Ticks and Tick-borne Diseases 10 (2019) 241–250

Contents lists available at ScienceDirect

Ticks and Tick-borne Diseases

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

Epidemiology of Lyme disease in Pennsylvania 2006–2014 using electronic health records

Katherine A. Moon[a,*], Jonathan Pollak[a], Annemarie G. Hirsch[a,b], John N. Aucott[c], Cara Nordberg[d], Christopher D. Heaney[a,e,f], Brian S. Schwartz[a,b,g]

[a] Department of Environmental Health and Engineering, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA
[b] Department of Epidemiology and Health Services Research, Geisinger, Danville, PA, USA
[c] Johns Hopkins School of Medicine, Department of Medicine, Division of Rheumatology, Baltimore, MD, USA
[d] Department of Biomedical and Translational Informatics, Geisinger, Danville, PA, USA
[e] Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA
[f] Department of International Health, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA
[g] Johns Hopkins School of Medicine, Department of Medicine, Baltimore, MD, USA

ARTICLE INFO

Keywords:
Lyme disease
Lyme borreliosis
Post-treatment Lyme disease syndrome
Epidemiology
Surveillance
Electronic health record

ABSTRACT

Lyme disease is the most common vector-borne disease in the United States. Electronic health record (EHR)-based research on Lyme disease is limited. We used Geisinger EHR data from 479,344 primary care patients in 38 Pennsylvania counties in 2006–2014 to compare EHR-based Lyme disease incidence rates to surveillance incidence rates, evaluate individual and community risk factors for incident Lyme disease, and to characterize the proportion of cases with diagnoses consistent with post-treatment Lyme disease syndrome in the EHR ($PTLDS_{EHR}$). We primarily identified Lyme disease cases using diagnosis codes, serologic testing order codes, and medication orders but also completed subgroup analyses among those with positive serology and those with both diagnosis code and antibiotic treatment. We compared annual incidence rates from the EHR to surveillance by age, sex, and county. In case-control analyses, we compared cases to randomly selected controls (5:1) frequency-matched on year, age, and sex. We identified 9657 cases of Lyme disease, including 1791 cases with positive serology and 4992 cases with both diagnosis code and antibiotic treatment. Annual incidence rates in the EHR were 4.25–7.43 times higher than surveillance. In adjusted analyses, white non-Hispanic race/ethnicity (vs. black, Hispanic, or other) was associated with higher odds of Lyme disease (odds ratio [OR]: 2.06, 95% confidence interval [CI]: 1.73–2.44). Medical Assistance insurance use (always vs. never; OR: 0.77, 95% CI: 0.68–0.88), and higher community-level socioeconomic deprivation (quartile 4 vs. 1 OR: 0.50 (95% CI: 0.42–0.59) were associated with lower odds of Lyme disease. Within 4–52 weeks after Lyme disease diagnosis, 20.8% (n = 735) of cases with a diagnosis code and treatment had a diagnosis of malaise or fatigue, pain, or cognitive difficulties not present in the past 26 weeks. These results highlight the utility of EHR data for epidemiologic research on Lyme disease for case-finding, surveillance, risk factor evaluation, and characterization of PTLDS using EHR data.

1. Introduction

Lyme disease, caused by *Borrrelia burgdorferi* and transmitted by the blacklegged tick, is the most common vector-borne disease in the United States (US) (Adams et al., 2016). For persons who live in or have recently travelled to an endemic area, diagnosis can be made based on erythema migrans, the hallmark rash of early Lyme disease (Wormser et al., 2006). Although uncomplicated Lyme disease generally resolves if treated promptly, Lyme disease can progress within weeks to months to more serious complications, including arthritis, meningitis, neuropathy, and carditis (Wormser et al., 2006). Up to 20% of cases in treatment trials experience persistent subjective symptoms after antibiotic treatment, called post-treatment Lyme disease syndrome (PTLDS) (Marques, 2008).

In 2015, there were 38,069 surveillance cases of Lyme disease in the US (Adams et al., 2016). However, the US annual incidence rate was recently estimated to be ten-fold higher than reported by surveillance (Nelson et al., 2015). Surveillance for Lyme disease is resource-intensive for local health departments and is limited by low rates of

* Corresponding author at: Dept. of Environmental Health & Engineering, 615N Wolfe St., W7604, Baltimore, MD, USA.
E-mail addresses: kmoon9@jhu.edu (K.A. Moon), jpollak2@jhu.edu (J. Pollak), aghirsch@geisinger.edu (A.G. Hirsch), jaucott2@jhmi.edu (J.N. Aucott), cmnordberg@geisinger.edu (C. Nordberg), cheaney1@jhu.edu (C.D. Heaney), bschwar1@jhu.edu (B.S. Schwartz).

https://doi.org/10.1016/j.ttbdis.2018.10.010
Received 27 March 2018; Received in revised form 28 September 2018; Accepted 24 October 2018
Available online 26 October 2018
1877-959X/

Table 1
Lyme disease case definitions.

Case group	Case definition	Number
All cases (high sensitivity)	ICD-9 diagnosis code **OR** (serology CPT test order code, *excluding orders during emergency or inpatient encounters with negative test results*[a] **&** Antibiotic order within ± 30 days of sample draw, *excluding orders with respiratory disease diagnoses*[b]*)*	9657
Subgroups of all cases defined above:		
Positive test cases (high specificity for later stage disease)	Positive IgG Western blot within ± 90 days of incident date[d] **OR** Positive EIA **&** positive IgM Western blot[c] within ± 30 days of incident date[d]	1791
Diagnosis code & treatment cases (high specificity)	ICD-9 diagnosis code on or within 90 days after incident date[d] **&** Antibiotic order within ± 30 days of incident date[d]	4992

Abbreviations: ICD, International Classification of Diseases; CPT, Current Procedural Terminology; EIA, enzyme immunoassay; IgG, immunoglobulin G; IgM, immunoglobulin M.

[a] To improve the specificity of our primary case definition in the absence of a Lyme ICD code, we excluded Lyme serology CPT codes from eligibility in the emergency or inpatient setting, where Lyme disease serology tests likely had a low pre-test probability of disease.

[b] To improve the specificity of our primary case definition in the absence of a Lyme ICD code, we excluded antibiotic orders if diagnosis codes linked to the medication orders were respiratory disease (ICD-9: 460–519) or otitis media (ICD-9: 381.x or 382.x), since these are common diagnoses treated with the same antibiotics as Lyme disease.

[c] The CDC two-tiered testing protocol considers IgG Western blot positive test results only when preceded by an EIA positive test. The vast majority of positive IgG Western blots (> 89%) were accompanied by a positive EIA, but we also included positive IgG Western blots without EIA (10%) or with a negative EIA (< 1%) because we thought it was possible that the initial positive EIA was obtained and recorded outside of Geisinger.

[d] The Lyme disease incident date was the first date of either a: 1) Lyme disease ICD-9 diagnosis code or 2) CPT code for Lyme disease serology.

reporting by health care providers (Schiffman et al., 2018; White et al., 2018). Electronic health record (EHR) systems could improve upon traditional Lyme disease surveillance by improving case-finding, reducing provider burden, and increasing timeliness of health system reporting (Birkhead et al., 2015; Mac Kenzie et al., 2016). Although insurance claims and EHR data both contain rich longitudinal clinical data from large populations, including clinical diagnoses, medication orders, and test orders, EHRs also have laboratory test results, more detailed data on individual-level comorbidities, and clinical notes (Casey et al., 2016).

Previous EHR applications to Lyme disease epidemiology have been limited (Eliassen et al., 2017a; Ichikawa et al., 2017; Lantos et al., 2015). We aimed to demonstrate the utility of EHR data to advance Lyme disease epidemiology with three objectives: 1) to compare Lyme disease incidence rates identified in the EHR to those reported via routine surveillance; 2) to evaluate individual and community risk factors for incident EHR-based Lyme disease; and 3) to characterize Lyme disease cases with diagnoses consistent with PTLDS in the EHR ($PTLDS_{EHR}$).

2. Methods

2.1. Study population

Geisingeris an integrated health system that provides primary care services at community practice clinics and hospitals in central and northeastern Pennsylvania. For this study, we used retrospective EHR data on 479,344 individuals with a Geisinger primary care provider, contact with the EHR between January 1, 2006 and December 31, 2014, and a geocoded address within the 38-county study area. The Institutional Review Board at the Geisinger Health System approved this study.

2.2. Lyme disease case definitions

We defined incident Lyme disease cases based on *International Classification of Diseases* (ICD, 9th Revision) diagnostic codes linked to encounter and medication orders, *Current Procedural Terminology* (CPT) codes for Lyme disease serology orders, antibiotic orders, and Lyme disease serological test results. For laboratory diagnosis of Lyme disease, the Centers for Disease Control (CDC) and the Infectious Disease Society of America (IDSA) recommend an initial enzyme immunoassay (EIA) or immunofluorescence assay, and if positive or equivocal, then a confirmatory Western blot for immunoglobulin M or G (IgM and IgG) (Centers for Disease Control & Prevention (CDC), 1995). An IgG Western blot is considered confirmatory regardless of sample draw timing, whereas an IgM Western blot is only recommended if signs and symptoms have been present for less than a month (Seriburi et al., 2012).

We used three case definitions for Lyme disease in this study: a highly sensitive primary case definition and two subgroups with higher specificity (Table 1). For our primary case definition, we defined a Lyme disease case as either: 1) a Lyme disease ICD code (ICD-9 088.81) linked to an encounter or medication order; or 2) both a CPT code for a Lyme disease serologic test (EIA or Western blot) and an antibiotic order appropriate for Lyme disease within 30 days before or after the sample draw. We used two exclusion criteria to reduce false positives among individuals without a Lyme disease diagnosis. Regarding testing, it was apparent to us that in the emergency or inpatient settings, Lyme disease serologic testing was obtained in patients with a broad range of symptoms, many of which would define a low pre-test probability of disease. We thus excluded CPT codes from eligibility if the linked test result was negative and the test was ordered in an inpatient or emergency encounter. Regarding treatment, Lyme disease-appropriate antibiotics are used for respiratory diseases (ICD-9: 460–519) or otitis media (ICD-9: 381.x or 382.x), which are also very common. We assigned the Lyme disease incident date as the first date of either a Lyme disease ICD or CPT code. Appropriate treatment was defined by IDSA-recommended first or second line antibiotics (Wormser et al., 2006) and three antibiotics either closely related to recommended treatments or that were historical treatments (Nelson et al., 2015).

For the first higher specificity subgroup, positive test cases, we hypothesized that Lyme disease cases with positive test results would have a higher positive predictive value for later stage Lyme disease (Table 1). For this analysis, we defined a positive Lyme disease serological test as either an IgG positive test (either alone, with positive EIA, or with negative EIA), or an IgM positive test with a positive EIA. The vast majority of cases with positive IgG Western blots (> 89%)

were accompanied by a positive EIA, but we also included positive IgG Western blots without EIA (10%) or with a negative EIA (< 1%) because we thought it was possible that the initial positive EIA was obtained outside of Geisinger. In the second higher specificity subgroup, diagnosis code and treatment cases, we hypothesized that cases with both a diagnosis code and antibiotic treatment would have a higher positive predictive value than our primary case definition and would include both early and later stage cases (Table 1). For this subgroup, we included only Lyme disease cases with both a Lyme disease ICD-9 code, linked to an encounter or medication order on their incident date or within the next 90 days and a Lyme disease-appropriate antibiotic order within ± 30 days of their incident date.

2.3. Lyme disease controls

Individuals were eligible for control selection up to the calendar year prior to evidence of Lyme disease, defined as any of the following: selection as a case; diagnosis code for Lyme disease or tick bite, Lyme disease serology CPT code; or diagnosis code for a non-infected insect bite where there was also a Lyme disease-appropriate antibiotic order within ± 30 days. We randomly selected without replacement five control outpatient or emergency encounter dates for each case and frequency matched to cases by age (< 1, 1– < 5, 5– < 15, 15– < 25, 25– < 40, 40– < 65, ≥65 years), sex, and encounter year. For analyses of the two Lyme disease case subgroups, subgroup cases were compared to a random sample of all control encounters after repeat frequency-matching.

2.4. $PTLDS_{EHR}$

Clinically, PTLDS is largely a diagnosis of exclusion, requiring documentation of prior Lyme disease, appropriate treatment, and onset of unexplained, subjective symptoms within six months after a Lyme disease diagnosis that persist for at least six months after completion of antibiotic treatment (Wormser et al., 2006). Fatigue, pain, and memory or cognitive difficulties are the most common patient-reported symptoms associated with PTLDS (Aucott, 2015). To identify the frequency of Lyme disease cases with diagnoses consistent with PTLDS in the EHR ($PTLDS_{EHR}$), we examined diagnosis codes linked to encounters and medication orders within three categories (malaise or fatigue, pain, and cognitive difficulties) occurring in the four to 52 weeks after a Lyme disease incident date. For each of the three categories, we considered a diagnosis to be new if there were no diagnoses in that category in the 26 to four weeks prior to a Lyme disease incident date (Online Appendix Fig. A.1). Diagnoses during the first four weeks were excluded to allow for the two to four weeks of recommended antibiotic treatment (Wormser et al. 2006). Diagnoses in the four weeks before the Lyme disease incident date were excluded to avoid capturing diagnoses related to infection before treatment. In controls, diagnoses that met the criteria for $PTLDS_{EHR}$ were ascertained using the same criteria around their control encounter. For this analysis, we restricted to individuals with at least 26 weeks of observation in the EHR before their case or control selection date and at least 52 weeks of observation after, and without prior evidence of chronic fatigue, fibromyalgia, or chronic pain based on diagnosis codes linked to encounters, medication orders, or problem lists.

2.5. Lyme disease surveillance data

Pennsylvania surveillance data for 2006–2014 were obtained from the Pennsylvania Department of Health (PADOH), including reported Lyme disease cases and population at risk from the 2010 US Census by county, age, and sex (Pennsylvania Department of Health, 2017). Cases represent the Lyme disease surveillance case definition in use at the time. Beginning in 2008, surveillance cases include probable cases in addition to confirmed cases, and a single-tier IgG was considered positive laboratory evidence. In 2011, the case definition clarified that a two-tier IgM positive test result should only be interpreted as positive within the first 30 days of illness onset.

2.6. Individual-level covariates

Individual-level covariates were obtained from the EHR. All EHR variables were time-varying based on the case or control encounter date except for sex and race/ethnicity. We considered use of a Medical Assistance insurance payer (e.g., Medicaid) for an outpatient encounter as a proxy for low family socioeconomic status (Casey et al., 2017). We estimated a person's overall comorbidity burden using the Charlson index based on ICD-9 diagnostic codes linked to encounters, medication orders, and problem lists (Quan et al., 2005). We used the number of outpatient encounters within the past year as a measure of overall health care utilization.

We used ICD-9 diagnostic codes linked to encounter and medication orders to assess whether an individual was diagnosed with malaise or fatigue within the past year, myalgia or arthralgia within the past year, or allergy or atopy ever in the past. We excluded the 30 days prior to the case incident date or control encounter to avoid diagnoses related to active Lyme disease infection. We examined recent history of malaise or fatigue and myalgia or arthralgia because these are non-specific and subjective symptoms that overlap with those of early Lyme disease (Tibbles and Edlow, 2007), and we hypothesized that an association between either of these nonspecific diagnoses with Lyme disease among all cases but not among the higher specificity subgroups could provide evidence of false positives in the primary case definition. For allergy or atopy, we hypothesized that persons with these conditions may be more likely to develop pruritus at the tick bite site (Schwartz et al., 1991), which might allow early recognition and removal of a tick, and thus may lower the risk of Lyme disease.

2.7. Community socioeconomic deprivation

The study area included a range of community settings from very low density rural areas (townships) to walkable small towns (boroughs), automobile-dependent suburban subdivisions, and urban areas in Pennsylvania. We geocoded each individual's home address to assign them to one of three community types based on US Census minor civil divisions and census tracts (townships [n = 585], boroughs [n = 277], and census tracts within cities [n = 128]). We believe that this definition of community captures the cultural and behavioral space significant to residents in our study region (Liu et al., 2013) and is consistent with our prior work (Casey et al., 2013).

We used an index of community-level socioeconomic deprivation modified from the Townsend index to reflect the socioeconomic factors relevant to our study region (Liu et al., 2013). For each community, we z-transformed and summed the values of six indicators (proportion of the population with less than a high school education, unemployed, not in labor force, in poverty, receiving public assistance, and proportion of households without a car) derived from American Community Survey five-year estimates in 2006–2010 and 2010–2014 (U.S. Census Bureau, 2015, 2010). Encounters prior to 2011 were assigned values from 2006-2010, and values from 2010-2014 were used thereafter.

2.8. Statistical analysis

We calculated Geisinger EHR Lyme disease annual incidence rates for comparison to PADOH surveillance incidence rates. Each person under observation in a year contributed one person-year to the denominator of the incidence rate. Surveillance incidence rates were calculated using Lyme disease cases and population at risk restricted to our 38-county study area. We calculated the incident rate ratio of Geisinger EHR to PADOH surveillance incidence rates by year, county of residence, age, and sex.

In the case-control analysis, we estimated the association between each risk factor and incident Lyme disease in three multivariable logistic regression models, one for each Lyme disease case definition. We used robust standard errors clustered by the individual's community (township, borough, urban census tract). Final models adjusted for age (in the previously-described matching categories), sex, year, and all individual-level and community-level risk factors. We dichotomized race/ethnicity because fewer than 10% of persons were non-white or Hispanic. We initially evaluated associations with Medical Assistance considering a binary indicator (ever vs. never user) and categories of proportion of time under observation in the EHR that an individual used a Medical Assistance payer (e.g., 0, 0.1–33.3%, 33.4%–66.7%, 66.8%–99.9%, 100%), as previously described (Casey et al., 2017). In final models, we present associations with Medical Assistance as proportion of time on Medical Assistance after collapsing five categories into three (always, sometimes, or never user) based on the strength of the association and model Akaike Information Criteria.

We estimated the frequency of $PTLDS_{EHR}$ in Lyme disease cases and controls and compared Lyme disease cases with and without $PTLDS_{EHR}$ diagnoses by all individual and community risk factors using ANOVA or Pearson's chi-square tests. Analysis was conducted using Stata version 14 (StataCorp., College Station, TX, USA).

3. Results

3.1. Lyme disease cases

We identified a total of 9657 cases of Lyme disease in the Geisinger EHR 38-county study area between January 2006 and December 2014 (Fig. 1, Table 1). Of these cases, 1791 (18.5%) cases had a positive serological test, and 4992 (51.7%) cases had both a Lyme disease diagnosis code and an order for an appropriate antibiotic.

3.2. Comparison of EHR to PADOH surveillance incidence rates

EHR-derived annual incidence rates were between 4.25–7.43 times higher than the PADOH incidence rates for all study years (Online Appendix Table A.1). Incidence rates were higher in men in both data sources (Online Appendix Table A.2). PADOH rates were highest in the youngest (< 15 y) and oldest (≥55 y) ages, whereas EHR rates were highest in older persons (35– < 55 and ≥55 y), and lowest in younger persons (< 15 and 15– < 35 y) (Online Appendix Table A.3). The age distribution of Lyme disease incidence rates was more similar to PADOH rates when incidence rates were restricted to positive test cases and diagnosis code and treatment cases (Online Appendix Fig. A.2). Incidence rates varied considerably by county: incidence rate ratios varied between 2.69- and 23.23-fold higher in the EHR (Fig. 2, Online Appendix Table A.4).

3.3. Case-control analysis

In unadjusted comparisons to controls, Lyme disease cases were more likely to be of white non-Hispanic race/ethnicity, more likely to live in townships, less likely to use Medical Assistance insurance, and less likely to live in communities with high socioeconomic deprivation (Table 2 for all cases and Online Appendix Table A.5 for two case subgroups).

In adjusted analyses of the primary case definition, white non-Hispanic race/ethnicity was associated with higher odds of Lyme disease, while higher community socioeconomic deprivation and use of Medical Assistance were associated with lower odds of Lyme disease (Table 3). Compared with winter, summer was associated with the

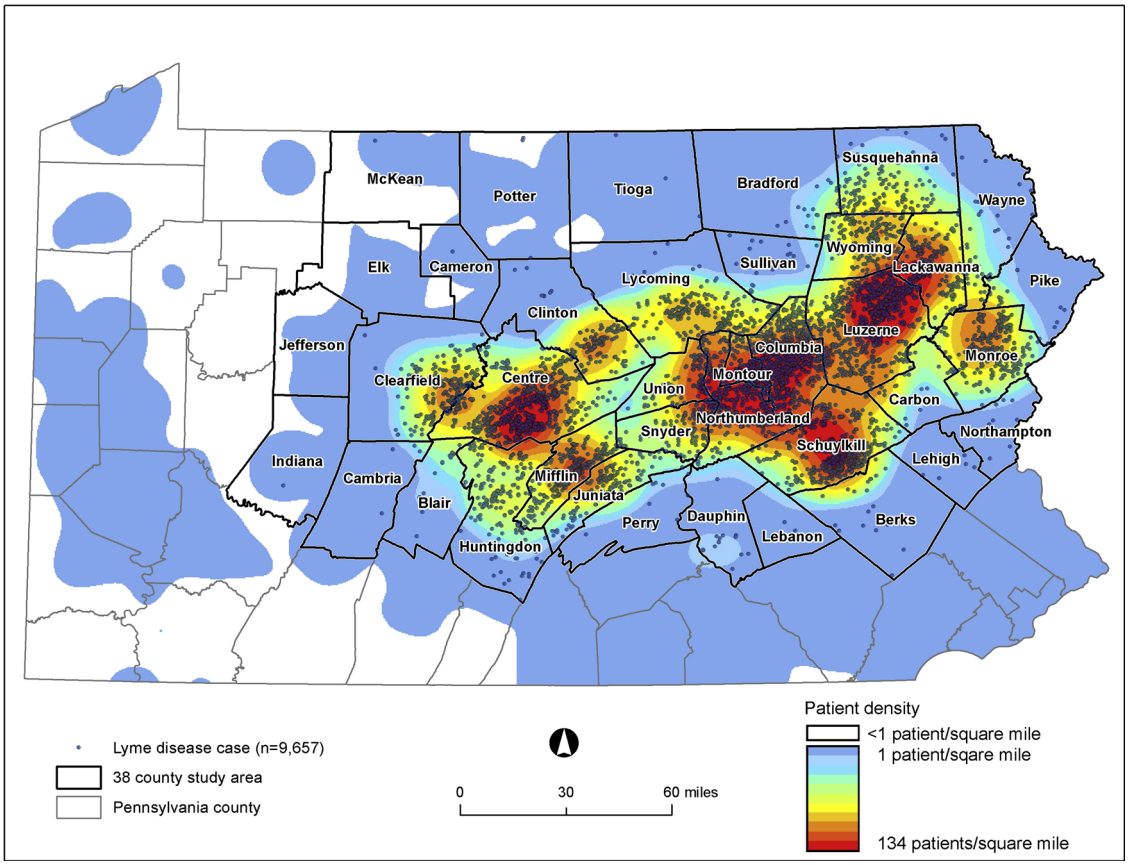

Fig. 1. Spatial distribution of Geisinger primary care patient EHR Lyme disease cases 2006–2014 and Geisinger patient density. Abbreviations: EHR, electronic health record. PADOH, Pennsylvania Department of Health. PA, Pennsylvania.

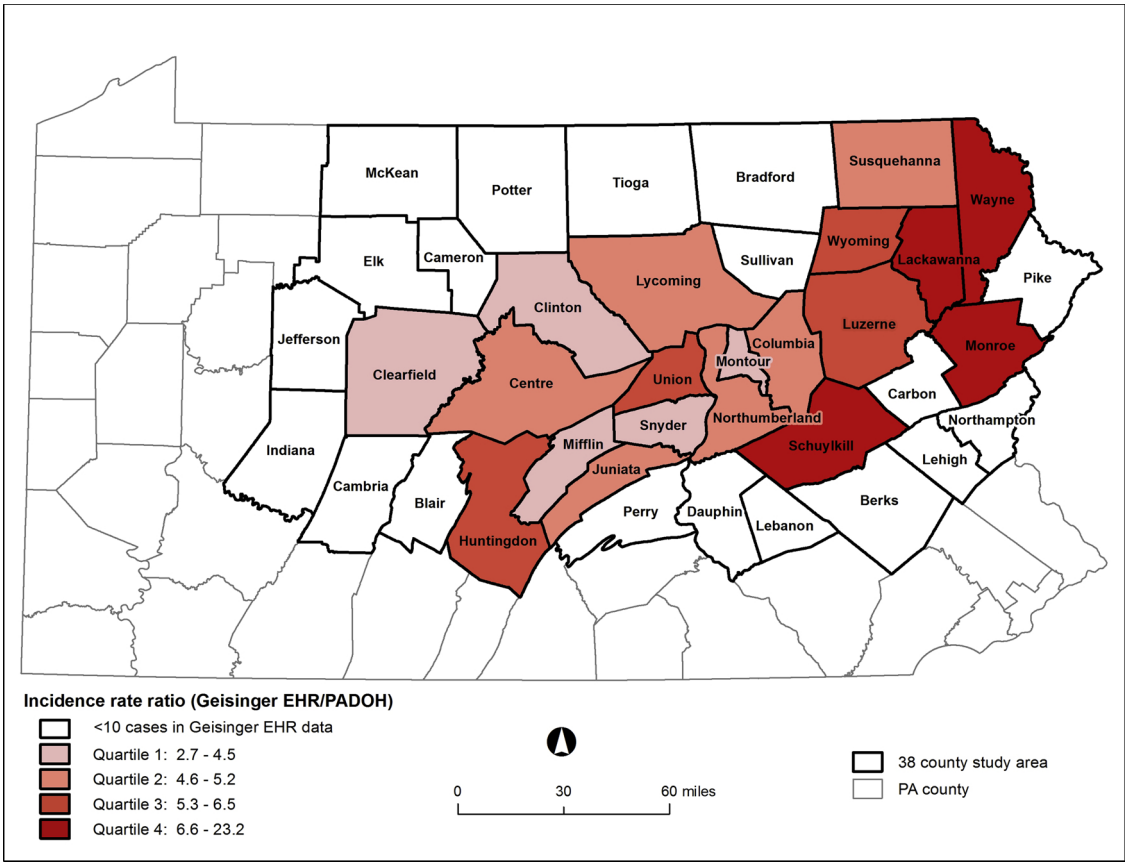

Fig. 2. Ratio of Geisinger primary care patient EHR incidence rates to surveillance incidence rates. Abbreviations: EHR, electronic health record. PADOH, Pennsylvania Department of Health. PA, Pennsylvania. Ratio of Lyme disease incidence rates in the Geisinger EHR to Pennsylvania Department of Health in 2014 by county.

highest odds of Lyme disease diagnosis, followed by fall and spring. Higher Charlson comorbidity index scores were associated with lower odds of Lyme disease. Outpatient visits in the past year were associated with higher odds of Lyme disease. History of malaise or fatigue within the past year, myalgia or arthralgia within the past year, and ever allergy or atopy were associated with higher odds of Lyme disease.

Among positive test cases, associations were consistent with the primary case definition with a few exceptions (Table 3). The association with Medical Assistance was slightly attenuated. Allergy or atopy was no longer associated with Lyme disease, and the association with summer strengthened, while the association with spring was no longer present. Among diagnosis code and treatment cases, associations were consistent with the primary case definition except that associations with spring, summer, and fall were stronger.

3.4. Frequency of $PTLDS_{EHR}$ diagnoses

The proportion of Lyme disease cases with $PTLDS_{EHR}$ was 22.8% (n = 1556) in all cases, 21.6% (n = 253) in positive test cases, and 20.8% (n = 735) in diagnosis code and treatment cases among individuals without a history of chronic pain, fibromyalgia, or chronic fatigue and with at least 26 weeks of EHR contact before and at least 52 weeks after Lyme disease incident dates (Appendix Table A.6). In comparison, the frequencies of diagnoses that met the criteria for $PTLDS_{EHR}$ in controls ranged from 13.2% to 14.6%. Among Lyme disease cases with $PTLDS_{EHR}$, new diagnoses of pain (77.4%) were the most common, followed by fatigue or malaise (32.9%) and memory or cognitive diagnoses (3.9%). Compared to cases without $PTLDS_{EHR}$, Lyme disease cases with $PTLDS_{EHR}$ were older, more likely to be female, had more comorbidities and prior outpatient visits, were more likely to have a diagnosis of malaise or fatigue in the past year, and were more likely to have a history of allergy or atopy (Table 4 for all cases and diagnosis code and treatment cases and Appendix Table A.7 for positive test cases).

4. Discussion

To our knowledge, this is the first study to use EHR data to estimate Lyme disease incidence rates for comparison to surveillance incidence rates, to evaluate individual and community risk factors for Lyme disease, and to characterize cases with and without $PTLDS_{EHR}$. First, in a comparison of Lyme disease incidence rates derived from the EHR and state-reported surveillance, we observed substantially higher rates of Lyme disease in the Geisinger primary care patient population and some differences by age and county. Second, in case-control analyses, race/ethnicity and measures of socioeconomic status were associated with Lyme disease. Comparing associations across three Lyme disease case definitions with varying sensitivity and specificity, we observed some differences in associations that could be explained by disease stage or misdiagnosis. Third, we identified a subgroup of Lyme disease cases with new diagnoses of fatigue, pain, and cognitive difficulties in the year following their Lyme disease diagnosis, and observed some differences in age, sex, and comorbidity burden. Taken together, these analyses demonstrate the utility of EHR data to improve surveillance and understanding of Lyme disease epidemiology.

Our estimate of the ratio of under-reporting compared to surveillance was generally consistent with the previous literature. Medical record review studies in endemic areas have identified underreporting ratios between three and twelve-fold (as reviewed in Schiffman et al. (2018)). A national insurance claims study using a sensitive case

definition, based on diagnosis codes and antibiotic orders, estimated an underreporting ratio of ten-fold (Nelson et al., 2015). In contrast, an analysis of insurance claims that applied a more specific case definition that would exclude early Lyme disease (ICD code, CPT code, and antibiotic order) and compared only to confirmed surveillance cases reported only about a two to three-fold higher incidence rate in high prevalence states (Tseng et al., 2015).

Table 2
Comparison of selected study variables among Lyme disease case encounters and year-, age-, and sex-matched control encounters in the Geisinger primary care patient EHR 2006–2014.

Characteristic	All cases (n = 9657)	Controls (n = 48,285)
Year, n (%)		
2006	323 (3.3%)	1615 (3.3%)
2007	369 (3.8%)	1845 (3.8%)
2008	627 (6.5%)	3135 (6.5%)
2009	905 (9.4%)	4525 (9.4%)
2010	898 (9.3%)	4490 (9.3%)
2011	1472 (15.2%)	7360 (15.2%)
2012	1264 (13.1%)	6320 (13.1%)
2013	1723 (17.8%)	8615 (17.8%)
2014	2076 (21.5%)	10,380 (21.5%)
Age, years		
mean (SD)	40.4 (22.3)	40.7 (22.9)
< 1, n (%)	9 (0.1%)	45 (0.1%)
1– < 5, n (%)	495 (5.1%)	2475 (5.1%)
5– < 15, n (%)	1447 (15.0%)	7235 (15.0%)
15– < 25, n (%)	855 (8.9%)	4275 (8.9%)
25– < 40, n (%)	1504 (15.6%)	7520 (15.6%)
40– < 65, n (%)	3941 (40.8%)	19,705 (40.8%)
≥65, n (%)	1406 (14.6%)	7030 (14.6%)
Sex, n (%)		
Female	4640 (48.0%)	23,200 (48.0%)
Male	5014 (51.9%)	25,080 (51.9%)
Unknown	3 (< 1%)	5 (< 1%)
Season, n (%)		
Winter (Dec–Feb)	996 (10.3%)	11,791 (24.4%)
Spring (March–May)	1658 (17.2%)	12,062 (25.0%)
Summer (June–August)	4583 (47.5%)	11,777 (24.4%)
Fall (Sept–Nov)	2420 (25.1%)	12,655 (26.2%)
Race/ethnicity, n (%)		
White non-Hispanic	9378 (97.1%)	45,000 (93.2%)
Hispanic	85 (0.9%)	1266 (2.6%)
Black non-Hispanic	139 (1.4%)	1511 (3.1%)
Other	46 (0.5%)	427 (0.9%)
Unknown	9 (0.1%)	81 (0.2%)
Medical Assistance insurance		
Never user	7752 (80.3%)	37,446 (77.6%)
Sometimes user	1518 (15.7%)	7850 (16.3%)
Always user	387 (4.0%)	2989 (6.2%)
History of selected conditions		
Malaise or fatigue in the past year, n (%) (a)	635 (6.6%)	1735 (3.6%)
Myalgia or arthralgia in the past year, n (%) (a)	1409 (14.6%)	4737 (9.8%)
Allergy or atopy ever in the past, n (%) (a)	4981 (51.6%)	22,138 (45.8%)
Charlson comorbidity index, n (%)		
0	3296 (34.1%)	16,687 (34.6%)
1–2	3382 (35.0%)	16,127 (33.4%)
3–4	1793 (18.6%)	8293 (17.2%)
≥5	1186 (12.3%)	7178 (14.9%)
Outpatient encounters in past year, mean (SD)	5.0 (4.5)	4.3 (4.1)
Community type, n (%)		
Township	7085 (73.4%)	28,915 (59.9%)
Borough	2124 (22.0%)	14,209 (29.4%)
City	448 (4.6%)	5161 (10.7%)
Community socioeconomic deprivation index		
Quartile 1, n (%)	2655 (27.5%)	10,144 (21.0%)
Quartile 2, n (%)	3024 (31.3%)	12,032 (24.9%)
Quartile 3, n (%)	2400 (24.9%)	13,051 (27.0%)
Quartile 4, n (%)	1578 (16.3%)	13,058 (27.0%)

Abbreviations: SD: standard deviation; EHR, electronic health record.
Notes:
(a) Past ICD-9 diagnosis codes in encounters and medication orders, excluding 30 days before case or control encounter date:
Fatigue or malaise: 780.71 (chronic fatigue syndrome), 780.79 (other malaise and fatigue).
Myalgia or arthralgia: 729.1 (myalgia and myositis, unspecified), 719.4x (pain in joint, arthralgia), 725 (polymyalgia rheumatica).
Allergy or atopy: V14.x (personal history of allergy to medicinal agents), V15.0x (personal history of allergy other than to medicinal agents, and other than to medicinal agents); 477.x (allergic rhinitis); 691.x (atopic dermatitis and related conditions); 692.xx (contact dermatitis and other eczema); 693.x (dermatitis due to substances taken internally); 995.27 (other drug allergy); 995.[367] (allergy unspecified, anaphylactic shock, other adverse food reactions); 708.x (urticaria).

The higher incidence rates in our EHR population compared to surveillance is likely due to the EHR capturing early Lyme disease cases diagnosed by erythema migrans that are substantially underreported by health care providers, and to a lesser extent, misclassification of laboratory-reported positive serological tests without clinical information. The EHR incidence rate may also be higher than the surveillance incidence rate due to overdiagnosis or because our estimate of the EHR population at risk could be smaller than the true population at risk. The EHR population at risk was based on individuals who sought care within a calendar year, which is likely biased toward women, children, and persons who are sicker on average (Klompas et al., 2017). Our estimates of the EHR incidence of Lyme disease may be underestimated in children, who have regular contact with the health system but who may be more likely to be diagnosed in urgent care or emergency settings outside of Geisinger, and over-estimated in older persons, who are more likely to be over-diagnosed with Lyme disease. Misclassification may be more likely in the overall case group, compared to the positive test cases or diagnosis code and treatment cases. Some of the annual variability between the Geisinger EHR and the PADOH incidence rates could be related to the changes in the surveillance case definition over time.

In the case-control analysis, our findings of higher odds of Lyme disease among white non-Hispanic individuals and those living in high socioeconomic status communities are consistent with previous literature (Bacon et al., 2008; Cromley et al., 1998; Fix et al., 2000; Jackson et al., 2006; Messier et al., 2015). We also observed a lower risk of Lyme disease in those who used Medical Assistance. To our knowledge, no previous study has evaluated the association between socioeconomic status at the individual or household level with Lyme disease. Low-density semi-rural and suburban communities close to wooded, brushy, or grassy landscapes favored by ticks are more likely to be white and affluent (Fix et al., 2000; Killilea et al., 2008) but differences by socioeconomic status and race in outdoor leisure activities, access to care, or awareness of Lyme disease may also play a role. There is some evidence, however, that African-Americans may be at higher risk than whites for delayed diagnosis and treatment of Lyme disease (Fix et al., 2000).

The observed seasonality differences across case definitions were expected. Diagnosis of later stage Lyme disease is more likely later in the season and higher specificity case definitions would be expected to

Table 3
Adjusted associations (odds ratio [95% CI]) of selected study variables with incident Lyme disease in the Geisinger primary care patient EHR 2006–2014.

Characteristic	All cases 9657 cases/48,285 controls	Positive test cases 1791 cases/8955 controls[a]	Diagnosis code & treatment cases 4992 cases/24,960 controls[a]
Race/ethnicity			
Black, Hispanic or other	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
White non-Hispanic	2.06 (1.73, 2.44)	2.28 (1.65, 3.15)	2.24 (1.84, 2.73)
Medical Assistance insurance			
Never user	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
Sometimes user	0.93 (0.86, 1.01)	0.84 (0.71, 1.00)	0.89 (0.80, 0.99)
Always user	0.77 (0.68, 0.88)	0.86 (0.65, 1.13)	0.72 (0.61, 0.85)
Community socioeconomic deprivation index			
Quartile 1 (Least deprived)	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
Quartile 2	0.99 (0.86, 1.13)	0.97 (0.80, 1.18)	0.95 (0.81, 1.12)
Quartile 3	0.74 (0.64, 0.84)	0.67 (0.54, 0.83)	0.65 (0.55, 0.76)
Quartile 4 (Most deprived)	0.50 (0.42, 0.59)	0.48 (0.36, 0.64)	0.44 (0.35, 0.54)
Season			
Winter (Dec–Feb)	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
Spring (March–May)	1.63 (1.49, 1.77)	0.97 (0.76, 1.23)	2.19 (1.92, 2.50)
Summer (June–August)	4.71 (4.31, 5.15)	9.55 (7.88, 11.57)	7.83 (6.91, 8.87)
Fall (Sept–Nov)	2.34 (2.16, 2.53)	2.84 (2.31, 3.50)	3.17 (2.79, 3.60)
Charlson comorbidity index			
0 (No comorbidities)	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
1–2	1.00 (0.92, 1.10)	0.87 (0.69, 1.09)	0.92 (0.82, 1.04)
3–4	0.88 (0.78, 0.99)	0.80 (0.60, 1.05)	0.88 (0.75, 1.03)
> 4 (Most comorbidities)	0.52 (0.45, 0.60)	0.34 (0.24, 0.46)	0.46 (0.38, 0.56)
Outpatient encounters in past year, per visit	1.04 (1.03, 1.05)	1.03 (1.01, 1.04)	1.03 (1.02, 1.04)
Malaise or fatigue in the past year			
No	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
Yes	1.67 (1.50, 1.85)	1.65 (1.27, 2.15)	1.38 (1.17, 1.63)
Myalgia or arthralgia in the past year			
No	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
Yes	1.40 (1.29, 1.52)	1.28 (1.06, 1.53)	1.35 (1.21, 1.50)
Allergy or atopy ever in the past			
No	1.00 (Reference)	1.00 (Reference)	1.00 (Reference)
Yes	1.15 (1.10, 1.22)	0.93 (0.82, 1.05)	1.14 (1.06, 1.23)

Abbreviations: EHR, electronic health record.
Controls were frequency matched by year, age, and sex. Odds ratios adjusted for frequency-matching variables (encounter year [2006–2014], age [< 1, 1– < 5, 5– < 15, 15– < 25, 25– < 40, 40– < 65, ≥65 years], sex, and all variables as shown in the above table. Robust standard errors were clustered by individual's community (census tract in cities, borough, or township). Final model sample sizes after excluding a small number of missing values in age, sex, and race/ethnicity (cases/controls): All cases (9648/48,204); Positive test cases (1786/8943); Diagnosis code and treatment cases (4987/24,915).
[a] For case-control analyses of Lyme disease case subgroups, subgroup cases were compared to a randomly selected sample of control encounters after repeated frequency matching by encounter year, age, and sex.

have stronger seasonal associations. The association with prior outpatient visits may be related to more opportunity for both diagnosis and overdiagnosis (Marques, 2008; Steere et al., 1993). Recent history of non-specific subjective symptoms associated with Lyme disease, malaise or fatigue and myalgia or arthralgia, were associated with higher odds of Lyme disease. A recent study found that Lyme disease cases who received non-standard extended (≥5 weeks) antibiotic therapy had higher odds of several non-specific signs and symptoms prior to Lyme disease diagnosis compared to those who received a standard course of antibiotics (Tseng et al., 2017). These findings suggest that there could be a non-specific prodrome in a subset of individuals diagnosed later with Lyme disease, or that there is a subgroup who seek care for non-specific signs and symptoms who are eventually diagnosed with Lyme disease. Instead of the negative association we hypothesized for allergy or atopy, we observed a positive association that was attenuated only among positive test cases. It may be that skin reactions in early Lyme disease are initially misdiagnosed as allergy.

Clinical studies of PTLDS have estimated a frequency of 10–20% after Lyme disease infection and treatment (Marques, 2008), which is consistent with our findings. Few previous studies have examined whether the frequency of new or worsened non-specific symptoms after Lyme disease diagnosis and treatment is higher than healthy controls, and findings have been mixed (Steere et al., 2016). A recent study of nation-wide insurance claims found that 34.4% of Lyme disease cases had at least one diagnosis potentially related to PTLDS within one year of Lyme disease diagnosis, but this study did not exclude pre-existing diagnoses and used a broader definition of PTLDS (Adrion et al., 2015). Other studies, including two prospective cohort studies of erythema migrans and healthy controls (Cerar et al., 2010; Eliassen et al., 2017b), reported that Lyme diseases cases had similar or lower frequencies of PTLDS than controls. In our study, the proportion of controls who met the definition for $PTLDS_{EHR}$ was relatively high, confirming prior findings that these symptoms are common in the general population. The proportion who met the definition of $PTLDS_{EHR}$ was approximately 9% higher in Lyme disease cases compared to controls. The proportion of cases who had pre-existing diagnoses before Lyme disease diagnosis that are possibly relevant to $PTLDS_{EHR}$ (e.g., malaise or fatigue, myalgia or arthralgia) was also higher than in controls (the prevalence of any of these diagnoses was 3.6% higher in cases compared to controls). However, this difference does not fully explain the 9% difference in frequency of $PTLDS_{EHR}$ by Lyme disease status.

There were important limitations to our study. Individuals in this analysis may have received care from non-Geisinger providers, which would not be reflected in the EHR data. We did not review free text in clinical notes to further identify the presence or absence of signs or symptoms. Although studies often use a Lyme diagnosis code alone to

identify cases, a recent study recommended using complementary approaches (Rutz et al., 2018), supporting our additional use of serology test orders and antibiotic treatment for case identification. Our primary case definition, that required a Lyme disease diagnosis code or a serology CPT code and antibiotic order, was likely highly sensitive but probably included false positives for several reasons. Diagnosis codes may be recorded inappropriately, such as when diagnosis was later ruled out or because of a coding error (Sickbert-Bennett et al., 2010). Health providers often order serological tests reflexively, even where a negative test would be expected (Brett et al., 2014). Antibiotics used for treatment of Lyme disease are commonly prescribed for unrelated conditions, and prescriptions may not have been filled and completed as ordered. Serology can improve positive predictive value for disseminated Lyme disease but has important limitations. Requiring serology excludes early Lyme disease diagnosed by erythema migrans. There is a risk of a false positive if an individual has a low prior probability of disease or a false negative if ordered too early prior to a robust antibody response (Mead, 2015), and individuals can remain IgG positive after previous infection (Hilton et al., 1999). To define PTLDS$_{EHR}$, we ascertained preexisting diagnoses using a 22-week period

Table 4
Characteristics of Geisinger primary care patients with Lyme disease with and without PTLDS$_{EHR}$ 2006–2014.

	All cases[a]			Diagnosis code & treatment cases[a]		
	With PTLDS$_{EHR}$[b] (n = 1556)	Without PTLDS$_{EHR}$[b] (n = 5284)	p-value	With PTLDS$_{EHR}$[b] (n = 735)	Without PTLDS$_{EHR}$[b] (n = 2808)	p-value
Year, n (%)						
2006	49 (3.1%)	241 (4.6%)	0.098	12 (1.6%)	87 (3.1%)	0.034
2007	72 (4.6%)	258 (4.9%)		38 (5.2%)	104 (3.7%)	
2008	123 (7.9%)	443 (8.4%)		49 (6.7%)	236 (8.4%)	
2009	174 (11.2%)	636 (12.0%)		81 (11.0%)	349 (12.4%)	
2010	179 (11.5%)	614 (11.6%)		98 (13.3%)	344 (12.3%)	
2011	292 (18.8%)	1017 (19.2%)		134 (18.2%)	558 (19.9%)	
2012	266 (17.1%)	831 (15.7%)		140 (19.0%)	450 (16.0%)	
2013	333 (21.4%)	1072 (20.3%)		158 (21.5%)	607 (21.6%)	
2014	68 (4.4%)	172 (3.3%)		25 (3.4%)	73 (2.6%)	
Age, years						
mean (SD)	45.5 (19.8)	39.0 (23.1)	< 0.001	44.2 (21.5)	35.5 (24.5)	< 0.001
< 1, n (%)	0 (0.0%)	5 (0.1%)	< 0.001	0 (0.0%)	3 (0.1%)	< 0.001
1– < 5, n (%)	26 (1.7%)	351 (6.6%)		20 (2.7%)	263 (9.4%)	
5– < 15, n (%)	143 (9.2%)	921 (17.4%)		96 (13.1%)	675 (24.0%)	
15– < 25, n (%)	126 (8.1%)	422 (8.0%)		58 (7.9%)	221 (7.9%)	
25– < 40, n (%)	245 (15.7%)	761 (14.4%)		96 (13.1%)	302 (10.8%)	
40– < 65, n (%)	756 (48.6%)	2071 (39.2%)		341 (46.4%)	958 (34.1%)	
≥65, n (%)	260 (16.7%)	753 (14.3%)		124 (16.9%)	386 (13.7%)	
Sex, n (%)						
Female	846 (54.4%)	2500 (47.3%)	< 0.001	389 (52.9%)	1317 (46.9%)	0.004
Male	709 (45.6%)	2784 (52.7%)		346 (47.1%)	1491 (53.1%)	
Unknown	1 (0.1%)	0 (0.0%)		0 (0.0%)	0 (0.0%)	
Season, n (%)						
Winter (Dec–Feb)	203 (13.0%)	564 (10.7%)	< 0.001	74 (10.1%)	206 (7.3%)	0.014
Spring (March–May)	324 (20.8%)	1035 (19.6%)		132 (18.0%)	509 (18.1%)	
Summer (June–August)	618 (39.7%)	2396 (45.3%)		343 (46.7%)	1459 (52.0%)	
Fall (Sept–Nov)	411 (26.4%)	1289 (24.4%)		186 (25.3%)	634 (22.6%)	
Race/ethnicity, n (%)						
White non-Hispanic	1508 (96.9%)	5144 (97.4%)	0.044	712 (96.9%)	2737 (97.5%)	0.14
Hispanic	11 (0.7%)	55 (1.0%)		5 (0.7%)	30 (1.1%)	
Black non-Hispanic	28 (1.8%)	63 (1.2%)		14 (1.9%)	30 (1.1%)	
Other	9 (0.6%)	15 (0.3%)		4 (0.5%)	7 (0.2%)	
Unknown	0 (0.0%)	7 (0.1%)		0 (0.0%)	4 (0.1%)	
Medical Assistance insurance, n (%)						
Never user	1257 (80.8%)	4255 (80.5%)	0.39	596 (81.1%)	2231 (79.5%)	0.43
Sometimes user	251 (16.1%)	828 (15.7%)		116 (15.8%)	463 (16.5%)	
Always user	48 (3.1%)	201 (3.8%)		23 (3.1%)	114 (4.1%)	
History of selected conditions, n (%)						
Malaise or fatigue in the past year	125 (8.0%)	325 (6.2%)	0.008	53 (7.2%)	136 (4.8%)	0.011
Myalgia or arthralgia in the past year	246 (15.8%)	746 (14.1%)	0.096	103 (14.0%)	348 (12.4%)	0.24
Allergy or atopy ever in the past	940 (60.4%)	2909 (55.1%)	< 0.001	444 (60.4%)	1578 (56.2%)	0.040
Charlson comorbidity index, n (%)						
0	357 (22.9%)	1912 (36.2%)	< 0.001	191 (26.0%)	1157 (41.2%)	< 0.001
1–2	603 (38.8%)	1824 (34.5%)		264 (35.9%)	881 (31.4%)	
3–4	360 (23.1%)	928 (17.6%)		170 (23.1%)	476 (17.0%)	
≥5	236 (15.2%)	620 (11.7%)		110 (15.0%)	294 (10.5%)	
Outpatient encounters in past year, mean (SD)	6.2 (4.9)	5.1 (4.3)	< 0.001	6.0 (4.9)	4.8 (3.8)	< 0.001
Community type, n (%)						
Township	1107 (71.1%)	3903 (73.9%)	0.050	547 (74.4%)	2134 (76.0%)	0.46
Borough	363 (23.3%)	1150 (21.8%)		164 (22.3%)	571 (20.3%)	
City	86 (5.5%)	231 (4.4%)		24 (3.3%)	103 (3.7%)	

(*continued on next page*)

Table 4 (*continued*)

	All cases[a]			Diagnosis code & treatment cases[a]		
	With $PTLDS_{EHR}$[b](n = 1556)	Without $PTLDS_{EHR}$[b](n = 5284)	p-value	With $PTLDS_{EHR}$[b] (n = 735)	Without $PTLDS_{EHR}$[b] (n = 2808)	p-value
Community socioeconomic deprivation (CSD)						
Quartile 1, n (%)	417 (26.8%)	1492 (28.2%)	0.38	202 (27.5%)	864 (30.8%)	0.21
Quartile 2, n (%)	486 (31.2%)	1704 (32.2%)		238 (32.4%)	912 (32.5%)	
Quartile 3, n (%)	396 (25.4%)	1267 (24.0%)		185 (25.2%)	620 (22.1%)	
Quartile 4, n (%)	257 (16.5%)	821 (15.5%)		110 (15.0%)	412 (14.7%)	

Abbreviations: SD: standard deviation; EHR, electronic health record; $PTLDS_{EHR}$, post-treatment Lyme disease syndrome in the EHR.
Malaise or fatigue (ICD-9): 780.71 (chronic fatigue syndrome); 780.79 (other malaise and fatigue).
Pain (ICD-9): 719.4x (pain in joint, arthralgia); 729.1 (myalgia and myositis, unspecified); 729.5 (pain in limb); 780.96 (generalized pain); 338.19 (acute pain not elsewhere classified); 338.29 (chronic pain not elsewhere classified); 338.4 (chronic pain syndrome).
[a]Sample restricted to cases and controls under observation in the EHR for at least 26 weeks before and at least 52 weeks after their case or control selection date, and without a history of chronic fatigue, fibromyalgia, or chronic pain: All cases (n = 6840); Diagnosis code & treatment cases (n = 3543).
[b]An individual was considered to have a $PTLDS_{EHR}$ diagnosis if they had at least one diagnosis code in one of three categories, defined below, in the 4–52 weeks after their Lyme disease incident date and none in the 26–4 weeks prior to their Lyme disease incident date.
Memory or cognitive difficulties (ICD-9): 780.93 (memory loss); 799.59 (other signs and symptoms involving cognition).

(26–4 weeks before Lyme disease diagnosis), which may have undercounted individuals with preexisting diagnoses. We selected 26 weeks to allow for sufficient time to seek care for preexisting conditions without sacrificing sample size and potential bias by requiring EHR contact before and after Lyme disease diagnosis. We used fixed 4-week time windows to exclude diagnoses during active infection and antibiotic treatment, respectively, which could have both under- and overestimated the frequency of $PTLDS_{EHR}$. Some non-specific symptoms associated with PTLDS, such as cognitive difficulties, may not be well documented in EHR diagnostic codes.

This study also had several strengths. We combined Geisinger EHR data, including rich longitudinal clinical information on individual-level risk factors, with spatially-linked community-level data over a large area that is endemic for Lyme disease in a state with the highest number of reported cases (Centers for Disease Control & Prevention (CDC), 2017). Geisinger EHR uses EPIC software (Minneapolis, MN), the most common commercial EHR system in the US (McCurry, 2014). We linked laboratory test results, which are not available in administrative claims data, with diagnostic, test order, and medication order codes to create multiple case definitions with varying sensitivity and specificity. The Geisinger EHR captures primary and inpatient care in a network of hospitals, outpatient, and urgent care settings and includes patients using private insurance, Medical Assistance, or without insurance. Studies of Lyme disease using EHR data from patients referred to tertiary care settings (Ichikawa et al., 2017) are likely to miss a substantial proportion of uncomplicated Lyme disease diagnosed and treated in primary and urgent care. Geisinger primary care patients are representative of the study region with respect to age, sex, race/ethnicity, and rural residence (Liu et al., 2013).

Future EHR research on Lyme disease could utilize unstructured clinical note text to improve case-finding for Lyme disease, especially for early Lyme disease, and ascertainment of PTLDS, including severity, quality-of-life measures, and attribution to Lyme disease. Future research is needed to identify risk factors for delays in diagnosis and treatment of Lyme disease, and to characterize risk factors for and outcomes after PTLDS. In partnership with public health agencies, EHR-based reporting and data sharing systems have the potential to lessen the reporting burden and improve the efficiency of both infectious and chronic disease surveillance (Birkhead et al., 2015; Klompas et al., 2017).

5. Conclusions

Lyme disease is a common vector-borne disease with a considerable public health burden, especially in highly endemic areas like Pennsylvania. Research with longitudinal clinical EHR data from a large primary care population linked to secondary community data can improve Lyme disease epidemiology through improved case-finding, surveillance, risk factor identification, and characterization of $PTLDS_{EHR}$.

Funding

This work was supported by the Steven & Alexandra Cohen Foundation (Stamford, CT, USA). The sponsors had no role in the study design, data collection, analysis and interpretation of data, writing of the manuscript, nor in the decision to submit the article for publication.

Conflict of interest

None.

Data sharing

To protect patient privacy, individual-level medical record data containing protected health information (PHI) are not available.

Acknowledgements

We gratefully acknowledge the assistance of several people at Geisinger: Joseph DeWalle for spatial variable creation and cartography, Dione Mercer for program management, and Dustin Hartzel for EHR data extraction.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ttbdis.2018.10.010.

References

Adams, D.A., Thomas, K.R., Jajosky, R.A., Foster, L., Baroi, G., Sharp, P., Onweh, D.H., Schley, A.W., Anderson, W.J., Group, for the N.N.I.C, 2016. Summary of notifiable infectious diseases and conditions — United States, 2015. MMWR Morb. Mortal. Wkly. Rep. 64, 1–143. https://doi.org/10.15585/mmwr.mm6453a1. 2017.
Adrion, E.R., Aucott, J., Lemke, K.W., Weiner, J.P., 2015. Health care costs, utilization and patterns of care following lyme disease. PLoS One 10https://doi.org/10.1371/journal.pone.0116767. e0116767.
Aucott, J.N., 2015. Posttreatment Lyme disease syndrome. Infect. Dis. Clin. N. Am. 29, 309–323. https://doi.org/10.1016/j.idc.2015.02.012.
Bacon, R.M., Kugeler, K.J., Mead, P.S., Centers for Disease, C., Prevention, 2008. Surveillance for Lyme disease–United States, 1992-2006. MMWR Surveill. Summ.

57, 1–9.
Birkhead, G.S., Klompas, M., Shah, N.R., 2015. Uses of electronic health records for public health surveillance to advance public health. Annu. Rev. Public Health 36, 345–359. https://doi.org/10.1146/annurev-publhealth-031914-122747.
Brett, M.E., Hinckley, A.F., Zielinski-Gutierrez, E., Mead, P.S., 2014. U.S. healthcare providers' experience with Lyme and other tick-borne diseases. Ticks Tick Borne Dis. 5, 404–408. https://doi.org/10.1016/j.ttbdis.2014.01.008.
Casey, J.A., Curriero, F.C., Cosgrove, S.E., Nachman, K.E., Schwartz, B.S., 2013. High-density livestock operations, crop field application of manure, and risk of community-associated methicillin-resistant Staphylococcus aureus infection in Pennsylvania. JAMA Intern. Med. 173, 1980–1990. https://doi.org/10.1001/jamainternmed.2013.10408.
Casey, J.A., Pollak, J., Glymour, M.M., Mayeda, E.R., Hirsch, A.G., Schwartz, B.S., 2017. Measures of SES for electronic health record-based research. Am. J. Prev. Med. 54, 430–439. https://doi.org/10.1016/j.amepre.2017.10.004.
Casey, J.A., Schwartz, B.S., Stewart, W.F., Adler, N.E., 2016. Using electronic health records for population health research: a review of methods and applications. Annu. Rev. Public Health 37, 61–81. https://doi.org/10.1146/annurev-publhealth-032315-021353.
Centers for Disease Control & Prevention (CDC), 2017. Reported Cases of Lyme Disease by State or Locality, 2006-2016. . https://www.cdc.gov/lyme/stats/tables.html.
Centers for Disease Control & Prevention (CDC), 1995. Recommendations for test performance and interpretation from the Second National Conference on Serologic Diagnosis of Lyme Disease. MMWR Morb. Mortal. Wkly. Rep. 44, 590–591.
Cerar, D., Cerar, T., Ružić-Sabljić, E., Wormser, G.P., Strle, F., 2010. Subjective symptoms after treatment of early Lyme disease. Am. J. Med. 123, 79–86. https://doi.org/10.1016/j.amjmed.2009.05.011.
Cromley, E.K., Cartter, M.L., Mrozinski, R.D., Ertel, S.H., 1998. Residential setting as a risk factor for Lyme disease in a hyperendemic region. Am. J. Epidemiol. 147, 472–477.
Eliassen, K.E., Berild, D., Reiso, H., Grude, N., Christophersen, K.S., Finckenhagen, C., Lindbaek, M., 2017a. Incidence and antibiotic treatment of erythema migrans in Norway 2005-2009. Ticks Tick Borne Dis. 8, 1–8. https://doi.org/10.1016/j.ttbdis.2016.06.006.
Eliassen, K.E., Hjetland, R., Reiso, H., Lindbaek, M., Tschudi-Madsen, H., 2017b. Symptom load and general function among patients with erythema migrans: a prospective study with a 1-year follow-up after antibiotic treatment in Norwegian general practice. Scand. J. Prim. Health Care 35, 75–83. https://doi.org/10.1080/02813432.2017.1288812.
Fix, A.D., Pena, C.A., Strickland, G.T., 2000. Racial differences in reported Lyme disease incidence. Am. J. Epidemiol. 152, 756–759.
Hilton, E., DeVoti, J., Benach, J.L., Halluska, M.L., White, D.J., Paxton, H., Dumler, J.S., 1999. Seroprevalence and seroconversion for tick-borne diseases in a high-risk population in the northeast United States. Am. J. Med. 106, 404–409. https://doi.org/10.1016/S0002-9343(99)00046-7.
Ichikawa, O., Glicksberg, B.S., Kidd, B.A., Li, L., Dudley, J.T., 2017. Learning and mapping Lyme disease patient trajectories from electronic medical data for stratification of disease risk and therapeutic response. bioRxiv. https://doi.org/10.1101/239020.
Jackson, L.E., Hilborn, E.D., Thomas, J.C., 2006. Towards landscape design guidelines for reducing Lyme disease risk. Int. J. Epidemiol. 35, 315–322. https://doi.org/10.1093/ije/dyi284.
Killilea, M.E., Swei, A., Lane, R.S., Briggs, C.J., Ostfeld, R.S., 2008. Spatial dynamics of lyme disease: a review. Ecohealth 5, 167–195. https://doi.org/10.1007/s10393-008-0171-3.
Klompas, M., Cocoros, N.M., Menchaca, J.T., Erani, D., Hafer, E., Herrick, B., Josephson, M., Lee, M., Payne Weiss, M.D., Zambarano, B., Eberhardt, K.R., Malenfant, J., Nasuti, L., Land, T., 2017. State and local chronic disease surveillance using electronic health record systems. Am. J. Public Health 107, 1406–1412. https://doi.org/10.2105/AJPH.2017.303874.
Lantos, P.M., Branda, J.A., Boggan, J.C., Chudgar, S.M., Wilson, E.A., Ruffin, F., Fowler, V., Auwaerter, P.G., Nigrovic, L.E., 2015. Poor positive predictive value of lyme disease serologic testing in an area of low disease incidence. Clin. Infect. Dis. 61, 1374–1380. https://doi.org/10.1093/cid/civ584.
Liu, A.Y., Curriero, F.C., Glass, T.A., Stewart, W.F., Schwartz, B.S., 2013. The contextual influence of coal abandoned mine lands in communities and type 2 diabetes in Pennsylvania. Health Place 22, 115–122. https://doi.org/10.1016/j.healthplace.2013.03.012.
Mac Kenzie, W.R., Davidson, A.J., Wiesenthal, A., Engel, J.P., Turner, K., Conn, L., Becker, S.J., Moffatt, S., Groseclose, S.L., Jellison, J., Stinn, J., Garrett, N.Y., Helmus, L., Harmon, B., Richards, C.L., Lumpkin, J.R., Iademarco, M.F., 2016. The promise of electronic case reporting. Public Health Rep. 131, 742–746. https://doi.org/10.1177/0033354916670871.
Marques, A., 2008. Chronic Lyme disease: a review. Infect. Dis. Clin. N. Am. 22, 341–360. https://doi.org/10.1016/j.idc.2007.12.011. vii–viii.
McCurry, R., 2014. CVS To Transition EHR To EpicCare EMR. Heal. IT Outcomes.. (Accessed 15 April 2018). https://www.healthitoutcomes.com/doc/cvs-to-transition-ehr-to-epiccare-emr-0001.
Mead, P.S., 2015. Epidemiology of Lyme disease. Infect. Dis. Clin. N. Am. 29, 187–210. https://doi.org/10.1016/j.idc.2015.02.010.
Messier, K.P., Jackson, L.E., White, J.L., Hilborn, E.D., 2015. Landscape risk factors for Lyme disease in the eastern broadleaf forest province of the Hudson River valley and the effect of explanatory data classification resolution. Spat Spatio-temporal Epidemiol. 12, 9–17. https://doi.org/10.1016/j.sste.2014.10.002.
Nelson, C.A., Saha, S., Kugeler, K.J., Delorey, M.J., Shankar, M.B., Hinckley, A.F., Mead, P.S., 2015. Incidence of clinician-diagnosed Lyme disease, United States, 2005-2010. Emerg. Infect. Dis. 21, 1625–1631. https://doi.org/10.3201/eid2109.150417.
Pennsylvania Department of Health, 2017. Enterprise Data Dissemination Informatics Exchange. (Accessed 6 June 2017). http://www.statistics.health.pa.gov/StatisticalResources/EDDIE/Pages/EDDIE.aspx.
Quan, H., Sundararajan, V., Halfon, P., Fong, A., Burnand, B., Luthi, J.C., Saunders, L.D., Beck, C.A., Feasby, T.E., Ghali, W.A., 2005. Coding algorithms for defining comorbidities in ICD-9-CM and ICD-10 administrative data. Med. Care 43, 1130–1139.
Rutz, H., Hogan, B., Hook, S., Hinckley, A., Feldman, K., 2018. Exploring an alternative approach to Lyme disease surveillance in Maryland. Zoonoses Public Health 65, 254–259. https://doi.org/10.1111/zph.12446.
Schiffman, E.K., McLaughlin, C., Ray, J.A.E., Kemperman, M.M., Hinckley, A.F., Friedlander, H.G., Neitzel, D.F., 2018. Underreporting of Lyme and other tick-borne diseases in residents of a high-incidence county, Minnesota, 2009. Zoonoses Public Health 65, 230–237. https://doi.org/10.1111/zph.12291.
Schwartz, B.S., Ford, D.P., Childs, J.E., Rothman, N., Thomas, R.J., 1991. Anti-tick saliva antibody: a biologic marker of tick exposure that is a risk factor for lime disease seropositivity. Am. J. Epidemiol. 134, 86–95. https://doi.org/10.1093/oxfordjournals.aje.a115996.
Seriburi, V., Ndukwe, N., Chang, Z., Cox, M.E., Wormser, G.P., 2012. High frequency of false positive IgM immunoblots for Borrelia burgdorferi in clinical practice. Clin. Microbiol. Infect. 18, 1236–1240. https://doi.org/10.1111/j.1469-0691.2011.03749.x.
Sickbert-Bennett, E.E., Weber, D.J., Poole, C., MacDonald, P.D., Maillard, J.M., 2010. Utility of international classification of diseases, ninth revision, clinical modification codes for communicable disease surveillance. Am. J. Epidemiol. 172, 1299–1305. https://doi.org/10.1093/aje/kwq285.
Steere, A.C., Strle, F., Wormser, G.P., Hu, L.T., Branda, J.A., Hovius, J.W., Li, X., Mead, P.S., 2016. Lyme borreliosis. Nat. Rev. Dis. Prim. 2, 16090. https://doi.org/10.1038/nrdp.2016.90.
Steere, A.C., Taylor, E., McHugh, G.L., Logigian, E.L., 1993. The overdiagnosis of Lyme disease. JAMA 269, 1812–1816.
Tibbles, C.D., Edlow, J.A., 2007. Does this patient have erythema migrans? JAMA 297, 2617–2627. https://doi.org/10.1001/jama.297.23.2617.
Tseng, Y.J., Cami, A., Goldmann, D.A., DeMaria Jr., A., Mandl, K.D., 2015. Using nationwide health insurance claims data to augment Lyme disease surveillance. Vector Borne Zoonotic Dis. 15, 591–596. https://doi.org/10.1089/vbz.2015.1790.
Tseng, Y.J., DeMaria Jr., A., Goldmann, D.A., Mandl, K.D., 2017. Claims-based diagnostic patterns of patients evaluated for Lyme disease and given extended antibiotic therapy. Vector Borne Zoonotic Dis. 17, 116–122. https://doi.org/10.1089/vbz.2016.1991.
U.S Census Bureau, 2010. American Community Survey, 2010. American Community Survey 5-Year Estimates. using American FactFinder. http://factfinder2.census.gov. (Accessed 9 May 2017).
U.S. Census Bureau American Community Survey, 2015. American Community Survey 5-Year Estimates. using American FactFinder. http://factfinder2.census.gov. (Accessed 9 May 2017).
White, J., Noonan-Toly, C., Lukacik, G., Thomas, N., Hinckley, A., Hook, S., Backenson, P.B., 2018. Lyme disease surveillance in New York state: an assessment of case underreporting. Zoonoses Public Health 65, 238–246. https://doi.org/10.1111/zph.12307.
Wormser, G.P., Dattwyler, R.J., Shapiro, E.D., Halperin, J.J., Steere, A.C., Klempner, M.S., Krause, P.J., Bakken, J.S., Strle, F., Stanek, G., Bockenstedt, L., Fish, D., Dumler, J.S., Nadelman, R.B., 2006. The clinical assessment, treatment, and prevention of lyme disease, human granulocytic anaplasmosis, and babesiosis: clinical practice guidelines by the Infectious Diseases Society of America. Clin. Infect. Dis. 43, 1089–1134. https://doi.org/10.1086/508667.