FC, Espino JU, et al. The emerging science of very early detection of disease outbreaks. J Public Health Manag Pract. 2001;7(6):51–9.
2. Labrie J. Self-care in the new millennium: American attitudes towards maintaining personal health. Consumer Healthcare Products Association, 2001. Available at: http://www.chpa-info. org/pdfs/CHPA%20Final%20Report%20revised%20(03-20)_.pdf. Accessed Sept 8, 2003.

3. McIsaac WJ, Levine N, Goel V. Visits by adults to family physicians for the common cold. J Fam Pract. 1998;47:366–9.

4. Zeng X, Wagner MM. Modeling the effects of epidemics on routinely collected data. Proc AMIA Symp. 2001:781–5.

5. Corso PS, Kramer MH, Blair KA, Addiss DG, Davis JP, Haddix AC. Cost of illness in the 1993 waterborne Cryptosporidium outbreak, Milwaukee, Wisconsin. Emerg Infect Dis. 2003;9:426–31.

6. Goldstein A. Strategic tracking of sniffles. Scientists on alert for terrorism monitor area health factors. Washington Post. March 28, 2003:A10.

7. Perez-Pena R. An early warning system for diseases in New York. New York Times. <http://www.nytimes.com/2003/04/04/nyregion/04WARN.html?ex=1053489600&en=a50eec53d50d3da6&ei=5070>. Accessed April 04, 2003.

8. Wagner MM, Robinson JM, Tsui F-C, Espino JU, Hogan WR. Design of a national retail data monitor for public health surveillance. J Am Med Inform Assoc. 2003;10:409–18.

9. Welliver RC, Cherry JD, Boyer KM, et al. Sales of nonprescription cold remedies: a unique method of influenza surveillance. Pediatr Res. 1979;13:1015–7.

10. Magruder S, Florio E. Evaluation of over-the-counter pharmaceutical sales as a possible early warning indicator of public health. Johns Hopkins Univ Appl Phys Lab Techn Dig. 2003. 2003;24(4): (in press).

11. Angulo FJ, Tippen S, Sharp DJ, et al. A community waterborne outbreak of salmonellosis and the effectiveness of a boil water order. Am J Public Health. 1997;87:580–4.

12. Stirling R, Aramini J, Ellis A, et al. Waterborne cryptosporidiosis outbreak, North Battleford, Saskatchewan, Spring 2001. Can Commun Dis Rep. 2001;27(22):185–92.

13. Rodman J, Frost F, Davis-Burchat L, Fraser D, Langer J, Jakubowski W. Pharmaceutical sales: a method of disease surveillance? J Environ Health. 1997;Nov:8–14.

14. 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(1):43–54.

15. 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.

16. Nappert G, Barrios JM, Zello GA, Naylor JM. Oral rehydration solution therapy in the management of children with rotavirus diarrhea. Nutr Rev. 2000;58(3 pt 1):80–7.

17. Practice parameter: the management of acute gastroenteritis in young children. American Academy of Pediatrics, Provisional Committee on Quality Improvement, Subcommittee on Acute Gastroenteritis. Pediatrics. 1996;97:424–35.

18. Lobanov B. Classification of Russian vowels spoken by different speakers. J Acoust Soc Am. 1971;49:606–8.

19. Proakis J, Manolakis D. Digital Signal Processing: Principles, Algorithms, and Applications (ed 2). New York: Macmillan Publishing Company, 1992, .

20. Tsui FC, Wagner MM, Dato V, Chang CC. Value of ICD-9 coded chief complaints for detection of epidemics. Proc AMIA Symp. 2001:711–5.

21. Roberts S. Control chart tests based on geometric moving averages. Technometrics. 1959;1:97–101.

22. Morton AP, Whitby M, McLaws ML, et al. The application of statistical process control charts to the detection and monitoring of hospital-acquired infections. J Qual Clin Pract. 2001;21:112–7.

23. Ford-Jones E, Mindorff C, Langley J, et al. Epidemiologic study of 4684 hospital-acquired infections in pediatric patients. Pediatr Infect Dis J. 1989;8:668–75.

24. Raymond J, Aujard Y. Nosocomial infections in pediatric patients: a European, multicenter, prospective study. Infect Control Hosp Epidemiol. 2000;21:260–3.

25. Rotz LD, Khan AS, Lillibridge SR, Ostroff SM, Hughes JM. Public health assessment of potential biological terrorism agents. Emerg Infect Dis. 2002;8:225–30.