ORIGINAL ARTICLE

Lyme Disease Surveillance Using Sampling Estimation: Evaluation of an Alternative Methodology in New York State

G. Lukacik[1], J. White[1], C. Noonan-Toly[1], C. DiDonato[2] and P. B. Backenson[1]

[1] New York State Department of Health, Albany, NY, USA
[2] NTT Data Inc., Albany, NY, USA

Impacts

- Lyme disease (LD) is highly endemic throughout much of New York State, leading to significant surveillance burden on local health department staff.
- In response to this burden, a system was developed to estimate county-level LD cases based on a 20% sample of positive LD laboratory reports. This alternative to traditional surveillance was intended to reduce investigative workload while maintaining the ability to track LD trends.
- The results of this study indicate that the estimating procedure is accurate and efficient in estimating LD cases at the county level. Use of case estimates for LD should be considered as a useful surveillance alternative by health decision makers at the national level in states with endemic LD.

Keywords:
Lyme disease; surveillance; sampling estimation

Correspondence:
J. White. New York State Department of Health, Empire State Plaza, Corning Tower Room 651, Albany, NY 12237, USA. Tel.: 518.473.4439; Fax: 518.474.7381; E-mail: Jennifer.white@health.ny.gov

Received for publication September 23, 2015

doi: 10.1111/zph.12261

Summary

In the 14-year period from 1993 to 2006, New York State (NYS) accounted for over one-quarter (27.1%) of all confirmed Lyme disease (LD) cases in the United States. During that time period, a nine-county area in south-east NYS accounted for 90.6% of the reported LD cases in the state. Based on concerns related to diminishing resources at both the state and local level and the increasing burden of traditional LD surveillance, the NYS Department of Health (DOH) sought to develop an alternative to traditional surveillance that would reduce the investigative workload while maintaining the ability to track LD trends by developing a system to estimate county-level LD cases based on a 20% random sample of positive laboratory reports. Estimates from this system were compared to observed cases from traditional surveillance for select counties in 2007–2009 and 2011. There were no significant differences between the two methodologies in six of nine evaluations conducted. In addition, in 93 of 98 (94.9%) demographic, symptom and other variable proportion comparisons made between the two methodologies in 2009 and 2011, there were no significant differences found. Overall, using sampling estimates was accurate and efficient in estimating LD cases at the county level. Use of case estimates for LD should be considered as a useful surveillance alternative by health policy makers for states with endemic LD.

Introduction

Lyme disease (LD), caused by the spirochaete *Borrelia burgdorferi*, is the most common vector-borne disease in the United States (US), and in 2013, it was the fifth most common nationally notifiable disease (CDC, 2014). Symptoms of LD can include headache, fever, fatigue, chills, muscle and joint pain, and swollen lymph nodes. A skin rash called erythema migrans (EM), commonly referred to as a 'bull's-eye' rash, occurs in 70–80% of those infected (Steere and Sikand, 2003). Left untreated, some patients experience fatigue, arthritis or cognitive difficulties, or all of these symptoms (Feder et al., 2007).

In New York State (NYS), LD is considered to be endemic in every county outside of New York City (NYC). For the 14-year period from 1993 to 2006, NYS accounted for over one-quarter (27.1%) of all confirmed LD cases in the US (CDC, 2008a). Over that time span, a nine-county area

© 2016 Blackwell Verlag GmbH

in southeast NYS accounted for 90.6% of the LD cases reported in NYS (excluding NYC). Annual county-level LD incidence has been reported as high as 1583 cases per 100 000 population (one LD case for every 63.1 residents, Columbia County) (NYSDOH, 2002a) while annual county case counts have reached as high as 1720 cases (Dutchess County) (NYSDOH, 2002b).

Reporting of human cases and positive human laboratory findings of LD in NYS is mandated by law (10NYCRR 2.10, PHL 2012, NYC Health Code Articles 11 and 13). Electronic laboratory reporting, mandated by state law in 2008, has been shown to substantially increase the number of LD reports under investigation (CDC, 2008b). Using the Electronic Clinical Laboratory Reporting System (ECLRS) in NYS, laboratories can electronically transmit all positive LD results to a single centralized database at NYSDOH where reports are automatically distributed by county of residence to the appropriate county local health departments (LHDs). Following an initial increase when the law went into effect in 2008, the number of laboratories reporting positive LD laboratory results remained stable from 2008 through 2011.

In NYS, excluding NYC, from 2003 to 2005, there were an average 27 075 investigations of positive LD laboratory reports per year, with the two highest volume counties (Dutchess and Westchester) each averaging more than 5200 investigations annually over that period. Investigation of LD cases that are found to meet the LD national surveillance case definition [Council of State and Territorial Epidemiologists (CSTE), 1996, 2008] are entered into the NYS Communicable Disease Electronic Surveillance System (CDESS). The system is a secure, web-based repository to which LHD staff submit reportable infectious disease case reports. Lyme disease cases in CDESS can be broken down into three categories: (i) physician reports (most often with a diagnosis of EM rash); (ii) positive laboratory reports where LD is the only testing performed; and (iii) positive laboratory reports that accompany reports of another tick-borne disease (e.g. a positive LD report accompanied by a positive anaplasmosis report).

In higher incidence counties in NYS, adherence to the requirement for complete investigation of all potential LD cases is burdensome, affecting public health staff's ability to conduct prevention activities as well as other essential public health work. Highly endemic counties in NYS may have to investigate thousands of positive LD laboratory reports each year, each of which requires provider follow-up. As LD is not communicable from person to person, public health authorities at both the State and LHDs felt that resources could be better spent on LD prevention or on investigation and control of other communicable diseases. Due to the substantial increase in case investigations – the number of positive LD laboratory reports requiring investigation more than doubled from 2005 (18 420 reports) to 2008 (38 503 reports) – NYSDOH sought to reduce the workload of overburdened LHDs while still maintaining the ability to track LD trends. Therefore, NYSDOH aimed to better balance resources and burden, yet maintain precision and accuracy by creating an alternative surveillance method.

Methods

In 2006, retrospective analyses were conducted prior to the implementation of alternative surveillance to evaluate various sampling strategies to estimate county-level LD cases. These initial evaluations showed that the proportion of ten data elements in CDESS [sex, three age categorical factors (LD modal: 5–14, 20–29 and 60–69), arthritis, EM rash, laboratory test conducted, season (June through August), and two geographic factors (one based on ZIP code population and one on ZIP code location)] examined in three different sample sizes (10%, 20% and 33%) were rarely significantly different from the population proportion for cases in five counties (Albany, Dutchess, Orange, Putnam and Westchester counties) over a 5-year period (2001–2005) (Table 1). While both the 10% and 33% samples provided an accurate representation of the population investigated, it was decided that the 20% sample size would be better suited to reducing burden while still maintaining a sufficient level of completed case investigations to allow for continued trend analysis.

Table 1. Evaluation of randomly selected sample sizes for sampling estimation, using retrospective data[a]

	Percentage of time sample significantly different than population* Sample size		
Factor	10%	20%	33.3%
Sex	3.1	2.5	1.7
Age 5–14	3.9	3.9	2.1
Age 20–29	2.5	2.4	1.1
Age 60–69	3.5	2.9	2.0
Arthritis	2.3	3.2	1.7
EM rash	3.6	2.1	1.7
Laboratory test conducted	4.4	1.9	1.7
Season	3.6	3.1	1.6
ZIP code (population size)	4.3	2.3	2.1
ZIP code (geographic)	4.5	2.8	1.3
All ten comparisons	3.6	2.7	1.7

*$P < 0.05$.

[a]Comparisons were 30 random samples of each sample size, for each of five counties (Albany, Dutchess, Orange, Putnam and Westchester), for each of 5 years (2001–2005).

© 2016 Blackwell Verlag GmbH

In 2007, four of the counties with significant LD investigative burden in 2006 (Dutchess, Orange, Suffolk, and Westchester counties) agreed to conduct LD surveillance via the sampling methodology. The remaining 53 non-NYC counties followed the traditional surveillance methodology. In 2009, 12 additional counties became eligible (≥50 LD cases on average over a 3-year period, a threshold chosen to address both incidence rate and investigative burden) and opted to conduct surveillance using sampling estimation bringing the total number of participating counties to 16 (28% of all NYS counties outside of NYC). By 2013, 19 (33%) counties were conducting the new methodology statewide (Fig. 1).

For the sampling methodology, a random number generator was used to generate and assign a 'select' value to 20% of all laboratory reports. Those records assigned the select value were assigned to counties for investigation. The one exception to this method was for multidisease laboratory reports (e.g. a positive LD report accompanied by a positive anaplasmosis report), which required investigation by the county. For the purpose of calculating estimates, persons with multiple ECLRS laboratory reports were eligible for assignment based on their initial positive laboratory test result annually.

After case investigation of select laboratory reports was completed, and case classification determined using the national surveillance case definition, estimates were calculated by multiplying confirmed and probable cases by five (the inverse sampling fraction). Cases in CDESS with no identified match in ECLRS (physician-reported EM only cases) and CDESS cases created from ECLRS multitick-borne disease reports were added to the calculated estimate.

To evaluate the accuracy of the estimates generated in 2007 and 2008, Dutchess County investigated all ECLRS reports (including both sampled and non-sampled reports) to allow for comparison with the traditional surveillance methodology (observed LD cases). This evaluation was repeated in 2009 for four counties: Onondaga, Rensselaer, Rockland and Washington and 2011 for three counties: Albany, Onondaga and Washington. The counties selected represent counties that are a mix of urban and rural and historically endemic and newly endemic. In 2011, NYSDOH staff performed traditional surveillance for participating counties (i.e. investigated 100% of positive LD laboratory reports). To test the accuracy of the estimates, the number of observed LD cases was compared to the calculated estimate. A 95% confidence interval was generated around the estimate; the accuracy of the estimate was determined based on whether the observed number of LD cases was contained within the confidence interval.

Thirteen demographic, symptom and other data elements were tested in 2009 and 2011 for differences using the *Z*-test for proportions. The test is used to determine whether two populations or groups differ significantly on a

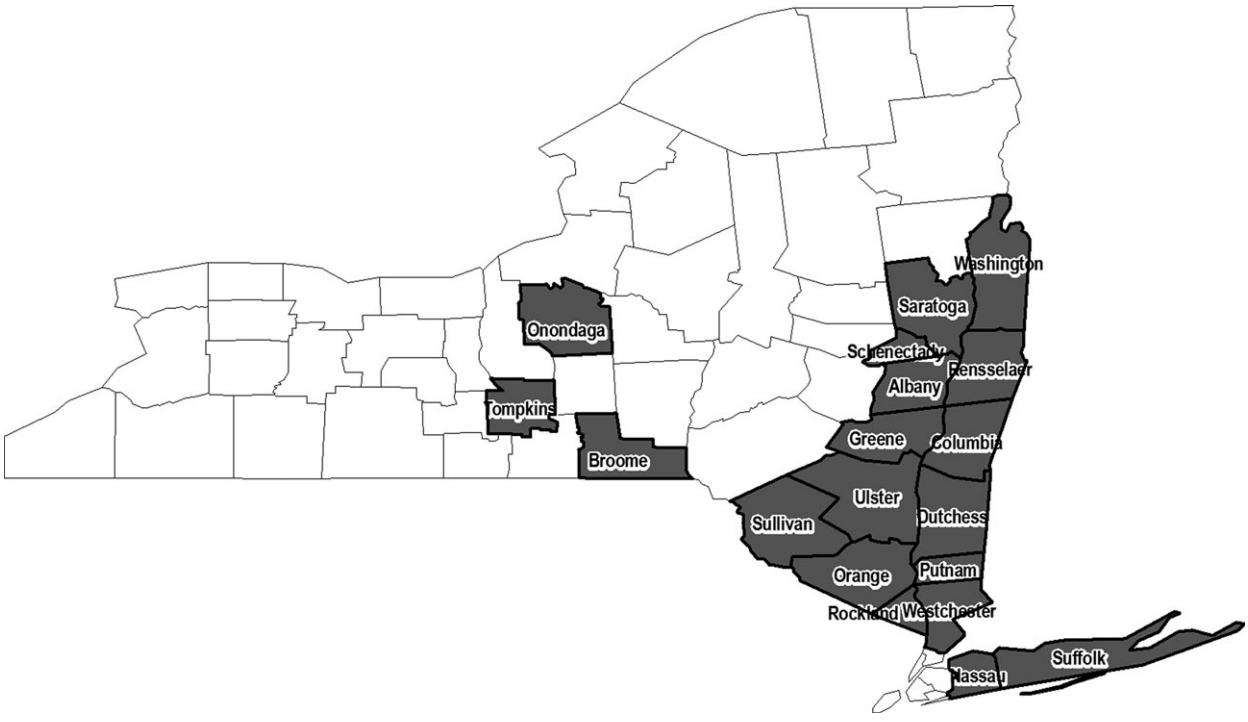

Fig. 1. Counties conducting Lyme disease surveillance using sampling estimation – 2013.

© 2016 Blackwell Verlag GmbH

single characteristic. The characteristics examined included sex, age, race, physician-diagnosed EM rash, arthritis with objective joint swelling, cranial neuritis, LD encephalomyelitis, lymphocytic meningitis, radiculoneuropathy, acute cardiac 2nd or 3rd degree atrial-ventricular conduction defect, physician-diagnosed LD, presence of laboratory data (in the CDESS record), and presence of tick bite.

Maps were made using ARCMAP 10.1 (Redlands, CA, USA). Analyses were conducted using SAS 9.3 (Cary, NC, USA).

Results

In 2007 and 2008, the number of observed cases in Dutchess County fell within the 95% interval calculated around the estimate (Table 2). In 2009, however, three of the four examined counties' (Onondaga, Rensselaer and Rockland) observed numbers fell outside the interval calculated around the estimate, indicating that they were significantly different from the estimate. A final review of three counties (Albany, Onondaga and Washington) in 2011 found that all observed case numbers were within the confidence interval around the estimate.

The representativeness of observed and estimated proportions of 13 CDESS demographic, symptom and other data elements was examined for 2009 and 2011. In 2011 (Table 3), of the 42 comparisons made in three counties, 38 were not significantly different at the $P < 0.05$ level. Data from 2009 showed similar correlation between observed and estimated proportions [98.2% of

Table 2. Observed and estimated Lyme disease case numbers in New York State by year and county in which both sampling estimation and standard surveillance were conducted

				95% confidence interval on estimate	
Year	County	Observed	Estimated	Lower	Upper
2007	Dutchess	276	310	253.2	366.8
2008	Dutchess	898	935	833.2	1036.8
2009	Onondaga	17[a]	55	40.4	69.6
2009	Rensselaer	148[a]	445	401.7	488.3
2009	Rockland	139[a]	465	422.2	507.8
2009	Washington	91	105	74.9	135.1
2011	Albany	292	245	185.9	304.1
2011	Onondaga	72	60	31.2	88.8
2011	Washington	134	140	103.8	176.2
2007	Total	276	310	253.2	366.8
2008	Total	898	935	833.2	1,036.8
2009	Total	395[a]	1070	999.4	1,140.6
2011	Total	498	445	369.5	520.5

[a]Not within 95% confidence interval.

Table 3. Lyme disease data element comparison – estimated versus observed percentages in three counties, 2011

	Albany		Onondaga		Washington	
	Estimated %	Observed %	Estimated %	Observed %	Estimated %	Observed %
Sex						
Male	47.7	54.5	60.3	68.8	58.6	56.2
Age						
≤18	18.9	18.7	16.2	33.8*	27.4	19.2*
19–49	40.4	34.3	67.6	32.5	20.3	24.6
50+	40.4	46.7	16.2	33.8	51.5	55.4
Average age	43.9	41.2	35.4	35.8	48.0	45.4
Race (white)	89.5	91.5	98.6	100.0	98.7	94.3
EM rash	51.2	54.8	63.2	61.3	71.4	72.7
Arthritis	18.6	14.2	14.7	15.0	15.4	16.5
Cranial neuritis	4.9	6.3	7.4	11.3	2.3	3.5
Encephalomyelitis	0.0	0.3	0.0	0.0	0.0	0.4
Lymphocytic meningitis	0.4	0.6	0.0	2.5	0.4	1.2
Radiculoneuropathy	0.7	3.0*	7.4	2.5	0.4	1.5
Acute cardiac defect	0.4	1.5	0.0	1.3	0.4	0.8
Physician-diagnosed LD[a]	87.0	91.0	100.0	93.8*	99.2	99.2
Lab data missing[b]	10.2	10.8	10.3	8.8	48.5	49.6
Tick bite	26.3	25.9	17.6	21.3	39.5	40.8

*$P < 0.05$.

[a]From CDESS report form: has a physician diagnosed this patient with Lyme disease?

[b]No laboratory information on CDESS form.

© 2016 Blackwell Verlag GmbH

comparisons made in four counties were not significantly different at the $P < 0.05$ level (data not shown)]. Overall, in the 2 years of examination, there was very little discrepancy (5.1%) between the observed and estimated proportions. The only substantive difference in the characteristics investigated occurred in categorical age in Onondaga County in 2011 where there was a small number of cases. Most characteristics which are important for tracking LD trends and targeting intervention (sex, average age, and all symptom characteristics except radiculoneuropathy in Albany County in 2011) were shown not to be significantly different across all counties in 2009 and 2011.

Discussion

In the 7 years (2007 through 2013) that sampling estimation has been used in NYS, evaluation of the accuracy of LD estimates at the county level was conducted in nine counties (total) in four separate years. The estimates generated for six of the counties were accurate as the observed number of cases were well within the confidence intervals around the estimates. Similar retrospective studies undertaken in Minnesota (MN) revealed that 20% and 50% samples of laboratory-reported LD produced estimated LD case counts that were similar to the observed counts (Friedlander et al., 2013). The results of the evaluations in NYS and MN revealed that the methodology used in both states could produce accurate case estimates with 20% of the traditional surveillance effort.

A comparison of proportions of 13 demographic, symptom and other characteristics in NYS in 2009 and 2011 showed that the proportions generated from estimates were highly representative of the observed proportions. Ten of the data elements examined (sex, average age, race, EM rash, arthritis, cranial neuritis, encephalomyelitis, lymphocytic meningitis, acute cardiac defect and tick bite) were not significantly different for both years in all seven counties. Similar results were also found in MN and MA when examining most demographic and risk factors over the 5-year analysis (Friedlander et al., 2013).

There are several limitations to this study. Electronic laboratory reports are assigned to county health departments for investigation in ECLRS based on the reported patient address. If patient addresses are not provided by the testing laboratory, ECLRS assigns the report to the county of the ordering provider or facility. This process may have led to reports being incorrectly assigned to a county different than the patients' residence, although the authors feel that this misclassification would be minimal. Another limitation involved the matching of ECLRS and CDESS cases. Due to name and date of birth data entry errors, a small number of records may not match for the same person in both ECLRS and CDESS.

Observed case counts were found outside the boundary calculated around the estimate in three counties in 2009. The authors feel the difference in the LD investigation procedure was the reason for the significant differences. Unlike other years in which evaluations were conducted, in 2009, due to competing public health priorities (including response to pandemic H1N1 influenza), there was a lag of several months between receipt of a positive laboratory report and initiation of an LD case investigation for cases not in the 20% sample. Conversely, the cases in the 20% sample were investigated within days of receipt of a positive laboratory report. Due to the length of time that elapsed between the patient's diagnosis and the initiation of the LD case investigation for the cases not in the 20% sample, the authors feel the medical practices may have been less likely to report LD cases. This lack of reporting may have been due to the additional effort required to search past medical records for clinical information related to the LD diagnosis or that case reporting was considered less critical many months after diagnosis. As shown in Table 2, the estimate bounds were much higher than the observed number of cases for three of the four counties in which there was a wide time discrepancy, as the investigations for the estimate calculation were conducted in a timelier manner. Despite the fact that the observed case counts were outside the boundary calculated around the estimate in these three counties, the estimated case counts were consistent with trends in surrounding counties.

One of the goals of LD sampling estimation is to allow for more staff time to be spent on more complex individual tick-borne case investigations. However, many county health departments reassigned staff to other communicable disease activities following their enrollment in the sampling system. This would be beneficial to counties in their investigation and attempted prevention of other diseases but would not increase the quality of the information garnered from LD sampling. The sampling system would also be unsuited for the detection of low or limited volume outcomes such as acute 2nd or 3rd degree cardiac atrio-ventricular conduction defect.

Additional limitations involve discrepancies with reported LD case numbers. Sampling estimates create variability both within the state (pre- and post-estimate same-county comparisons) and across states, although there has always been some degree of dissimilarity between states. At present, NYSDOH publishes estimates on the NYSDOH public website. CDC, with guidance from the CSTE, publishes non-estimated case counts for NYS. The discrepancy has caused confusion among the public and the media. CDC-reported LD case counts for NYS have appeared to diminish over the last several years (CDC, 2013), when in fact LD has not decreased in NYS. To date, CSTE has not supported the publication of estimates for LD.

© 2016 Blackwell Verlag GmbH

Overall, sampling estimation was efficient and accurate in estimating LD cases at the county level. From 2006 to 2013, there was a substantial increase in the number of LD laboratory reports in NYS, in counties that do and do not conduct sampling. For those counties conducting sampling, the investigative burden has been reduced without resulting in a loss of important LD trend data. It is our hope that public health policy makers will consider the use of case estimates for diseases such as LD, at least in states with a long standing history of endemic LD.

Acknowledgements

We thank the following for their contributions to this work: Dr. Dennis White, Dr. Russell Rockwell, Christen Hertzog, Andrew Rotans, Rebecca Hoen, Carol Lyn Piazza, Dr. Hwa-Gan Chang, Jacquelin Griffin, David DiCesare, Dr. Alison Hinckley, Sarah Hook and staff at participating local health departments.

References

CDC, 2008a: Reported Lyme disease cases by state, 1993–2007. Available at: http://www.aldf.com/LymeDiseaseReportedCasesbyState19932007.pdf (accessed on 25 August 2014).

CDC, 2008b: MMWR – Effect of Electronic Laboratory Reporting on the Burden of Lyme Disease Surveillance – New Jersey, 2001–2006. Available at: http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5702a4.htm (accessed on 17 September 2014).

CDC, 2013: Reported cases of Lyme disease by state or locality, 2003–2012. Available at: http://www.cdc.gov/lyme/stats/chartstables/reportedcases_statelocality. html (accessed on 5 November 2013).

CDC, 2014: Lyme Disease Data – Fast Facts. Available at: http://www.cdc.gov/lyme/stats/index.html (accessed on 7 January 2015).

CSTE, 1996, 2008: Revised national surveillance case definition for Lyme disease. Available at: http://c.ymcdn.com/sites/www.cste.org/resource/resmgr/PS/07-ID-11.pdf (accessed on 15 May 2013).

Feder, H. M., B. J. Johnson, S. O'Connell, E. D. Shapiro, A. C. Steere, and G. P. Wormser, 2007: A critical appraisal of chronic Lyme disease. *N. Engl. J. Med.* 357, 1422–1430.

Friedlander, H., K. Smith, and D. Nietzel, 2013: Sampling Lyme Disease Reports to Calculate Disease Estimates: Further Evaluation of an Approach to Reduce Surveillance Burden, Minnesota, 2005–2010. CSTE Annual Conference Proceedings, Pasadena, CA.

NYSDOH, 2002a: 2002 Communicable Disease – Annual Report, Rate per 100,000 Population by Disease and County – 2002. Available at: http://www. health.ny.gov/statistics/diseases/communicable/2002/rates2.htm (accessed on 9 July 2103).

NYSDOH, 2002b: 2002 Communicable Disease – Annual Report, Reported Cases by Disease and County – 2002. Available at: http://www.health.ny. gov/statistics/diseases/communicable/2002/cases2.htm (accessed on 9 July 2103).

Steere, A. C., and V. K. Sikand, 2003: The presenting manifestations of Lyme disease and the outcomes of treatment. *N. Engl. J. Med.* 348, 2472–2474.

 © 2016 Blackwell Verlag GmbH