Estimation of the Incidence of Lyme Disease

Grant L. Campbell,1 Curtis L. Fritz,1,2 Durland Fish,3 John Nowakowski,4 Robert B. Nadelman,4 and Gary P. Wormser4

The incidence of Lyme disease in most endemic areas is unknown but will be an important factor in determining the cost-effectiveness of Lyme disease vaccines in those areas. The authors developed a deterministic model with nine components to estimate the frequency of Ixodes scapularis tick bites and the resulting incidence of Lyme disease in residents of endemic areas. For each component, best point estimates and plausible ranges of values were based on the published literature, unpublished data, expert opinion, or a combination of the above. By using the mean, crude, annual total of 3,827 Lyme disease cases reported from the endemic county of Westchester, New York, in 1991–1994, a mean of 178,889 I. scapularis bites (20.4 per 100 person-years) and a mean of 10,632 incident Lyme disease cases (1.2 per 100 person-years) were estimated to have occurred per year. Results of a sensitivity analysis that used two different methods suggested that this deterministic model is reasonably robust. In conclusion, according to this model, the incidence of Lyme disease in Westchester County is several-fold higher than suggested by the current passive reporting system. Am J Epidemiol 1998;148:1018–26.

Lexodes; Lyme disease; theoretical models; tick infestations

Lyme disease is caused by infection with the spirochete Borrelia burgdorferi, which is transmitted to humans by the bites of ticks in the genus Ixodes (1). The principal vectors in the United States are Ixodes scapularis (formerly I. dammini) in the northeastern and north-central regions, and I. pacificus in the Pacific coastal region. Lyme disease is the most frequently reported arthropod-borne disease in the United States. Since voluntary national reporting began in 1982, more than 100,000 cases have been reported to the Centers for Disease Control and Prevention. A mean of more than 10,000 cases were reported per year during 1988–1996 (2). As progress is made toward developing an effective vaccine for preventing the disease in humans (3), defining populations that would most benefit from vaccination is a subject of growing interest. Lyme disease incidence in a given area will be the most important factor. For most areas, however, only crude estimates of incidence based on passive case reporting systems are available. More refined estimates of Lyme disease incidence based on active case surveillance are available for only a few areas (4). In turn, the factor that drives Lyme disease incidence in endemic areas is almost certainly the frequency of bites by ticks infected with B. burgdorferi. No surveillance system exists for the reporting of persons with recognized tick bites and no population-based studies of tick bite incidence have been reported. Estimates of the frequency of unrecognized tick bites, which clearly is a key factor in the epidemiology of Lyme disease (5), are unavailable.

In the current study, a simple deterministic model was developed to estimate the frequency and incidence of I. scapularis bites and the resulting incidence of Lyme disease in endemic areas. This model was then used to estimate these parameters during 1991–1994 in Westchester County, New York, which perennially is among the counties with the highest reported incidences of Lyme disease in the United States (1).

MATERIALS AND METHODS

Development of the deterministic model

The first step in estimating the incidence of Lyme disease in endemic areas was to diagram the universe of persons who experience Ixodes scapularis bites in
Lyme disease-endemic areas and its intersection with the universe of Lyme disease cases (figure 1). Next, an algorithm was developed to represent all such bites and subsequent Lyme disease-related events associated with them (figure 2). In constructing these figures and the resulting deterministic model, the following assumptions were made: 1) antimicrobial prophylaxis of recognized *I. scapularis* bites is 100 percent effective in preventing Lyme disease; 2) all persons with recognized recent *I. scapularis* bites who do not receive antimicrobial prophylaxis and who then develop Lyme disease are closely monitored and therefore diagnosed and treated in an early (≤1 month duration) clinical stage; 3) all Lyme disease patients diagnosed in an early clinical stage present with erythema migrans (EM), the characteristic skin lesion of early Lyme disease (1); 4) all patients with EM are seen by a health care provider and are correctly diagnosed; and 5) Lyme disease patients first diagnosed in a late (>1 month duration) clinical stage do not recall the tick bite that transmitted *B. burgdorferi* to them.

Branch A of the algorithm represents the frequency of recognized bites, which is the sum of the respective frequencies of recognized bites for which antimicrobial prophylaxis was not provided (branch A₁) and for which antimicrobial prophylaxis was provided (branch A₂). Branch B represents the frequency of unrecognized bites. Thus, the overall frequency of *I. scapularis* bites is represented by:

$$\text{BITE}_{\text{freq}} = A_1 + A_2 + B.$$ 

To express branches $A_1$, $A_2$, and $B$ as algebraic terms, nine components (a through i, table 1) were used:

- $A_1 = \frac{abde}{cf}$
- $A_2 = \frac{abde}{cf} \left( \frac{g}{1 - g} \right)$
- $B = \left[ \frac{abe(1 - d)}{c} + \frac{ab(1 - e)}{ci} \right] \frac{1}{h}$

**FIGURE 1.** Venn diagram representing the universe of persons experiencing *Ixodes scapularis* bites in Lyme disease-endemic areas and its intersection with the universe of Lyme disease cases. The proportionate area of the various domains is not to scale.
where the common term \( ab/c \) represents the number of diagnosed Lyme disease cases that are true cases. Model components \( b \) through \( i \) are diagrammed in figure 3. Using regional population totals \( P \) obtained from the US Census Bureau, conversion of \( I. scapularis \) bite frequency to incidence (bites per 100 person-years) was done in typical fashion:

\[
\text{BITE}_{\text{rate}} = \frac{\text{BITE}_{\text{freq}} \times 100}{P}
\]

It follows that: 1) the total number of EM cases that resulted from recognized \( I. scapularis \) bites for which antimicrobial prophylaxis was not provided is represented by \( A_1 f \); 2) the total number of EM cases that were prevented by antimicrobial prophylaxis of recognized bites is represented by \( A_2 f \); 3) the total number of new Lyme disease cases that resulted from unrecognized bites is represented by \( B h \); and, thus, 4) the overall frequency of new Lyme disease cases is represented by:

\[
\text{CASE}_{\text{freq}} = A_1 f + B h.
\]

Conversion of case frequency to incidence (cases per 100 person-years) was done in typical fashion:

\[
\text{CASE}_{\text{rate}} = \frac{\text{CASE}_{\text{freq}} \times 100}{P}
\]

**Best point estimates and plausible ranges of model components**

To estimate the value of model component \( a \), we used 1991–1994 Lyme disease case report totals from Westchester County (an average of 3,827 cases per year; range, 2,688 to 5,830; G. Jacquette, Westchester County Department of Health, personal communication, 1996). These were crude totals; no consideration was made for whether or not cases met the national surveillance case definition of Lyme disease (6). Plausible ranges and best point estimates for values of model components \( b \) through \( i \) were determined from the published literature, unpublished data, expert opinion of the current authors, or a combination of the above (table 1). For model component \( e \), it was assumed that a diagnosis of Lyme disease based on the presence of EM was generally more accurate than such a diagnosis in the absence of EM. Thus, the best point estimate for \( e \) was considered to be the proportion of Lyme disease cases that reportedly included EM (0.619) among all cases reported from New York in 1991–1994 that met the national surveillance case definition (12,937) (unpublished data), plus 10 percent. For model component \( h \), the minimum (0.02) of

\[
\text{CASE}_{\text{rate}} = \frac{\text{CASE}_{\text{freq}} \times 100}{P}
\]
TABLE 1. Definitions, plausible ranges, best point estimates, and methods used to determine best point estimates of components of a deterministic model of the average annual frequency of *Ixodes scapularis* bites and the incidence of Lyme disease (LD) in human populations

<table>
<thead>
<tr>
<th>Model component</th>
<th>Plausible range</th>
<th>Best point estimate</th>
<th>Method used to determine best point estimate</th>
<th>Reference or source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Symbol a</td>
<td>LD cases, diagnosed, reported*</td>
<td>0.321–0.53</td>
<td>0.307</td>
<td>Simple annual average</td>
</tr>
<tr>
<td>Symbol b</td>
<td>LD cases, diagnosed, true cases</td>
<td>0.08–0.23</td>
<td>0.160</td>
<td>Midpoint of range</td>
</tr>
<tr>
<td>Symbol c</td>
<td>LD cases, diagnosed, reported</td>
<td>0.14–0.52</td>
<td>0.300</td>
<td>Weighted average of raw values from references</td>
</tr>
<tr>
<td>Symbol d</td>
<td>LD cases, diagnosed, true cases, EM† recognized recent tick bite</td>
<td>0.53–0.83§</td>
<td>0.661§</td>
<td>Single point estimate</td>
</tr>
<tr>
<td>Symbol e</td>
<td>LD cases, diagnosed, true cases, EM</td>
<td>0.01–0.05</td>
<td>0.020</td>
<td>Weighted average of raw values from references</td>
</tr>
<tr>
<td>Symbol f</td>
<td>Persons with recognized <em>I. scapularis</em> bite, not prophylaxed, who then develop EM</td>
<td>0.22–0.57</td>
<td>0.267</td>
<td>Weighted average of raw values from references</td>
</tr>
<tr>
<td>Symbol g</td>
<td>Persons with recognized <em>I. scapularis</em> bite, prophylaxed</td>
<td>0.02–0.25§</td>
<td>0.135</td>
<td>Midpoint of range</td>
</tr>
<tr>
<td>Symbol h</td>
<td>Persons with unrecognized <em>I. scapularis</em> bite who then develop EM or late LD</td>
<td>0.35–0.65§</td>
<td>0.500</td>
<td>Single point estimate</td>
</tr>
</tbody>
</table>

* Crude total of case reports without consideration for whether or not cases meet the national surveillance case definition (6).
† Minimum is the proportion of all cases reported to the Westchester County Department of Health during 1991–1994 that met the national surveillance case definition (G. Jacquette, Westchester County Department of Health, personal communication, 1996).
‡ EM, erythema migrans; CDC, Centers for Disease Control and Prevention.
§ See text.

the plausible range was based on the assumption that unrecognized *I. scapularis* bites were at least twice as likely as recognized bites (0.01, the lower limit of the plausible range of component f, table 1) to result in the transmission of *B. burgdorferi*; the maximum (0.25) was based on the observation that, in Lyme disease-endemic regions of the northeastern United States, 20–35 percent (average, approximately 25 percent) of nymphal *I. scapularis* typically are infected with *B. burgdorferi* (24), and that more than 48 hours of feeding is usually required for spirochetal transmission (25). For model components for which only single point estimates were available (*e* and *i*), plausible ranges were arbitrarily considered to be the point estimates ±15 percent. For all rate calculations, the 1990 US Census Bureau figure for the total Westchester County population (874,866 persons) was used.

**Sensitivity analysis of the deterministic model**

Two methods were used to evaluate the sensitivity of the deterministic model. The first method, which used a form of bootstrap sampling (26), consisted of generating 10,000 random covariate patterns for the nine model components within their respective plausible ranges, calculation of the 10,000 corresponding BITE rate values by using a computer spreadsheet, and examination of the frequency distribution of these BITE rate values. The second method ("univariate incrementation") consisted of varying each model component in 10 percent increments throughout its plausible range while holding all other components constant at their respective best point estimates. The effects of these changes on the corresponding BITE rate values were evaluated by graphic inspection.

**RESULTS**

When this deterministic model was used, the estimated mean frequency of *I. scapularis* bites in Westchester County during 1991–1994 was 178,889 bites per year. The estimated frequency of recognized bites for which antimicrobial prophylaxis was not provided (model term $A_1$) was 82,339 per year (46.0 percent of all bites), which resulted in an estimated 1,647 incident Lyme disease cases per year ($A_1 f$).
estimated frequency of recognized bites for which antimicrobial prophylaxis was provided (term \( A_2 \)) was 29,993 per year (16.8 percent); prophylaxis was estimated to have prevented 600 Lyme disease cases per year (\( A_2f \)). The estimated frequency of unrecognized bites (term \( B \)) was 66,557 per year (37.2 percent), which resulted in an estimated 8,985 incident Lyme disease cases per year (\( Bh \)). Thus, recognized bites were calculated to be 1.7 times more common than unrecognized bites (\( [A_1 + A_2]/B \)) and a mean overall frequency of 10,632 new Lyme disease cases per year (CASE\(_{freq} = A_1f + Bh, \) CASE\(_{rate} = 1.2 \) cases per 100 person-years) were estimated to have occurred in Westchester County during 1991–1994.

Results of the sensitivity procedure involving bootstrap sampling of 10,000 combinations of the nine model components gave a mean BITE\(_{rate} \) estimate of 38.2 (range, 4.9–266.7; 5th-to-95th percentile range, 13.6–31.2) per 100 person-years. Overall, the value of BITE\(_{rate} \) was most sensitive to changes in model component \( h \), particularly at the lower end of its plausible range, and least sensitive to changes in \( e \) or \( i \).

**DISCUSSION**

A deterministic model with nine components was developed to estimate the incidence of Lyme disease and the frequency of \( I. \) scapularis bites of humans in Lyme disease-endemic areas. The results of tests of model sensitivity by two separate methods suggested that it is reasonably robust, i.e., that the output of the model is reasonably insensitive to small changes in the values of its components. Nevertheless, estimates derived from the model should be viewed as crude and preliminary because the model itself is imperfect, some of the assumptions involved in its construction may not be valid, and the methods for determining the plausible ranges and best point values of its components generally are imprecise.
FIGURE 4. Line graphs of the effects on the estimated mean yearly frequency of *Ixodes scapularis* bites (BITE$_{\text{mean}}$) in Westchester County, New York, during 1991–1994 of varying each of the nine deterministic model components (a through i) in 10 percent increments throughout its plausible range while holding the values of all other components constant at their respective best point estimates. For definitions of a through i, see table 1 and figure 3.
Using this model, we estimated that a mean of nearly 179,000 *I. scapularis* bites of persons in Westchester County, New York, occurred per year during 1991–1994. If no persons were bitten more than once in a given year, a mean of about one in five Westchester County residents sustained an *I. scapularis* bite each year. In Lyme disease-endemic areas, however, the number of persons who are multiply bitten in a given year may be significant. For example, in a study of antimicrobial prophylaxis of tick bites in Westchester County (27), 12 percent of persons who presented with *I. scapularis* bites were bitten again by this species within a 6-week observation period. In contrast, the current model predicts that if tick bites occurred randomly according to a standard Poisson distribution, only 1.2 percent of bitten persons would be expected to receive a second bite within 6 weeks (data not shown). In any case, the risk of being bitten by *I. scapularis* in Lyme disease-endemic areas appears to be non-uniform and depends on such factors as residential proximity to woodlands and other landscape features (28), avocation, and occupation.

To our knowledge, the only previous example of mathematical modeling of the frequency of *I. scapularis* bites in human populations was by Falco and Fish (19). Based on a total of 381 Lyme disease cases reported in Westchester County during 1985 and an estimated crude *I. scapularis* bite-to-reported case ratio of 35.5 (71/2), these authors estimated that some 13,500 *I. scapularis* bites occurred in the county that year. When the present and more complex model was used on 1991–1994 data, a crude *I. scapularis* bite-to-reported case ratio of 46.7 (178,889/3,827) was estimated. The fact that our estimate of the frequency of *I. scapularis* bites was more than 13-fold higher than that of Falco and Fish (19) could be due, at least in part, to the dramatic overall increase in the Westchester County population density of *I. scapularis* during 1985–1991. In fact, between 1984 and 1991, the overall density of immature *I. scapularis* in the county increased by as much as 34-fold, based on field estimates (29). In contrast, from 1980 to 1990, the human population of the county increased by less than 1 percent (US Census Bureau statistics). Despite a virtually stable human population size, the yearly totals of confirmed Lyme disease cases reported by Westchester County increased fairly steadily from 100 cases in 1986 (the year data collection began) to more than 1,500 cases in 1991 (D. Fox, New York State Department of Health, personal communication, 1998).

Using the current model, we estimated that a mean of 10,632 incident Lyme disease cases occurred per year in Westchester County during 1991–1994. Most of these cases (8,985 cases or 84.5 percent) are assumed to have resulted from unrecognized *I. scapularis* bites and, thus, were not reasonably preventable with prophylactic antimicrobials. Furthermore, an estimated 82,339 additional prescriptions for prophylactic antimicrobials would have been required per year to prevent the remaining 1,647 (15.5 percent) incident Lyme disease cases (i.e., 50 courses of antimicrobials per case of Lyme disease prevented, which derives directly from the best point estimate of model component f, table 1). An effective, safe, and convenient Lyme disease vaccine, if widely and appropriately administered to at-risk populations, would likely be preferable to an increased use of prophylactic antimicrobials in preventing Lyme disease cases. In the United States, the results of two large-scale efficacy trials of recombinant single protein vaccines for the prevention of Lyme disease in adult volunteers have been completed (3).

Of the six principal assumptions used in developing the current deterministic model, the first (that antimicrobial prophylaxis of recognized *I. scapularis* bites is 100 percent effective in preventing Lyme disease) may be the most defensible. Although no definitive evidence has been published to show that antimicrobials abort incubating *B. burgdorferi* infections, the view is supported by clinical experience with two other spirochetal but non-tickborne diseases, i.e., incubating syphilis and leptospirosis (30), and by the results of laboratory animal studies (31, 32). The animal studies suggest that *B. burgdorferi* usually multiplies locally at the site of tick inoculation for 2–14 days prior to dissemination (31) and that treatment of an *I. scapularis* bite site with topically applied antimicrobials within 2 days of detachment is highly effective in curing the infection (32). However, even if antimicrobial prophylaxis of recognized *I. scapularis* bites is assumed to be fully effective in preventing Lyme disease, the fact that not all patients who receive prophylactic antimicrobials take them as prescribed (22) weakens this assumption to some degree.

The second assumption, that all persons with recognized recent *I. scapularis* bites who do not receive antimicrobial prophylaxis and who then develop Lyme disease are closely monitored and therefore diagnosed and treated in an early clinical stage, may be optimistic. Undoubtedly, there are at least a few persons who sustain recognized *I. scapularis* bites, do not receive prophylactic antimicrobials, and are not closely monitored. A small fraction of these persons would be expected to subsequently develop clinical *B. burgdorferi* infections that are misdiagnosed or remain undiagnosed.

The third assumption is almost certainly simplistic.
in that, while the majority of patients with early *B. burgdorferi* infection probably manifest EM, a minority may instead manifest a nonspecific febrile illness ("summer flu"), facial palsy, or aseptic meningitis, in the apparent absence of EM (33–35). Lyme disease undoubtedly is misdiagnosed or remains undiagnosed in some of these patients. The fourth assumption is also clearly simplistic. Although most cases of EM seen by experienced health care workers in Lyme disease-endemic areas probably are diagnosed correctly, some EM lesions are morphologically atypical and difficult to diagnose accurately (36). The fifth and final assumption is also subject to criticism in that at least a few patients with Lyme disease first diagnosed in a late clinical stage probably can recall both a previous skin lesion compatible with EM and an antecedent tick bite at the site of that skin lesion.

Based on the current model, recognized *I. scapularis* bites were 1.7 times as common as unrecognized bites in Westchester County during 1991–1994. Because nymphal *I. scapularis* are very small and their bites may easily go unnoticed (37), this estimated ratio may seem counterintuitive. In areas such as Westchester County, however, where public awareness of ticks and the importance of their early removal is acute, it is at least conceivable that recognized *I. scapularis* bites have come to predominate over unrecognized bites. Unfortunately, no independent data currently are available with which to test this hypothesis.

In the current model, one variable which particularly influences the estimated ratio of recognized-to-unrecognized *I. scapularis* bites and for which a particularly imprecise best point estimate (13.5 percent) was available is *h*, i.e., the proportion of persons who sustain unrecognized *I. scapularis* bites who then develop Lyme disease (table 1, figure 3). For example, if *h* were decreased to a seemingly reasonable 8 percent and the values of all other model components were held constant, the estimated ratio of recognized-to-unrecognized *I. scapularis* bites in Westchester County during 1991–1994 would decrease from 1.7 to 1.0 (while the estimated mean BITEfreq would increase to 224,646 per year). Valid and important reasons probably exist to explain our current inability to estimate *h* with more precision, most notably an unknown incidence of subclinical infections. This, in turn, is most likely related to variability in the size of spirochetal inocula and to a variety of poorly understood characteristics of the pathogen and its human host.

Note added in proof

Recent evidence (38; G. P. Wormser et al., unpublished data) suggests that a more accurate best point estimate of model component *e*—the proportion of diagnosed, true Lyme disease cases that include EM—is 0.90. Substitution of this value in the current model gives an estimated mean annual frequency of 198,014 *I. scapularis* bites and a mean annual incidence of 8,867 new Lyme disease cases in Westchester County during 1991–1994. The difference between these estimates and those presented above is primarily due to the positive correlation of component *e* with the ratio of recognized to unrecognized tick bites ([*A1* + *A2*]/*B*).

ACKNOWLEDGMENTS

This work was supported in part by cooperative agreements U50/CCU206626-01 (Dr. Fish), U50/CCU210286-02 (Dr. Nadelman), and U50/CCU210280-02 (Dr. Wormser) with the Centers for Disease Control and Prevention, grant no. 28956 from the National Institute of Allergy and Infectious Disease (Dr. Fish), and grant nos. 41508 and 43135 from the National Institute of Arthritis and Musculoskeletal and Skin Diseases (Dr. Wormser).

The authors thank Dr. Paul Wileyto for helpful discussions.

REFERENCES
