Using administrative medical claims data to supplement state disease registry systems for reporting zoonotic infections

Stephen G Jones,[1] Steven Coulter,[2] William Conner[3]

[1]BlueCross and BlueShield of Tennessee, Chattanooga, Tennessee, USA
[2]BlueCross BlueShield of Tennessee Health Institute, Chattanooga, Tennessee, USA
[3]Forestry and Natural Resources Department, Clemson University, Georgetown, South Carolina, USA

Correspondence to
Dr Stephen G Jones, BlueCross and BlueShield of Tennessee, 1 Cameron Hill Circle, Bldg 2.1, Chattanooga, TN 37402, USA; stephen_jones@bcbst.com

ABSTRACT

Objective To determine what, if any, opportunity exists in using administrative medical claims data for supplemental reporting to the state infectious disease registry system.

Materials and methods Cases of five tick-borne (Lyme disease (LD), babesiosis, ehrlichiosis, Rocky Mountain spotted fever (RMSF), tularemia) and two mosquito-borne diseases (West Nile virus, La Crosse viral encephalitis) reported to the Tennessee Department of Health during 2000—2009 were selected for study. Similarly, medically diagnosed cases from a Tennessee-based managed care organization (MCO) claims data warehouse were extracted for the same time period. MCO and Tennessee Department of Health incidence rates were compared using a complete randomized block design within a general linear mixed model to measure potential supplemental reporting opportunity.

Results MCO LD incidence was 7.7 times higher ($p<0.001$) than that reported to the state, possibly indicating significant under-reporting ($\sim$196 unreported cases per year). MCO data also suggest about 33 cases of RMSF go unreported each year in Tennessee ($p<0.001$). Three cases of babesiosis were discovered using claims data, a significant finding as this disease was only recently confirmed in Tennessee.

Discussion Data sharing between MCOs and health departments for vaccine information already exists (eg, the Vaccine Safety Datalink Rapid Cycle Analysis project). There may be a significant opportunity in Tennessee to supplement the current passive infectious disease reporting system with administrative claims data, particularly for LD and RMSF.

Conclusions There are limitations with administrative claims data, but health plans may help bridge data gaps and support the federal administration's vision of combining public and private data into one source.

BACKGROUND AND SIGNIFICANCE

Notifiable diseases are infectious diseases which require regular, frequent, and timely reporting of individual diagnosed cases to aid in prevention and control (eg, Lyme disease (LD), giardiasis, salmonella).[1][2] Public health officials from state health departments collaborate annually with the Centers for Disease Control and Prevention (CDC) to determine which diseases should be listed. Reporting of disease cases by healthcare providers and laboratories is currently mandated only at the state level and therefore can vary from state to state.[2][3] In 2009, over 7000 cases of notifiable communicable diseases were reported to the Tennessee Department of Health (TDH) Communicable and Environmental Disease Services.[4] Although state regulations or contractual obligations may require the reporting of certain diseases, traditional passive reporting by the diagnosing clinician can be burdensome, incomplete, and delayed.[5] Thus, there is under-reporting of diseases as not all diagnosed or suspected cases are reported by healthcare providers,[6–10] and reporting can vary by physician specialty.[11] Many healthcare providers may not understand the importance of public health surveillance, and generally how, when, why, and what to report.[8][9]

With the recent passing of the Health Information Technology for Economic and Clinical Health (HITECH) Act and the Patient Protection and Affordable Care Act, there is a great opportunity to modernize healthcare delivery and improve the quality of care received in the USA. Through the HITECH Act, the Assistant Secretary for Planning and Evaluation for the Department of Health and Human Services is building a multi-payer, multi-claim database to support patient-centered research using public and private payer claims data. Combining public and private data into one source may alleviate certain problems surrounding existing disparate data sources and provide the ability to study less common conditions.[12] California is already planning for the integration of public and private data through their State Health Information Exchange Cooperative Agreement Program. A major initiative for them is to 'transmit and integrate data across multiple internal and external data sources' in order to 'prepare for and respond to emergencies, diseases, outbreaks, epidemics, and emerging threats.'[13]

Health insurance plans (ie, managed care organizations, MCOs) could play a major role in the reporting of infectious diseases.[8][14] Kaiser Permanente is one such entity already engaged in collaborative disease reporting with the CDC through their Vaccine Safety Datalink (VSD) Rapid Cycle Analysis (RCA) project. Their goal is to monitor potential adverse events (side effects) following vaccinations in near real time. Section 164.512(b) of the federal Health Information Portability Accountability Act (HIPAA) establishes safeguards allowing health plans to report such data. It states that all healthcare providers, health plans, and clearinghouses are permitted to disclose protected health information without individual authorization for the reporting of disease and vital events, and the conducting of public health surveillance, investigations, and interventions.[15]

Medical encounter data are recorded within the healthcare system every time a patient visits their doctor or hospital for a medical service, fills a prescription medicine, or seeks a consultation from a clinician. Considering that more than 253 million Americans have health insurance and will most likely utilize that service when needed,[16] administrative data captured in medical encounters could serve as a useful data supplement to the current disease reporting system. Zoonotic diseases are infectious diseases that can be transmitted from or through animals to humans, and arthropods often act as vectors for transmission. Zoonotic diseases are of significant concern regarding public health and account for approximately 75% of recently emerging infectious diseases; approximately 60% of all human pathogens originate from animals.[17] The aim of our study was to determine what, if any, opportunity may exist for supplementing the state infectious disease registry system with administrative medical claims data. We measure opportunity by comparing disease incidence rates across two data sources for five tick-borne (LD, babesiosis, ehrlichiosis, Rocky Mountain spotted fever (RMSF), tularemia) and two mosquito-borne (West Nile virus (WNV), La Crosse viral encephalitis) diseases known to occur in the southeastern USA. Any measurable differences between the two systems will indicate an opportunity for potential data sharing efforts.

MATERIALS AND METHODS
Case definition
Disease occurrence data within the proposed study area of Tennessee were collected from two separate data sources and tested for statistical differences. The first data source was administrative medical claims data obtained from BlueCross BlueShield of Tennessee, a large southeastern MCO located in Tennessee. The second data source was an extract provided by the TDH Center for Environmental and Communicable Diseases. A specific and separate case definition was needed for each data source due to the origin of the data.

Case definition—MCO
The participating MCO insures approximately 50% of the entire state's population. For the purposes of this study, cases are defined as all medical claims filed to the MCO having a primary or secondary arthropod-borne disease diagnosis code of interest with at least three separate corroborating events, using the member's first recorded occurrence (see below). Claims were extracted for the January 1, 2000—December 31, 2009 time period using the exact ICD-9 diagnosis code for disease identification (eg, 066.4: West Nile virus), not the more generic root level diagnosis available to providers (066: Other arthropod-borne viral diseases). In this manner, the minimum (ie, lower limit of) disease incidence according to MCO data is estimated .

Tick-borne diseases include:
▶ Babesiosis (ICD-9 code: 088.82)
▶ Borreliosis—LD (ICD-9 code: 088.81)
▶ Ehrlichiosis—human monocytic ehrlichiosis (HME) (ICD-9 code: 082.41)
▶ Rickettsiosis—RMSF (ICD-9 code: 082.0)
▶ Tularemia (ICD-9 code: 021)
Mosquito-borne diseases include:
▶ La Crosse viral encephalitis (ICD-9 code: 062.5)
▶ West Nile virus (WNV) (ICD-9 code: 066.4)

These diseases were selected because they represent a growing public health concern, are known to occur in the southeastern USA,[17] and provide a mix of relatively common cases (eg, borreliosis, rickettsiosis) and very rare cases (eg, babesiosis,

tularemia) for study. Using the MCO database, any patient receiving medical services for one of the selected diseases prior to the start of the study period (January 1, 2000) or after the study period (December 31, 2009) was removed from the analysis. To best ensure the diagnosis for the patient was in fact their first diagnosis, only the first recorded diagnosis date for each patient was retained. Any subsequent claims for the patient were removed and not considered in this study. To improve data validity, only claims with at least three instances of corroborating evidence were retained; here, corroboration occurred if the same diagnosis code in question was mentioned in separate line items on the same claim and/or on any follow-up claims. (A 'line item' is a term used to describe the structure of data storage on a medical claim. For every clinical procedure performed on a patient within a given claim, each procedure and an accompanying medical diagnosis will be recorded in separate line items.) For example, if a claim had only two line items with a primary diagnosis code of LD, this claim was not considered for study. This analysis utilized the exact diagnosis code for disease identification and allowed the minimum (ie, lower limit of) incidence according to MCO data to be estimated. Institutional Review Board approval was not needed for this study because only administrative claims data were examined.

Case definition—TDH
The second data source was an extract provided by the TDH, Center for Environmental and Communicable Diseases detailing all notifiable 'confirmed' or 'probable' disease cases reported to the State of Tennessee during the study period according to the CDC guidelines.[18] For example, the CDC case definitions for LD are as follows:
▶ Confirmed case—A case of erythema migrans ('bulls-eye' rash) with known exposure, or a case of erythema migrans with laboratory evidence of infection and without a known exposure, or a case with at least one late manifestation that has laboratory evidence of infection.
▶ Probable case—Any other case of physician-diagnosed LD that has laboratory evidence of infection.

This resulted in a comparison of medically diagnosed cases to CDC confirmed or probable disease cases. For example, a medical diagnosis for LD should be based on an individual's history of possible exposure to ticks that carry LD, the presence of typical signs and symptoms, and the results of blood tests. Data from different sources can be compared because specific ICD-9 diagnosis codes and not generic root-level codes were used to extract medical claims. Thus we assumed if a clinician codes at this detail, they have evidence to support the diagnosis beyond suspicion. The pathogens of interest are '[not] likely to be diagnosed clinically without confirmatory testing - either [providers] make the right diagnosis with lab confirmation or they never make the diagnosis at all' (Laura Cooley, MD, infectious disease specialist, personal communication, November 8, 2010). Furthermore, this very difference and ambiguity is the main focus of the study, which is to compare the difference between state reported incidence rates and medically diagnosed case rates to determine if claims data could support the current surveillance system. Because TDH serves as the compiler of all data sources to the state level, these data represent a theoretically complete set of reported cases for the state under the assumption that no under-reporting exists.

Statistical analyses
To estimate if and to what extent under-reporting of notifiable diseases exists, a randomized control block design was employed within a generalized linear mixed model (GLMM) approach to

compare TDH and MCO case counts. For the purposes of this study, the degree to which under-reporting is present is used to measure the amount of 'opportunity' that may exist in using administrative claims data to supplement the state registry system. GLMM models are particularly useful in estimating trends in disease rates where the response variable is not necessarily normally distributed.[19] [20] Input values into the models included a yearly (n=10) county (n=95) case total, which produced 950 observations for each data source. Separate models were built for each disease, and the response variable of interest was disease counts assumed to be Poisson distributed with a log-transformed population count as an exposure offset. Disease counts were expected to vary by county (ie, spatial heterogeneity) due to varying population denominators, socio-economic factors, and geographic and habitat character-istics.[21–24] Therefore, county was used as a blocking factor to remove the expected county-to-county variability when TDH values were compared with MCO values. Space (county) was considered a random effect, while time (year), data source (MCO vs TDH), and a time×data source interaction were considered fixed effects. Fixed effects were examined for statistical signifi-cance using the F test with an α level of 0.05. Variability in case counts across counties was tested using covariance tests.

RESULTS

Approximately 58 million medical claims were filed to the MCO during the 2000−2009 study period. Of these, 3535 claims contained a primary or secondary diagnosis for one of the seven described arthropod-borne diseases. After removal of invalid claims (patients without at least three separate ICD-9 entries for the disease, patients with claim dates starting or ending outside of the time period, duplicate patient entries, and patients having non-unique disease coding issues such as codes for RMSF and LD on the same claim), 1654 unique case-persons were distributed across the seven diseases of interest and remained for study. The majority of disease cases were LD (n=903; 55%), followed by RMSF (n=661; 40%). The remaining five diseases made up the residual 5% of disease cases. Three cases of babesiosis were found within the MCO claims data, specifically in Davidson, Lincoln, and Washington Counties.

Results from the GLMM suggest only LD and RMSF values are statistically different, as none of the other models converged (table 1). The average yearly numbers of medically diagnosed cases of LD from MCO data were 7.7 times higher than those reported to the state (F=835.44; p<0.0001). LD rates varied significantly over the 10-year study period (F=2.08; p=0.0283) and there was a significant temporal interaction with year×data

Table 1 Summary statistics and results from generalized linear mixed models for comparison of annual disease incidence per 100 000 population between MCO and TDH data

| | MCO | | TDH | | |
	Yearly mean	SD	Yearly mean	SD	F value
Lyme disease	3.76	0.80	0.49	0.13	835.44*
Babesiosis†	0.01	0.02	0.00	0.00	—
Rocky Mountain spotted fever	2.75	0.46	2.32	1.17	14.45*
Human monocytic ehrlichiosis†	0.06	0.06	0.59	0.34	—
Tularemia†	0.19	0.21	0.05	0.04	—
La Crosse viral encephalitis†	0.04	0.05	0.10	0.13	—
West Nile virus†	0.08	0.09	0.08	0.12	—

*Significant at p<0.05.
†Mixed models did not converge.
MCO, managed care organization; TDH, Tennessee Department of Health.

Table 2 Gender distribution across disease cases and data sources

| | MCO | | TDH | |
	Male n (%)	Female n (%)	Male n (%)	Female n (%)
LD	372 (44.1%)	531 (58.8%)	124 (42.5%)	168 (57.5%)
RMSF	348 (52.6%)	313 (47.4%)	783 (56.6%)	601 (43.4%)
HME	7 (50.0%)	7 (50.0%)	238 (68.4%)	110 (31.6%)
Tularemia	26 (61.9%)	16 (38.1%)	23 (79.3%)	6 (20.7%)
LACV	9 (90.0%)	1 (10.0%)	39 (65.0%)	21 (35.0%)
WNV	10 (47.6%)	11 (52.4%)	31 (67.4%)	15 (32.6%)
Totals*	772 (46.8%)	779 (53.2%)	1238 (57.3%)	921 (42.7%)

*6 TDH cases had unknown gender.
HME, human monocytic ehrlichiosis; LACV, La Crosse viral encephalitis; LD, Lyme disease; MCO, managed care organization; TDH; Tennessee Department of Health; RMSF, Rocky Mountain spotted fever; WNV, West Nile virus.

source (F=2.84; p=0.0026). Based on the residual pseudo-likeli-hood, a test of covariance suggests there is significant spatial variation in LD cases across the state (χ^2=84.8; p<0.0001). The average yearly numbers of medically diagnosed cases of RMSF from MCO data were 1.24 times higher than those reported to the state (F=14.45; p=0.0001). RMSF disease rates significantly varied over the 10-year study period (F=14.82; p<0.0001) and there was a significant temporal interaction with year×data source (F=10.14; p<0.0001). There was also significant spatial variation in RMSF cases across the state (χ^2=1135.01; p<0.0001). Gender distributions and the median ages of cases from MCO and TDH data were similar for LD and RMSF (tables 2 and 3). (Babesiosis was excluded as no TDH cases were recorded.)

Temporal trending indicates the aforementioned per 100 000 population rate differences varied from year to year (table 1), and MCO rates were not consistently higher throughout the study period (figure 1). LD rates from MCO data were consis-tently higher than TDH rates. TDH data indicated no evidence of babesiosis, but MCO data showed three separate cases in the years 2004, 2005, and 2009. RMSF rates from TDH data were lower than MCO rates for 2000−2005 and 2008, but were higher for 2006, 2007 and 2009 (figure 1).

DISCUSSION

Administrative medical claims data are an important resource for the research and surveillance of chronic diseases.[25] Our results suggest there may be an opportunity to supplement the current disease reporting system with administrative medical claims data. We make no attempts to quantify the number of cases shared between the data because our intent is to compare differences in disease rates at a population level, not an indi-vidual level. Thus at a minimum, the number of potential cases not reported to the state is the difference between the MCO rate and the TDH rate; therefore assuming a 100% overlap between MCO and TDH data, we identify the smallest possible benefit of utilizing MCO data. Although our personal interpretation, this work was recently presented to senior TDH

Table 3 Median age of recorded cases across disease categories and data sources

	MCO	TDH
Lyme disease	41	39
Rocky mountain spotted fever	36	44
Human monocytic ehrlichiosis	36	57
Tularemia	47	29
La Crosse viral encephalitis	8	8
West Nile virus	61	62

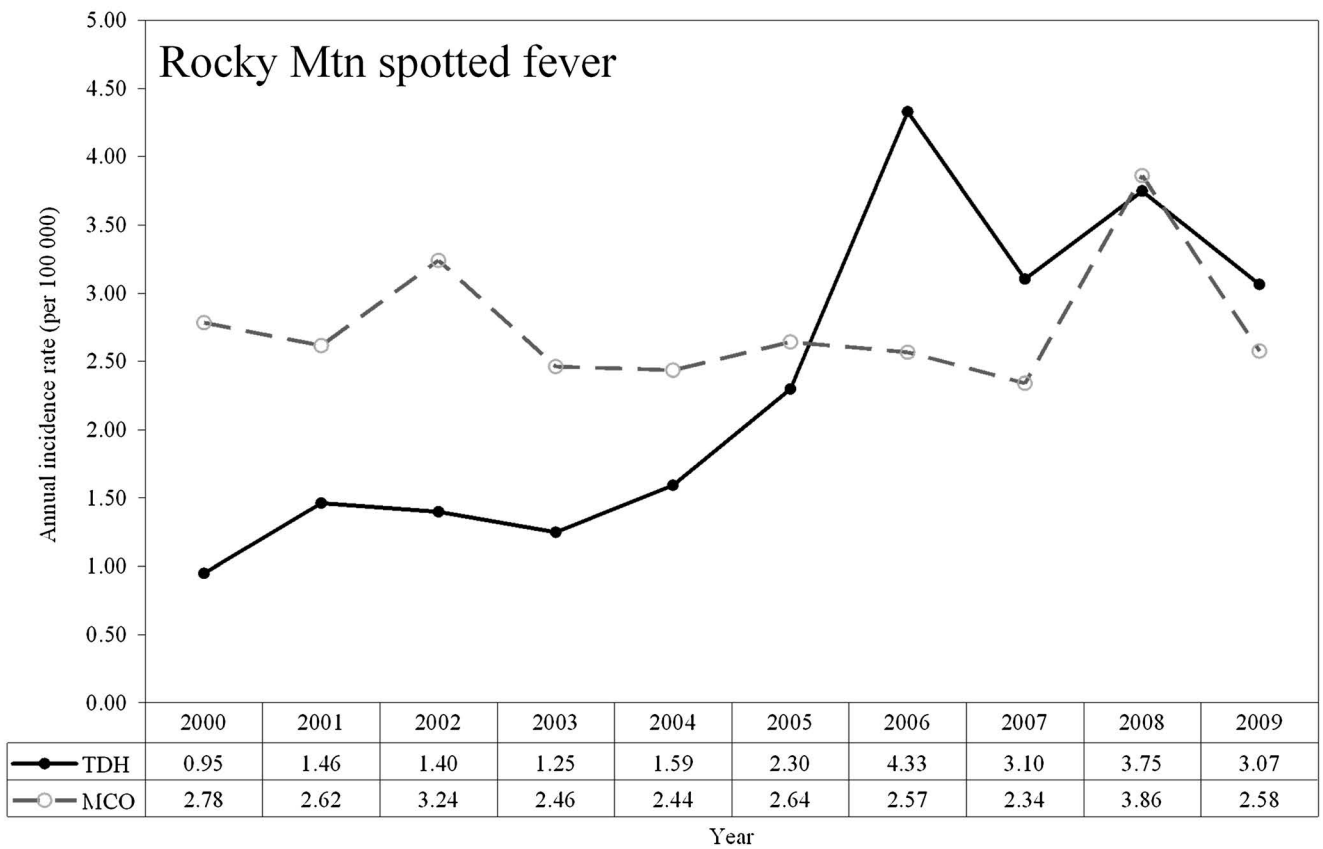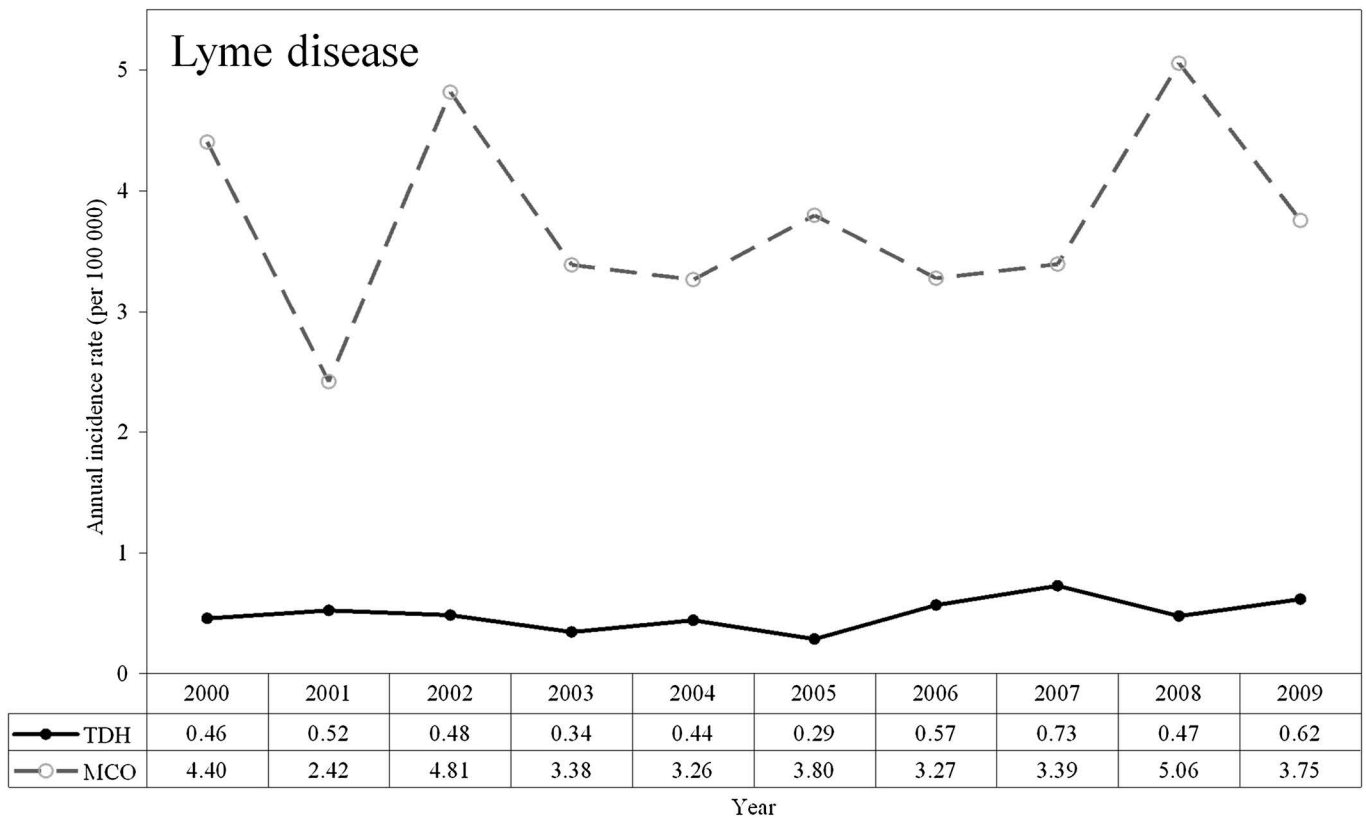MCO, managed care organization; TDH, Tennessee Department of Health.

Figure 1 Temporal comparison of Lyme disease (top) and Rocky Mountain spotted fever (bottom) incidence rates using managed care organization (MCO) medical claims data and Tennessee Department of Health (TDH) reported data.

epidemiologists/physicians who showed great enthusiasm and a desire to implement this type of data sharing. They recognize the importance of data enhancements and we are currently in discussions on how to best proceed. Kaiser Permanente and the CDC have laid the groundwork to engage in data sharing projects, and with great success having produced more than 75 scientific articles since the inception of its VSD program in 1990.[26] We envision a similar collaboration, providing data elements to the TDH such as ICD-9 diagnosis and CPT4-procedural codes and patient identifiers.

The authors understand the challenges that data sharing presents and the expected opposition by some to using claims data for case identification. We reduce the potential for claims error by including only exact diagnosis coding as opposed to root level coding, requiring at least three separate corroborations and retaining only the first recorded diagnosis. Admittedly, we are limited by the inability to relate, via a patient identifier, a medically diagnosed case to a CDC defined 'confirmed' or 'probable' case. However, it must be noted that, similar to MCO diagnosed cases, cases reported to the state are not necessarily laboratory confirmed either. According to the CDC guidelines, laboratory confirmation of LD is not necessary if the patient manifests with erythema migrans and the classic symptoms of the disease. Further, the sensitivity of ELISA testing on culture confirmed LD can vary from 37% to 70%, depending on the laboratory and methodology.[27]

LD rates reported to the state were below those of MCO administrative data, and the actual statewide prevalence rate over the study period may be 3.8 per 100 000 population, rather than the 0.49 derived from TDH statistics. This equates to an approximate 7.7-fold difference over the entire study period, resulting in an additional 1956 cases above the 292 reported to TDH. This suggests, on average, about 196 cases of LD go unreported each year in Tennessee and supports the body of evidence suggesting LD is under-reported, possibly up to 12-fold in some areas.[7 28] Although these diseases are required to be reported to the health department through the National Notifiable Disease Surveillance System, reporting is a voluntary process. It is known that in many cases, documentation of LD cases is incomplete, unavailable, and not submitted to the CDC.[29] However, the process of estimating a true prevalence rate is difficult, because there is also evidence suggesting LD cases are over-reported in areas not endemic for the disease,[30] possibly due to misdiagnoses[31 32] and the presence of clinical symptoms similar to those of other diseases such as southern tick associated rash illness.[30 33] This conflicting evidence does not invalidate the potential use of administrative claims data for supplemental reporting, but it does suggest the need for further investigation into integrating data sources and how best to leverage this additional information.

RMSF has been a reportable illness since the 1920s. RMSF rates were slightly higher according to the MCO data which suggest the actual number of cases in the state could have been 3.1 per 100 000 rather than 2.5 (an average difference of approximately 33 more cases per year). RMSF is the most severe and frequently reported tick rickettsial disease in the USA[17] and Tennessee is one of the top five states for RMSF transmission, accounting for approximately 12% of cases nationally. As with LD, the number of RMSF cases may be under-reported due to vague and/or asymptomatic infections,[34] and despite frequent laboratory testing and reports of RMSF, the true incidence in Tennessee is unknown.[35] Indirect immunofluorescence assay serologic testing is used by the CDC and most state laboratories, although this test commonly produces false positive and false negative results[36] and therefore cannot always provide definitive proof of RMSF in the early symptomatic phase. Additionally, diagnostic levels of antibodies do not appear until a week or more after the onset of symptoms, thus making early detection difficult. Prospective active surveillance for RMSF in regions where the disease is hyperendemic suggests as many as 50% of all cases (including confirmed but unreported deaths due to RMSF) are missed by passive surveillance mechanisms.[37]

Although statistical testing of babesiosis was inconclusive due to the small sample size, MCO data indicated at least three cases of babesiosis were diagnosed in Tennessee during the 2000–2009 study period. This is of interest because babesiosis has only been recently reported in Tennessee. This finding was recently discussed at the 2010 International Conference on Emerging Infectious Diseases, where the authors suggested they had discovered the first diagnosed case in Tennessee in 2009.[38] Additionally in 2010, the TDH acknowledged this report of possible infection as accurate.[39] Claims data suggest that one of the three cases also occurred in 2009, so it is possible, although not definite, that the Mosites *et al* finding[38] and the MCO claims finding are the same. Mosites *et al*[38] are now attempting to identify animal reservoir hosts and tick vectors. Data from the MCO could aid in this effort, and suggest at least two other cases occurred prior to this discovery.

As previously mentioned, a limitation in this particular study is the known tendency of some physicians to over-diagnose LD[31 32] particularly in non-endemic areas. While this could be an anomaly peculiar to LD, it also emphasizes the need for coordinated information sharing between health plans and the state health department. Additionally, if health plans contributed to the current reporting system using medical claims filings, the state could potentially investigate these claims for confirmation. These investigations could alleviate some of the concerns surrounding over- and mis-diagnosis of certain zoonotic diseases. Other limitations of this study include the inability to empirically filter out claims in error. Even though the data were filtered to include only cases with at least three line items, claims coding errors are still possible. However, administrative data have proven to be very reliable as a source of disease identification when compared to medical chart reviews.[40 41] We are further limited by the inability to relate, via a patient identifier, a medically diagnosed case based on claims data to a CDC defined 'confirmed' or 'probable' case. A patient could be coded with LD in the MCO claim system without necessarily having a laboratory confirmed diagnosis, or a physician could report a confirmed case without laboratory confirmation if the patient presented with erythema migrans and was recently in an endemic county.[42] Recent studies examining the differences between administrative data and notifiable disease data conclude administrative data could enhance the current passive surveillance registry system.[43 44] Lastly, although our final model included only a single random effect adjustment (county), we used a GLMM approach rather than a non-linear mixed model in order to provide more flexibility in adjusting for multiple random effects, including nested and crossed random effects during the development phase.

CONCLUSIONS

Data sharing to improve disease surveillance is not without challenges. However, significant opportunity may exist in Tennessee to supplement the current passive reporting system with administrative medical claims data, particularly for LD and RMSF and other rare infectious diseases. There are known limitations with using administrative claims data, but health plans may help bridge data gaps as well as support the federal

administration's vision of combining public and private data into one source. Additionally, the benefit to the public health system should take precedence over any potential data discrepancies at the individual level. State and local public health officials rely on healthcare providers, laboratories, and other public health personnel to report the occurrence of notifiable diseases to state and local health departments.[45] Missing from this system are health plans and the data they could provide to state and national surveillance efforts. Without such data, trends may not be accurately monitored, unusual occurrences of diseases might not be detected, and the effectiveness of intervention activities cannot be easily evaluated.

REFERENCES

1. **Government Accountability Office (GAO).** *Emerging Infectious Diseases Review of State and Federal Disease Surveillance Efforts: Report to the Chairman, Permanent Subcommittee on Investigations, Committee on Governmental Affairs, US Senate.* United States Government Accountability Office, 2004. GAO report# GAO-04−877. http://www.gao.gov/new.items/d04877.pdf
2. **Centers for Disease Control and Prevention (CDC NNDSS).** *National Notifiable Diseases Surveillance System.* 2010. http://www.cdc.gov/ncphi/disss/nndss/nndsshis.htm (accessed 21 Sep 2010).
3. **Koo D,** Wetterhall SF. History and current status of the National Notifiable Diseases System. *J Public Health Manag Pract* 1996;**2**:4−10.
4. **Tennessee Department of Health (TDH CEDS).** *Tennessee Department of Health, Communicable and Environmental Disease Services: Reportable Diseases and Events.* 2010. http://health.state.tn.us/ceds/notifiable.htm (accessed 5 Oct 2010).
5. **Doyle T,** Glynn M, Groseclose S. Completeness of notifiable infectious disease reporting in the United States: an analytical literature review. *Am J Epidemiol* 2002;**155**:866−74.
6. **Marier R.** The reporting of communicable diseases. *Am J Epidemiol* 1977;**105**:587−90.
7. **Meek J,** Roberts C, Smith E Jr, *et al.* Underreporting of Lyme disease by Connecticut physicians. *J Public Health Manag Pract* 1996;**2**:61−5.
8. **Koo D,** Caldwell B. The role of providers and health plans in infectious disease surveillance. *Eff Clin Pract* 1999;**2**:247−52.
9. **Figueiras A,** Lado E, Fernández S, *et al.* Influence of physicians' attitude on under-notifying infectious diseases: a longitudinal study. *Public Health* 2004;**118**:521−6.
10. **Young J.** Underreporting of Lyme disease. *N Engl J Med* 1998;**338**:1629.
11. **Campos-Outcalt D,** England R, Porter B. Reporting of communicable diseases by university physicians. *Public Health Rep* 1991;**106**:579−83.
12. **Office of the National Coordinator for Health Information Technology (ONC).** *Federal Health Information Technology Strategic Plan 2011−2015.* Washington, DC: Office of the National Coordinator for Health Information Technology (ONC), 2011.
13. **California Department of Public Health.** *Planning for Health Information Technology and Exchange in Public Health. UC Davis Health Informatics Seminar.* 2009. http://www.cdph.ca.gov/programs/HISP/Documents/UCD-InformaticsSeminar-HIE-CDPH-0809-LScott.pdf
14. **Rutherford GW.** Public health, communicable diseases, and managed care: will managed care improve or weaken communicable disease control? *Am J Prev Med* 1998;**14**(3 Suppl):53−9.
15. *Health Insurance Portability and Accountability Act (HIPAA) of 1996.* U.S. Department of Health & Human Services, 1996. Pub. L. No. 104-191, 110 Stat. 1936. http://aspe.hhs.gov/admnsimp/pl104191.htm
16. **DeNavas-Walt C,** Proctor B, Smith J. *US Census Bureau, Current population Reports, P60−238, Income, Poverty, and Health Insurance Coverage in the United States: 2009.* Washington, DC: US Government Printing Office, 2010.
17. **Centers for Disease Control and Prevention (CDC NCEZID).** *National Center for Emerging and Zoonotic Infectious Diseases.* 2010. http://www.cdc.gov/ncezid/ (accessed 21 Sep 2010).
18. **Tennessee Department of Health (TDH WebAim).** *Tennessee Department of Health, Communicable and Environmental Disease Services: The Communicable Disease Interactive Data Site.* 2010. http://health.state.tn.us/Ceds/WebAim/WEBAim_criteria.aspx
19. **Salah A,** Kamarianakis Y, Chlif S, *et al.* Zoonotic cutaneous leishmaniasis in central Tunisia: spatio-temporal dynamics. *Int J Epidemiol* 2007;**36**:991−1000.
20. **SAS®.** *Statistical Analysis Software.* Cary, NC, USA: Copyright © 2006-2008 by SAS Institute Inc, 2008.
21. **Kalluri S,** Gilruth P, Rogers D, *et al.* Surveillance of arthropod vector-borne infectious diseases using remote sensing techniques: a review. *PLoS Pathog* 2007;**3**:1361−71.
22. **Wimberly MC,** Baer AD, Yabsley MJ. Enhanced spatial models for predicting the geographic distributions of tick-borne pathogens. *Int J Health Geogr* 2008;**7**:15.
23. **Winters AM,** Eisen RJ, Lozano-Fuentes S, *et al.* Predictive spatial models for risk of West Nile virus exposure in eastern and western Colorado. *Am J Trop Med Hyg* 2008;**79**:581−90.
24. **Yang K,** Zhou XN, Wu XH, *et al.* Landscape pattern analysis and Bayesian modeling for predicting *Oncomelania hupensis* distribution in Eryuan County, People's Republic of China. *Am J Trop Med Hyg* 2009;**81**:416−23.
25. **Yiannakoulias N,** Schopflocher D, Svenson L. Using administrative data to understand the geography of case ascertainment. *Chronic Dis Can* 2009;**30**:20−8.
26. **Centers for Disease Control and Prevention (CDC).** *Vaccine Safety Datalink (VSD) Project.* http://www.cdc.gov/vaccinesafety/Activities/VSD.html
27. **Marangoni A,** Sparacino M, Cavrini F, *et al.* Comparative evaluation of three different ELISA methods for the diagnosis of early culture-confirmed Lyme disease in Italy. *J Med Microbiol* 2005;**54**:361−7.
28. **Coyle BS,** Strickland GT, Liang YY, *et al.* The public health impact of Lyme disease in Maryland. *J Infect Dis* 1996;**173**:1260−2.
29. **Bacon RM,** Kugeler KJ, Mead PS; Centers for Disease Control and Prevention (CDC). Surveillance for Lyme disease — United States, 1992−2006. *MMWR Surveill Summ* 2008;**57**:1−9.
30. **Rosen M.** *Investigating the maintenance of the Lyme disease Pathogen, Borrelia burgdorferi, and Its Vector, Ixodes scapularis, in Tennessee.* Master's Thesis, University of Tennessee, 2009. http://trace.tennessee.edu/utk_gradthes/554
31. **Steere AC,** Taylor E, McHugh GL, *et al.* The overdiagnosis of Lyme disease. *JAMA* 1993;**269**:1812−16.
32. **Svenungsson B,** Lindh G. Lyme borreliosis—an overdiagnosed disease? *Infection* 1997;**25**:140−3.
33. **Moncayo A.** Vector-borne diseases in Tennessee. *Tennessee Department of Health & Vanderbilt University Medical Center. Twenty Sixth Biennial State Public Health Vector Control Conference.* 2006. http://www.cdc.gov/ncidod/dvbid/westnile/conf/26thbiennialVectorControl/pdf/state/Tennessee.pdf
34. **Lacz N,** Schwartz R, Kapila R. *Rocky Mountain spotted Fever. MedScape eMedicine.* http://emedicine.medscape.com/article/1054826-overview (accessed 1 Dec 2010).
35. **Moncayo A,** Cohen S, Fritzen C, *et al.* Absence of Rickettsia rickettsii and occurrence of other spotted fever group rickettsiae in ticks from Tennessee. *Am J Trop Med Hyg* 2010;**83**:653−7.
36. **Georgia Epidemiology Report (GER).** *Common Tickborne Diseases in Georgia: Rocky Mountain Spotted Fever and Human Monocytic Ehrlichiosis.* Georgia State Health Dept, Division of Public Health, 2004:20:5. http://health.state.ga.us/pdfs/epi/gers/ger0504.pdf
37. **Wilfert C,** MacCormack J, Kleeman K, *et al.* Epidemiology of Rocky Mountain spotted fever as determined by active surveillance. *J Infect Dis* 1984;**150**:469−79.
38. **Mosites E,** Wheeler P, Pieniazek N, *et al.* Human babesiosis caused by a novel babesia parasite—Tennessee, 2009. International Conference on Emerging Infectious Diseases: *Unpublished Abstract Presented at the International Conference on Emerging Infectious Disease (ICEID) July 2010 Atlanta, GA.* 2010. http://www.iceid.org/images/stories/2010program.pdf
39. **Moncayo A.** *Human Babesiosis. 2010 One Medicine Symposium; Reality Bites: A One Medicine Approach to Vector-Borne Diseases.* 2010. http://www.ncagr.gov/oep/oneMedicine/noms/2010/Moncayo_Abelardo_Human_Babesiosis.pdf
40. **Hux JE,** Ivis F, Flintoft V, *et al.* Diabetes in Ontario: determination of prevalence and incidence using a validated administrative data algorithm. *Diabetes Care* 2002;**25**:512−16.
41. **Tu K,** Campbell N, Chen X, *et al.* Accuracy of administrative databases in identifying patients with hypertension. *Open Med* 2007. http://www.openmedicine.ca/article/view/17/23
42. **Centers for Disease Control and Prevention (CDC).** Notice to readers: recommendations for test performance and Interpretation from the Second National Conference on Serologic Diagnosis of Lyme disease. *MMWR Morb Mortal Wkly Rep* 1995;**44**:590−1.
43. **Yiannakoulias L,** Svenson LW. Differences between notifiable and administrative health information in the spatial—temporal surveillance of enteric infections. *Int J Med Inform* 2009;**78**:417−24.
44. **Jones S,** Conner W, Song B, *et al.* Comparing spatio-temporal clusters of arthropod-borne infections using administrative medical claims and state reported surveillance data. *Spat Spatiotemporal Epidemiol* 2012;**3**:205−13.
45. **Centers for Disease Control and Prevention (CDC).** Case definitions for infectious conditions under public health surveillance: recommendations and reports. *MMWR Recomm Rep* 1997;**46**:1−55.