Assessment of administrative claims data for public health reporting of *Salmonella* in Tennessee

Ellyn Marder1,2, Katie Garman1, Timothy F Jones1, John Dunn1, Stephen Jones3

ABSTRACT

In the USA, approximately 4% of the estimated 1 million *Salmonella* infections occurring annually are reported to public health. Administrative claims data from large health insurance companies capture disease-specific data which could potentially enhance public health surveillance. To determine the utility of medical claims data for public health reporting of *Salmonella*, we assessed medical claims data from BlueCross BlueShield of Tennessee (BCBST) members compared to Tennessee Department of Health (TDH) surveillance data. BCBST *Salmonella* cases diagnosed during 2007–2011 were matched to TDH *Salmonella* cases reported during the same time period. Matches and non-matches were validated using medical records. Of the 450 BCBST cases identified, 72% matched TDH cases. All culture-confirmed BCBST cases were reported to TDH. Non-matched BCBST cases included clinical diagnoses which were culture negative or not tested. Our findings indicate administrative claims data are not currently a viable mechanism for enhancing routine reporting of *Salmonella* infections.

Key words: *Salmonella*, public health surveillance, administrative claims

INTRODUCTION

*Salmonella* is a major cause of foodborne illness in the USA. Non-typhoidal *Salmonella* spp. infection causes acute gastroenteritis including nausea, vomiting, diarrhea, and fever. Although most infections cause mild illness, serious disease (eg, bloodstream infection) resulting in hospitalization or death can occur. Each year, approximately 42 000 cases of culture-confirmed *Salmonella* are reported in the USA.1 However, this represents only 4% of the >1 million *Salmonella* infections that are estimated to occur annually. Numerous steps must occur for a case to be reported through public health surveillance: an exposed person must develop illness, seek healthcare, and provide a stool specimen from which *Salmonella* is isolated, and then the positive culture must be reported to public health. Since most *Salmonella* infections are managed symptomatically, frequently, not all of these steps occur. Therefore, many *Salmonella* cases are unlikely to be ascertained by public health. Methods to ensure ascertainment of all cases are a priority for public health.2

In Tennessee, state law requires that all cases of salmonellosis be reported to the Department of Health. Under-reporting likely occurs as healthcare providers often do not report diseases and rely on laboratories to report.3 In an effort to enhance reporting, the Tennessee Department of Health (TDH) has conducted active, population-based surveillance as part of the Centers for Disease Control and Prevention’s (CDC) Foodborne Diseases Active Surveillance Network (FoodNet) since 2000. FoodNet is a collaborative program between CDC, 10 state health departments, the US Department of Agriculture’s Food Safety Inspection Service (USDA-FSIS), and the US Food and Drug Administration (FDA).

Other data sources have been explored to conduct or enhance public health surveillance activities. Public health agencies have previously used administrative data for infectious disease surveillance purposes, specifically syndromic disease surveillance.4,5 In addition to syndromic surveillance, which typically uses aberration detection to detect disease clusters, administrative data from large health insurance companies, which capture disease-specific claims data for billing purposes, could potentially be used for infectious disease case reporting.6,7 A previous study used BlueCross BlueShield of Tennessee (BCBST) administrative claims to identify differences in reporting rates of Lyme disease and other arthropod-borne infections between administrative claims and public health surveillance data in Tennessee.8 However, the authors did not directly compare data to describe the cases common to both systems. To determine the utility of medical claims data for public health case reporting of *Salmonella*, we assessed medical claims data from BCBST members compared to TDH surveillance data.
METHODS
BCBST is a managed care organization located in Tennessee serving approximately 50% of Tennesseans. BCBST collects data for every encounter a member has with the healthcare system (eg, visiting a doctor or hospital, laboratory testing) if an insurance claim is filed. International Classification of Diseases, 9th edition (ICD-9) codes and Current Procedural Terminology (CPT)-4 codes are recorded for each encounter, indicating diagnoses and procedures, respectively, associated with the encounter. However, administrative claims data are designed for managing claims and reimbursement rather than surveillance.

In this study, a BCBST case was defined as a member’s incident medical encounter during January 2007 to December 2011 with an ICD-9 code indicative of Salmonella infection (ICD-9: 003.0). A TDH case was defined as the incident Salmonella-positive culture for a person reported to TDH during January 2007 to December 2011. Subsequent Salmonella-positive cultures ≥30 days from the incident Salmonella-positive culture were considered a new case. Matched cases were BCBST cases identified in TDH data by an algorithm based on first name, last name, birth date, sex, pathogen, and a difference of ≤30 days between the claim date of service and TDH specimen collection date. Non-matched cases were BCBST cases not identified in TDH data by the matching algorithm. Matched and non-matched cases were validated using TDH surveillance data and medical records. Non-matched cases were considered reported to public health if there were ≤180 days between the incidence date of service and the closest TDH specimen collection date. Medical records were requested at least three times from diagnosing providers.

The study protocol was approved by the TDH and BCBST Institutional Review Boards. Analyses were completed using SAS software, V.9.3 (SAS Institute, Cary, North Carolina, USA). Differences in proportions were assessed by the \( \chi^2 \) test of proportions or Fisher’s exact test, depending on sample size, with an \( \alpha \)-level of 0.05. Differences in means were assessed with the t test.

RESULTS
There were 4630 Salmonella cases reported to TDH and 450 BCBST Salmonella cases identified during the study period. The mean annual incidence of Salmonella based on TDH data was 14.8 per 100,000 population (range: 12.8–17.1 per 100,000 population). The mean annual incidence of Salmonella among BCBST members was 2.9 per 100,000 population (range: 2.4–3.5 per 100,000 population) (figure 1). Of the 450 BCBST cases diagnosed during the study period, 325 (72.2%) matched TDH cases. For each year of the study period, at least 50% of the BCBST cases diagnosed in that year matched TDH cases using the matching algorithm. The BCBST matched cases displayed seasonality similar to the TDH cases with peaks occurring during the summer months, whereas the BCBST non-matched cases did not display the same seasonality (figure 2). BCBST matched cases displayed a similar bimodal age distribution to the TDH cases with at least 25% of cases under the age of 5 years and at least 25% of cases aged 20–64 years.

To assess the matching algorithm, we reviewed the 125 non-matches. Of these, 24 (19.2%) were reported to TDH but were not identified by the matching algorithm: 11 (45.8%) had a TDH specimen collection date >30 days different than the BCBST date of service, nine (37.5%) had mismatching demographics, including first name, last name, and date of birth, and four (16.7%) were unconfirmed cases (figure 3). Forty-three (12.3%) of the 350 BCBST cases reported to public health received a Salmonella diagnosis before Salmonella was isolated. Among these cases, the mean difference between the BCBST date of service and TDH specimen collection date was 3.3 days (median: 1.0 day, range: 1–41 days). Medical records were available for 45 non-matches. There were no statistically significant differences in age or gender between those for whom charts were and were not available. Of the 45 with available medical records, 15 (33.3%) BCBST Salmonella cases had documentation of a negative stool culture and 30 (66.7%) had no record of a laboratory stool culture included in the medical chart (figure 3).

DISCUSSION
We found that over 75% of the Salmonella cases identified through BCBST billing data were reported to TDH. All culture-confirmed BCBST cases were reported to TDH. Of the 22.4% of BCBST cases not reported to TDH, none were culture confirmed. Specific causes of infectious gastroenteritis can rarely be distinguished clinically, so other causes of acute gastroenteritis likely account for some of the non-matched cases which had a Salmonella diagnosis code in BCBST data.
Of the 45 BCBST non-matched cases for whom medical records were reviewed, all were clinical diagnoses which were culture negative or not tested. Among *Salmonella*-diagnosed BCBST members, we found that for every stool culture performed, there were two other *Salmonella*-diagnosed BCBST members who sought care and were not cultured. This is similar to previous research that determined there were approximately 2.3 persons with diarrhea who sought care for every stool culture request. Laboratory-based surveillance depends on stool cultures for patients with acute gastroenteritis, although this is not always indicated for clinical management. From a public health perspective, culture of stool of patients with acute gastroenteritis is beneficial for accurate surveillance and detection of outbreaks in need of public health action.

Active surveillance is dependent on laboratory surveillance, which introduces inherent lags between illness onset date and isolation date. Using medical claims data to supplement reporting might allow public health to identify potential cases or outbreaks sooner. The medical claims date of service was before the specimen collection date for 12.3% of cases reported to TDH. Over the course of a year, this means nearly 120 cases could potentially be investigated sooner and outbreaks might be identified faster.

This study has several limitations. Most importantly, BCBST cases are determined by ICD-9 codes, which represent medically diagnosed cases, whereas TDH cases are confirmed by laboratory testing. Although ICD-9 codes have been used for public health surveillance efforts previously, these attempts have been mostly limited to syndromic disease surveillance, which does not include individual case reporting. The BCBST *Salmonella* diagnosis is a sensitive but non-specific, non-standardized case definition which we attempted to use to inform a highly specific, standardized case definition for TDH. However, a North Carolina study determined that ICD-9 codes for many communicable diseases, including non-typhoidal *Salmonella*, had positive predictive values high enough to be useful for communicable disease surveillance. Additionally, medical records were available for only 40 (40.4%) cases not reported to TDH. Lastly, discrepancies exist between the BCBST and TDH databases. Cases were matched using two different dates: BCBST date of service, which represents the date a BCBST member had a healthcare encounter recorded with the *Salmonella* ICD-9 code, and the TDH specimen collection date, which represents the date a stool specimen was collected. To account for this difference, we matched cases using a ≤30-day window between specimen collection date and date of service.
Previous research has suggested that administrative claims data have the potential to enhance population health surveillance, but this needs to be assessed on a pathogen-specific basis.8,12 Our study assessed the utility of administrative claims data for public health case reporting of *Salmonella*. No culture-confirmed *Salmonella* cases were identified through BCBST that were not reported to TDH through routine mechanisms. A small proportion of the BCBST cases were reported sooner than they would have been detected otherwise, which might improve outbreak detection. Overall, the minimal benefit of using these administrative data for surveillance, and the substantial resources required to do so, make this a non-viable mechanism for routine public health surveillance for *Salmonella* infection.

**ACKNOWLEDGEMENTS**

The authors thank Anuradha Penaganti and TDH FoodCORE staff for their assistance. This study was supported in part by an appointment to the Applied Epidemiology Fellowship Program administered by the Council of State and Territorial Epidemiologists (CSTE).

**CONTRIBUTORS**

EM, TFJ, JD, and SJ conceived of and designed the study. EM analyzed the data. EM, KG, TFJ, JD, and SJ contributed to the writing and review of the paper.

**COMPETING INTERESTS**

None.

**ETHICS APPROVAL**

Tennessee Department of Health and BlueCross BlueShield of Tennessee Institutional Review Boards approved this study.

**PROVENANCE AND PEER REVIEW**

Not commissioned; externally peer reviewed.

**REFERENCES**


AUTHOR AFFILIATIONS

1Tennessee Department of Health, Communicable and Environmental Diseases and Emergency Preparedness, Nashville, Tennessee, USA
2CDC/CSTE Applied Epidemiology, Atlanta, Georgia, USA
3BlueCross BlueShield of Tennessee, Chattanooga, Tennessee, USA