Original Contribution

Contribution of Seasonality in Transmission of *Mycobacterium tuberculosis* to Seasonality in Tuberculosis Disease: A Simulation Study

Lucia C. Soetens*, Hendriek C. Boshuizen, and Hester Korthals Altes

* Correspondence to Lucia C. Soetens, National Institute for Public Health and the Environment (RIVM), P.O. Box 1, 3720 BA Bilthoven, the Netherlands (e-mail: loes.soetens@rivm.nl).

Initially submitted October 23, 2012; accepted for publication May 16, 2013.

A seasonal rise in tuberculosis (TB) notifications has been confirmed in several studies. Here, we examined one hypothesis for its cause: increased transmission of TB during wintertime due to crowding. Seasonality analysis was performed on actual and simulated notifications of clustered TB cases, which are considered to be representative of recent transmission, diagnosed from 1993 to 2004 in the Netherlands (n = 4,746). To test the hypothesis of winter crowding, notifications were simulated by adding patient delay and incubation period to an infection date randomly taken to be in winter in 80% of cases. The incubation periods were derived from frequency distributions for different TB disease localizations drawn from the literature. Seasonality analysis was performed using autocorrelation function plots and spectral analysis. Actual notifications showed strong seasonality in clustered TB and clustered extrapulmonary TB cases but not in clustered pulmonary TB cases. Analysis of simulated notifications revealed barely significant seasonality only in extrapulmonary TB cases. Our results suggest that increased transmission of TB during wintertime is unlikely to be the only cause of the seasonal peak in TB notifications. A factor closer to the notification date probably contributes to the seasonality observed in TB notifications.

Fourier analysis; incubation period; infectious diseases; molecular typing; seasons; tuberculosis; tuberculosis transmission

Abbreviations: EPTB, extrapulmonary tuberculosis; PTB, pulmonary tuberculosis; TB, tuberculosis.

Over the past decades, widespread research regarding the seasonality of tuberculosis (TB) has been conducted. A rather consistent seasonal pattern has been shown across different regions of the world. TB incidence shows a peak in spring (1–4) and a trough in autumn (2, 5–8). TB occupies a special position in infectious diseases with a seasonal pattern, having an incubation and generation time on the same order of magnitude as the duration of the seasonal cycle. Several mechanisms have been proposed to explain the peak in TB notifications in spring.

Three main causes may account for seasonality in notifications of an infectious disease: seasonality in reporting, seasonality in development of disease, and seasonality in time of infection. All of these factors may contribute to the seasonal pattern observed for *Mycobacterium tuberculosis* disease.

First, seasonality in reporting may contribute to local patterns of notification as in the study by Naranbat et al. (2), but it is an unlikely overall factor since seasonality in TB is observed in different countries and parts of the world. Second, seasonality in the development of disease may contribute to TB seasonality through lowered host immunity (9–12). Impaired immunity, resulting from vitamin D deficiency or influenza-like illnesses at the end of winter, would favor development of TB disease as well as reactivation. Third, seasonality in time of infection may be caused by an increased transmission during wintertime. This is hypothesized to be caused by indoor crowding (2, 3, 6). Arguments against this hypothesis state that the incubation period of TB appears to be too variable to lead to a seasonal pattern in onset of TB disease (13).

To our knowledge, no study has tried to assess the role of seasonality of transmission in TB seasonality. In the current work, we examined whether increased transmission during wintertime can result in a seasonal rise in TB notifications. To that end, we analyzed seasonality patterns in clustered pulmonary...
tuberculosis (PTB)/extrapulmonary tuberculosis (EPTB) cases from the Netherlands Tuberculosis Register (1993–2008), limiting our study to clustered TB cases because it is generally assumed that they represent recent transmission (14, 15). We simulated a situation in which most of the transmission occurs in winter; we then derived the incubation period from distributions in the available literature, as well as patient delay from the Netherlands Tuberculosis Register. We then analyzed the seasonality patterns in the simulated notification dates. The question we tried to answer with this simulation was whether the observed pattern of seasonality in notifications can be reproduced by assuming that 80% of the yearly transmission occurs in winter, using the knowledge on incubation period from the literature.

MATERIALS AND METHODS

Data

In the Netherlands, TB is a notifiable disease. The Netherlands Tuberculosis Register, held by the Royal Netherlands Tuberculosis Foundation until 2012 (now by the National Institute for Public Health and the Environment), has recorded data on patient characteristics, risk factors, and treatment outcomes since 1993. Data on age, sex, country of birth, localization of TB disease, and patient delay of cases registered during the period 1993–2004 were used in this study.

All M. tuberculosis complex isolates of patients diagnosed in the Netherlands are sent to the Tuberculosis Reference Laboratory of the National Institute for Public Health and the Environment for species identification, DNA fingerprinting, and drug-susceptibility testing. During the period 1993–2004, DNA fingerprinting was performed through standard restriction fragment length polymorphism typing with IS6110 as a probe. Subtyping, using the polymorphic guanine-cytosine–rich sequence probe, was performed when standard typing resulted in fewer than 5 IS6110 copies. If patients showed an identical DNA fingerprint and were less than 2 years apart, they were marked as clustered. We used data on clustering and culture sample date for all patients diagnosed from 1993 to 2004.

Data on TB notifications after 2004 were available from both the Netherlands Tuberculosis Register and the Tuberculosis Reference Laboratory. However, a different patient delay registry method was used after 2004, so we excluded patients registered after 2004.

The data in the 2 databases were matched by sex, date of birth, postal area code, and year of diagnosis. In this study, seasonality of (recent) transmission was examined. Therefore, we included only clustered TB cases, since clustering is taken as a marker for recent transmission. Exclusion criteria were unknown TB main localization and diagnosis of both PTB and EPTB.

Overview methods

Before we started the simulation study, we investigated whether the seasonality pattern in notifications of TB disease described earlier (13) was also present in the clustered TB cases. We used spectral analysis, as described below, to assess seasonality. We then wished to reproduce a possible seasonal pattern through a simulation study. We introduced seasonality in transmission, where most infections occurred during winter: 80% of the patients in a given year were infected in December, January, or February of the year prior to the notification year, while the remaining 20% were infected between March and November. A sensitivity analysis was performed on the proportion of patients infected in winter (“high-transmission period”) and the length of this period. Subsequently, we estimated an incubation period for every patient that depended on his/her TB disease localization and imputed patient delay (time between onset of symptoms and first physician visit) if the data on delay were missing. Both of these periods were added to the simulated infection date to obtain a simulated notification date for every patient. We performed seasonality analysis on these simulated notification dates to establish whether seasonality in transmission can lead to seasonality in notifications.

Calculation of incubation period and patient delay

Incubation period. Since the exact time of infection is mostly unknown in TB patients, it is hard to determine the exact length of the incubation period. Extrapulmonary disease localization is a strong risk factor for a short incubation period (16). Therefore, we computed the incubation time according to disease localization. In the Netherlands Tuberculosis Register, TB disease localization falls into 9 categories: “primary TB,” “respiratory TB (pleural TB),” “pulmonary TB,” “TB meningitis, “miliary TB,” “intestinal TB,” “urogenital TB,” “bone/joint TB,” and “TB in other organs” (mainly lymph-node TB). To obtain estimates for incubation time per disease localization, we used clinical data from 2 studies (17, 18). The study by Poulsen et al. (18) was conducted on a small island, allowing for reconstruction of the transmission network, thereby giving an estimate of the time between infection and disease manifestation. Wallgren (17) closely followed the clinical signs typical of the different TB disease stages in his patients and derived the incubation period from these clinical signs. For patients diagnosed with primary TB, pleural TB, TB meningitis, pulmonary TB, or bone/joint TB, this estimation was done 100 times by randomly drawing 100 values from frequency distributions described in the studies mentioned above. A more detailed description of this procedure is given in Web Appendix 1 (available at http://aje.oxfordjournals.org/), part 1.

No distributions are available, however, for the incubation periods of intestinal/urogenital TB (also known as abdominal TB), miliary TB, and TB in other organs. We assumed that the frequency distribution of the incubation period of miliary TB was the same as that for TB meningitis, based on a statement by Wallgren (17). For patients with abdominal TB and TB in other organs, the average incubation period was derived using the overall median incubation period of 1.26 years (16) (Web Appendix 1, part 2). We performed a sensitivity analysis for this assumption using 3 different scenarios on the ratio between incubation periods of abdominal TB and TB in other organs (Web Appendix 1, part 2).

Patient delay. To limit data loss, we imputed missing patient delay values using the Multivariate Imputation by Chained Equations (MICE) technique (19), available in the MICE package of R, version 2.15.0 (R Foundation for Statistical Computing, Vienna, Austria). The variables age, sex,
ethnicity, diagnosis, TB localization, clustering status, and year of diagnosis were used as predictors for missing patient delays. The imputed values were assessed for convergence. Furthermore, we plotted the densities of both observed and imputed values to check whether imputed values were reasonable.

Seasonality analysis

The seasonality of both observed and simulated notifications of TB disease in clustered TB cases was studied. First, autocorrelation between notifications at different time lags was calculated using the autocorrelation function. Second, we performed a Fourier spectral analysis. This method, based on the underlying sine and cosine functions of a time series, has been described extensively by Cancho-Candela et al. (20). Spectral analysis produces periodograms in which the spectral density is determined for every frequency. The spectral density at a certain frequency quantifies the contribution of that frequency to the overall periodic behavior of the time series. To confirm a yearly pattern, the spectral density at a frequency of 1/12 months$^{-1}$ should be significantly greater than the spectral density at the other frequencies. Significance was calculated through a permutation test (Web Appendix 2).

All analyses were carried out using R statistical software, version 2.15.0.

Pooling results

The random draws of the incubation period distributions resulted in 100 new data sets. In addition, the sensitivity analysis of the incubation period was performed on 3 different scenarios.

We pooled the $P$ values associated with the outcome of spectral analysis by considering them to be an estimate derived from an imputation. We used the standard pooling rules, which amounts to averaging the $P$ values. The standard pooling rules assume that values to be pooled come from a normal distribution; however, $P$ values come from a uniform distribution under the null hypothesis. Therefore, we tested this procedure in a simulation study and observed that it performed well despite the violation of this assumption (data not shown).

RESULTS

Study population

Matching patient data from the Netherlands Tuberculosis Register to DNA fingerprint data from the laboratory of the National Institute for Public Health and the Environment yielded a database with a total of 10,299 culture-positive TB patients diagnosed during the period 1993–2004 (Figure 1).
Of these 10,299 patients, only the clustered cases with a known TB main localization and a diagnosis of PTB or EPTB were included in this study (n = 4,746 (46.1%)) (Figure 1). Table 1 gives the basic characteristics of the study population.

### Seasonality in actual notifications

Figures 2 and 3 show the time series and periodograms of clustered TB, clustered EPTB, and clustered PTB notifications for 1993–2004. Seasonality was significant in all clustered cases (spectrum value $f_{1/12} = 367.13; P < 0.001$) (Figure 3) and in clustered EPTB cases (spectrum value $f_{1/12} = 193.08; P < 0.001$) (Figure 3), and there was a peak in notifications from March to June (Figure 4). No seasonal pattern was present in clustered PTB cases (spectrum value $f_{1/12} = 47.42; P = 0.169$).

As described in the study by Korthals Altes et al. (13), a great trough in notifications was observed in December (Figure 4). To examine whether the seasonal pattern in all clustered TB and clustered EPTB cases was driven by this trough, we again analyzed the data by means of Fourier spectral analysis with December and January notifications replaced by the averaged consecutive December and January notifications. This analysis yielded no different results; the seasonal pattern in clustered TB and clustered EPTB cases remained significant (data not shown), which implied that the observed seasonality was not solely driven by the winter trough.

### Patient delay and incubation period

Information on patient delay was missing in approximately one-third of all cases (33.8%). We imputed these missing data to limit data loss. Quality control of the imputed data set assessing convergence and density plots of observed and imputed values revealed that the imputed data were realistic. The median patient delay was 4.0 weeks (interquartile range, 1.0–8.0).

The incubation period derived from the literature was shortest for primary TB (1.4 months) and certain manifestations of EPTB disease (4.1 and 6.8 months for TB meningitis/miliary TB and pleural TB, respectively) compared with the incubation period of PTB and bone/joint TB (23.1 and 22.0 months, respectively) (Table 2). Miliary TB was assumed to have the same incubation period distribution as TB meningitis (Web Appendix 1, part 2). Under the assumption of equal incubation periods for abdominal TB and TB in other organs, an incubation period of 30.2 months was estimated for these localizations (Table 2).

### Seasonality in simulated notification date

We simulated that 80% of all transmission takes place in December, January, and February (Figure 5). In Figure 6, the monthly numbers of simulated cases are shown for 5 random simulations. In all cumulated clustered cases and clustered PTB cases, no clear seasonal peak was observed. In clustered EPTB cases, a slight elevation was seen around June. To confirm this observation, seasonality was tested with autocorrelation function plots and spectral analysis. Autocorrelation function plots of the simulated notifications revealed no clear seasonal pattern (Web Appendix 3). Results of the spectral analysis are presented in Figure 7, displaying 5 randomly picked simulations from 100 simulations. Periodograms showed no significant peak at the frequency $1/12$ months$^{-1}$ ($P_{\text{pooled}} = 0.57$) for all clustered TB cases, implying that seasonality was not demonstrated for the simulated notifications of clustered TB cases. Furthermore, seasonality could not be established in clustered PTB cases ($P_{\text{pooled}} = 0.18$). In simulated notifications of clustered EPTB cases, however, a significant peak was confirmed at the frequency $1/12$ months$^{-1}$ ($P_{\text{pooled}} = 0.03$).

Sensitivity analysis, examining the effects of varying incubation period of TB in other organs, showed similar results in the seasonality analysis of the simulated notifications for the different scenarios (Web Appendix 4). Therefore, results of the seasonality analysis in this paper are given only for the most probable scenario, which assumes that the incubation periods are similar for abdominal TB and TB in other organs (Web Appendix 1, part 2). Sensitivity analysis on the proportion of patients infected during the high-transmission period and length of this period showed that seasonality in notifications depends on both the proportion of patients infected in winter and the length of the high-transmission period. Seasonality in simulated EPTB notifications is lost when the proportion of patients infected is lower and the length of the high-transmission period increases (Web Appendix 5). Seasonality in simulated PTB and overall TB notifications remains unaffected.

### DISCUSSION

In our study, seasonality of notifications with a peak from March to June was confirmed in clustered TB cases and clustered EPTB cases but not in clustered PTB cases. This finding
closely follows the observation by Korthals Altes et al. (13) that a seasonal peak is confirmed in EPTB cases but not in PTB cases. We tried to reproduce this seasonality by simulating enhanced transmission in winter and deriving notification time series using incubation time estimates with pooled data from clinical studies (17, 18). Incubation period distributions were strongly skewed to the right for most TB disease localizations, with an overall shorter incubation period for primary TB and certain manifestations of EPTB disease compared with the incubation period of PTB and bone/joint TB. We did reproduce the seasonality patterns in actual notifications of clustered EPTB and PTB cases when simulating seasonal transmission. No pattern was detected for clustered PTB notifications, but a seasonal pattern was detected for clustered EPTB notifications. We failed to reproduce the seasonal pattern seen in actual notifications of all clustered TB cases when assuming increased winter transmission. This means that a strong seasonal pattern in transmission is detected only in the notification dates of clustered EPTB cases.

This study had several limitations. First and most notably, the method used for establishing the incubation periods for the different TB localizations was based on limited knowledge,

![Figure 2. Time series of all clustered (solid line), clustered extrapulmonary (dashed line), and clustered pulmonary (dotted line) tuberculosis notifications in the Netherlands, 1993–2004.](image)

![Figure 3. Periodograms of all clustered (A), clustered extrapulmonary (B), and clustered pulmonary (C) tuberculosis notifications in the Netherlands, 1993–2004. The dotted line is placed at the frequency $1/12$ months$^{-1}$. P values were $<0.001$, $<0.001$, and 0.169 for clustered, clustered extrapulmonary, and clustered pulmonary tuberculosis cases, respectively.](image)
requiring several assumptions. For example, the incubation period estimates for primary TB, pleural TB, PTB, TB meningitis, and bone/joint TB were based on studies carried out in the 1950s, because there have been no recent studies on this subject. It is possible that incubation periods have changed over time, owing to changing strains and changes in disease management. Moreover, the resolution of the data was sometimes very low (i.e., measured in years), which contributed to the variability in incubation period estimation.

In addition, we made plausible assumptions about the incubation period of abdominal TB, miliary TB, and TB in other organs. For example, intestinal TB and urogenital TB were assumed to have the same incubation period distribution because they can both be classified as abdominal TB, but strong support for this and several other assumptions is lacking. However, this simplification will not influence the results greatly because only 4% of all TB cases are classified as abdominal

<table>
<thead>
<tr>
<th>TB Localization</th>
<th>No. of Persons</th>
<th>Average Incubation Period (µi), months</th>
<th>95% CI for µi</th>
<th>Weighting Factor (ai)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary TB</td>
<td>225</td>
<td>1.4</td>
<td>0.9, 1.8</td>
<td>0.02</td>
</tr>
<tr>
<td>Pulmonary TB</td>
<td>5,999</td>
<td>23.1</td>
<td>7.8, 58.3</td>
<td>0.63</td>
</tr>
<tr>
<td>Pleurisy/respiratory tract TB</td>
<td>857</td>
<td>6.8</td>
<td>2.9, 13.1</td>
<td>0.09</td>
</tr>
<tr>
<td>Bone/joint TB</td>
<td>383</td>
<td>22.0</td>
<td>8.2, 47.3</td>
<td>0.04</td>
</tr>
<tr>
<td>Meningitis/miliary TB</td>
<td>319</td>
<td>4.1</td>
<td>2.1, 7.4</td>
<td>0.03</td>
</tr>
<tr>
<td>Abdominal TB</td>
<td>416</td>
<td>30.2</td>
<td>7.6, 89.5</td>
<td>0.04</td>
</tr>
<tr>
<td>TB in other organs</td>
<td>1,345</td>
<td>30.2</td>
<td>7.6, 89.5</td>
<td>0.14</td>
</tr>
<tr>
<td>Subtotal</td>
<td>9,544</td>
<td></td>
<td></td>
<td>1.00</td>
</tr>
<tr>
<td>Unknown</td>
<td>755</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>10,299</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; TB, tuberculosis.

a An average incubation period and weighting factor were used.
b Number of cases in the study population before nonclustered cases and cases with unknown TB localization were excluded.
c Average incubation period was based on the scenario in which it is assumed that the incubation period of abdominal TB is similar to the incubation period of TB in other organs (Web Appendix 5).

Table 2. Average Estimated Incubation Period and Frequency of Cases by Tuberculosis Disease Localization in the Study Population, the Netherlands, 1993–2004a
TB (Table 2). Similarly, we assumed that miliary TB has the same incubation period distribution as TB meningitis, but only 3% of all cases are classified as TB meningitis/miliary TB (Table 2). TB in other organs, mostly lymph node TB with an unknown incubation period, accounts for a greater fraction of all TB cases (14%; Table 2). However, the sensitivity study examining the effects of varying incubation periods for TB in other organs yielded no different results.

Second, the peak in simulated notifications of clustered EPTB cases does not synchronize with the peak in actual clustered EPTB notifications. The simulated peak in July–August–September (Figure 6) does not match the actual peak in March–April–June–July (Figure 4). There are 2 explanations for this mismatch. First, clustered cases, as defined by the Netherlands Tuberculosis Register, have been diagnosed less than 2 years apart. In contrast, we estimated incubation periods based on studies spanning a much longer study period (15 years in the study by Borgdorff et al. (16)). This might overestimate the incubation period relating to the clustered cases. A second explanation for this mismatch could be that not elevated transmission in winter but rather another factor, such as seasonal immunosuppression, causes the seasonal pattern seen in actual TB notifications.

A third limitation is the doubtful validity of the patient delay data. We noticed that many patients reported a delay of 4 or 8 weeks, whereas delays such as 3 or 9 weeks were reported less often. This is quite logical given that a very plausible answer to the question “When did you first have complaints?” would be “around a month ago” or “around 2 months ago.” However, the impact of inaccuracy in patient delay will probably be negligible, considering the length of the incubation period (approximately an 8-week-patient delay vs. an average incubation period of 1.8 years) (16).

Finally, we included all clustered cases in our study and therefore also included source cases, which might not be linked to recent transmission. This will only be a problem for clustered PTB cases, since clustered EPTB cases are all secondary cases because EPTB is not considered infectious (21). Because the seasonal pattern in clustered notifications is driven by the seasonality in EPTB cases, this is not a real problem. It is possible that a seasonal pattern in recent PTB infections is obscured by the inclusion of index cases that might be remote infections. However, the simulated PTB notifications (assuming they would be recent infections) failed to show seasonality and thus left our conclusion unchanged.

The limitations of this study uncover some gaps in knowledge and possible targets for further research.

The first target would be more research on the incubation period of various forms of TB. We are aware of the fact that this research is difficult because the incubation period is long. However, molecular epidemiologic studies on epidemiologically linked clustered cases (16) could help. If such a study could be carried out in a country where various forms of EPTB are more prevalent, more precise estimates of the incubation period of different TB disease localizations could be obtained.

Furthermore, this study showed that increased transmission during wintertime, although it may well occur, cannot explain the seasonal peak in TB notifications, as was speculated by Leung et al. (22). This finding is in line with our observation that clustered and unique EPTB cases have an equally strong seasonal pattern, suggesting that transmission has little influence on TB seasonality (23). Therefore, an additional or other factor, such as seasonality of disease onset or reporting, contributes to TB seasonality. We think it is important to further investigate the underlying cause of seasonality in TB disease because it will improve our understanding of the host-pathogen interaction and be useful in the context of disease control (24). An interesting approach would be to incorporate the elements of seasonal forcing in a dynamic model (e.g., as done with measles by Conlan and Grenfell (25)) to investigate the possible roles of each element.

To summarize, when 80% of transmission occurs in winter, a seasonal peak is seen in EPTB cases but not in PTB cases or TB cases overall. Two findings stand out. First, this peak in clustered EPTB cases does not exactly coincide with the actual peak in clustered EPTB cases, when we assume that transmission...
should be highest in December–February (the coldest months of the year in the Northern Hemisphere). Second, winter crowding does not lead to a seasonal peak in clustered TB cases, while notifications do peak seasonally. These findings suggest 2 things. First, winter crowding would explain EPTB seasonality, but the timing of the peak might be off because we might have overestimated incubation time. Second, winter crowding, if it does contribute to seasonality, does not cause a pattern strong enough to lead to a seasonal peak in clustered TB cases. Thus, an additional factor must be involved to explain this peak. This suggests that a factor closer to the notification date is likely to cause the seasonal pattern observed in TB notifications. Whether it is at the level of disease development, seasonal vitamin D insufficiency/decreased immunity, or reporting remains to be investigated.

ACKNOWLEDGMENTS

Author affiliations: Epidemiology and Surveillance Unit, National Institute for Public Health and the Environment, Bilthoven, the Netherlands (Lucia Soetens, Hester Korthals Altes); and Department of Statistics, Mathematical Modeling and Data Logistics, National Institute for Public Health and the Environment, Bilthoven, the Netherlands (Hendriek Boshuizen).

We thank the members of the Royal Netherlands Tuberculosis Foundation, especially Henrieke Schimmel (now at the National Institute for Public Health and the Environment), for their contribution to data collection and for providing the data. We also thank Dr. Ronald Geskus for valuable information on incubation times for different tuberculosis disease localizations. Finally, we thank Dr. Jacco Wallinga and Rolf Ypma for fruitful discussions.

Conflict of interest: none declared.

REFERENCES