

Empirical Evidence for Synchrony in the Evolution of TB Cases and HIV+ Contacts among the San Francisco Homeless

MOJDEH MOHTASHEMI^{1, 2*} and L. MASAE KAWAMURA³

¹MITRE Corporation, 7515 Colshire Dr., McLean, VA, 22102 USA; ²MIT Computer
Science and AI Laboratory, 77 Mass. Ave., Cambridge, MA 02139-4307 USA; ³TB
Control Section, San Francisco Dept. of Public Health, Ward 94, 1001 Potrero Ave., San
Francisco, CA 94110 USA

* To whom correspondence should be addressed. E-mail: mojdeh@mit.edu

**There are no prior or duplicate publications of this manuscript and no financial
conflicts of interest. This manuscript has been approved by all authors.**

Background

The reemergence of tuberculosis (TB) in the mid 1980s in many parts of the world, including the United States, is often attributed to the emergence and rapid spread of human immunodeficiency virus (HIV) and acquired immunodeficiency syndrome (AIDS). Although it is well established that TB transmission is particularly amplified in populations with high HIV prevalence, the epidemiology of interaction between TB and HIV is not well understood. This is partly due to the scarcity of HIV related data, a consequent of the voluntary nature of HIV status reporting and testing, and partly due to current practices of screening high risk populations through separate surveillance programs for HIV and TB. The San Francisco Department of Public Health, TB Clinic Program, has been conducting active surveillance among the San Francisco high risk populations since the early 1990s. We present extensive TB surveillance data on HIV and TB infection among the San Francisco homeless to investigate the association and synchrony between the TB cases and their HIV+ contacts.

Methods and Findings

We used wavelet coherence and phase analyses applied to TB surveillance data from January 1993 through December 2005, to establish and quantify statistical association and synchrony in the highly non-stationary and ostensibly non-periodic waves of HIV and TB infection in San Francisco. When stratified by homelessness, and similar analyses were performed on data from the homeless and non-homeless populations in San Francisco, we found that the evolution of TB cases and their HIV+ contacts is highly coherent over time and locked in phase at a specific periodic scale among the San Francisco homeless, but no significant association was observed for the non-homeless. Finally, to verify that contact tracing alone does not account for the observed dynamics, wavelet analysis was applied to other data types on the homeless population.

Conclusions

This study confirms the hypothesis that the dynamics of HIV and TB are significantly intertwined and that HIV is likely a key factor in the sustenance of TB transmission among the San Francisco homeless. The findings of this study underscore the importance of contact

tracing in detection of HIV+ individuals that may otherwise remain undetected, and thus highlights the ever increasing need for HIV related data and an integrative approach to monitoring high risk populations with respect to HIV and TB transmission.

Introduction

San Francisco has the highest rate of TB in the United States, nearly three to four times the national rate (1). While North America is among the lowest risk regions of the world for TB transmission (2, 3), in some areas of San Francisco the TB rate is as high as the developing world with rates ranging from 100 to 200 per 100,000 (1). After decades of decline in the TB incidence due to advances in health and medicine in the mid 20th century, San Francisco, as many parts of the world, witnessed a gradual reemergence of TB in the mid 1980s. This reemergence has been attributed to a number of risk factors including increased homelessness and prevalence of HIV and AIDS (4). In the early 1990s, the San Francisco Department of Public Health (SFDPH), TB Control Program (TBCP), adopted new and intensified control measures to significantly reduce the transmission of *Mycobacterium tuberculosis* in San Francisco (5). While the number of TB cases has been steadily declining since the early 1990s in the San Francisco general population, exclusive of the homeless (non-homeless) (Figure 1A), the rate of decline has been much slower and more variable for the homeless population, in spite of extensive active surveillance among the homeless by the SFDPH TBCP (Figure 1B). Although this disparity in trends has been mostly attributed to the prevalence of HIV and AIDS among the indigent population in San Francisco (6, 7), and many impoverished regions of the

world (8-17), few studies have been properly designed and documented to establish a direct epidemiological link between TB and HIV.

In a nation-wide study, exclusive of California, the Centers for Disease Control and Prevention (CDC) reported that 9% of the TB patients were also HIV⁺ in 2005 (18). For the San Francisco general population, a recent study using molecular genotyping estimated that 13.7% of TB cases from 1991 to 2002 were attributed to HIV infection (19). For the San Francisco homeless, in general, and the TB infected homeless, in particular, studies pertaining to the prevalence of HIV are scarce. In a cross sectional study of the San Francisco inner city shelters and free meal programs, the HIV seroprevalence was estimated at 8.5% (95% Confidence Interval (CI): 7%-10%) among the homeless (6). The authors found that 19% of the study subjects who were seropositive for HIV were co-infected with TB. In a recent study among the San Francisco urban homeless and marginally housed, the HIV seroprevalence was estimated to be five times higher among the indigent young adults than those of the San Francisco general population (7). The overall HIV seroprevalence was estimated at 10.5% and 29.6% among men who had sex with men. To date, we are not aware of any documented study on the HIV prevalence, or seroprevalence, among the TB infected homeless population.

The objective of this study is three-fold. First, using extensive TB surveillance data in San Francisco we demonstrate that the prevalence of HIV infection is likely higher for the TB infected homeless population than those reported for the general population of the United States, the San Francisco general population, and the San Francisco homeless.

Second, using wavelet time series analysis of the homeless TB cases and their HIV+ contacts, we show that the dynamics of the two infectious diseases are significantly intertwined and synchronized at a specific periodic band, implying the importance of contact tracing in detection of unknown HIV+ contacts, and a likely indicator of epidemiological interaction between TB and HIV among the San Francisco homeless. Third, we argue that these findings have direct implications for the HIV and TB surveillance; they highlight the importance of HIV related data and call for an integrative approach to the TB and HIV surveillance in high risk populations.

Methods

Data

The dataset for this study consists of comprehensive information on individuals that have been diagnosed by the SFDPH TBCP with active TB (TB cases), and their reported contacts (PPD+ and PPD-)¹, all identified from January 1993 to December 2005. For the homeless population, there were a total of 387 TB cases and 613 reported contacts in the data. For the non-homeless, there were a total of 2,329 TB cases and 2,003 reported contacts. The TBCP routinely collects comprehensive information on TB cases including demographics (e.g., age, gender, race), individuals they have been in contact or exposed to (contacts), duration of exposure to each reported contact, personal risk factors (e.g., intravenous drug use, HIV status, alcohol intake), laboratory results (e.g., PPD skin test,

¹ PPD+ contacts refer to individuals with positive purified protein derivative (PPD) skin test for TB indicating exposure to the *Mycobacterium tuberculosis*. Likewise, PPD- contacts are individuals with negative TB skin test.

chest x-ray), date of diagnosis, and primary residence. Contact tracing is a key component of active and ongoing surveillance by the SFDPH TBCP. TB patients are brought to the TB clinic as they are identified. Contact tracing follows immediately the identification and diagnosis of a TB case. Reported contacts of each TB patient are then tracked and brought to the TB clinic for medical and laboratory examinations and interviews. During these interviews, the SFDPH TBCP documents wide-ranging information including personal, demographic, risk factors (e.g., HIV status), and exposure related information. The SFDPH TBCP providers determine the HIV status of the TB cases and their contacts from their medical records. If the HIV status cannot be determined, the SFDPH TBCP will offer HIV testing. Approval was obtained from the SFDPH TBCP to use its routinely collected TB surveillance data before any analysis was performed.

To examine the association between TB and HIV infection among the San Francisco homeless, contact network data comprised of monthly number of TB cases and their reported contacts who are HIV+ but PPD- (HIV+/PPD-) were retrieved for each population. Because we seek to determine the significance of interdependency in the dynamics of HIV and TB infection, PPD+ contacts were excluded from the analysis to reduce bias in statistical correlation.

Wavelet Coherence and Phase Analysis

We used wavelet time series analysis (20-24) to investigate the relationship between the evolution of TB cases and their reported HIV+/PPD- contacts. In this study, each time

series undergoing wavelet transform was logarithmically transformed after adding a constant of one to each value in the series. The log transformed series were then standardized to have zero mean and unit variance before wavelet analysis was performed. To determine the statistical association or correlation between the spectra of a pair of time series (24), we performed wavelet coherence analysis. The resulting coherence values can identify regions of high correlation between a pair of time series at various periodic components. The phase difference between the two time series was then computed to quantify the statistical tendency for the two signals to be phase locked, and determine the temporal lag between the two (25, 26), if any. Statistical significance of the results were assessed using bootstrap methods (27) similar to those used in (28, 29). For a complete mathematical treatment of the wavelet coherence and phase analysis see Appendix A.

Wavelet coherence and phase analysis have proven to be effective analytic tools for studying the association and synchrony in the evolution of time series data with periodic components. In particular, wavelet phase analysis has been used to study the spatio-temporal patterns of movement of measles epidemics from large cities to smaller towns in the pre- and post-vaccination era in England and Wales (28), and the relationship and spatial synchrony between the dengue incidence in Bangkok and the rest of Thailand (29). It has also been used to analyze weekly state-specific excess mortality rate from pneumonia and influenza (P&I) from the lower 49 contiguous states in the United States over the past 30 years (30); the association and synchrony between cholera incidence, Indian Oscillation Index (IOI), and rainfall across five West African countries (31); and

the association and synchrony between the Influenza Like Illnesses (ILI) consultation rates and laboratory surveillance data (32).

A major theme common in aforementioned investigations is that their data, while non-stationary, comprise strong periodic components including those inflicted by the effect of seasonality. In contrast, the TB and HIV time series data documented by the SFDPH TBCP are highly non-stationary and do not comprise any apparent periodic component (Figure 4A).

Results

Trends in HIV Reporting

While the prevalence, or seroprevalence, of HIV infection among the TB infected homeless population in San Francisco remains unknown, it is widely believed to be higher than that of the general population of the United States, exclusive of California (18), that of the San Francisco general population (19), or even that of the San Francisco homeless (6, 7). Indeed, a count of the TB cases co-infected with HIV, derived from the SFDPH TBCP data, strongly supports this hypothesis and reveals contrasting trends between the homeless and non-homeless populations of San Francisco. In the SFDPH TBCP data spanning the 13-year period from January 1993 to December 2005, 37% of the homeless TB cases are also HIV+ (Table 1). The actual rate, however, is likely higher, since 17% of the TB cases in the data have unknown HIV status (Table 1). During the same period, only 10% of the non-homeless TB cases were reported as HIV+, while 54% of them had missing HIV status (Table 1). In 2005, on average, about 10% of

all TB cases (homeless and non-homeless) in the SFDPH TBCP data were reported as HIV+ $\left(\frac{13}{131} \approx 0.1\right)$ (see last row in Table 1), which is much closer to the average value for the nation, exclusive of California, as reported by the CDC for that year (18). When stratified by homelessness, however, 44% of the homeless TB cases are reported as HIV+ versus 4% for the non-homeless, an 11-fold difference (see last row in Table 1), suggesting that the average value is a misleading measure of disease frequency.

At the same time, the SFDPH TBCP data reveals disparity in trends of the HIV status reporting between the homeless and non-homeless. Because testing or reporting of the HIV status is not mandatory, missing information pertaining to the HIV status of TB cases and their contacts is a limiting factor for investigating the impact of HIV on TB or the association between them. According to the CDC, the HIV status reporting among TB patients has improved in the United States general population, exclusive of California (18). However, when the SFDPH TBCP data are stratified according to each subpopulation, contrasting dynamics are observed. While the HIV status reporting has improved among the homeless TB cases (Figure 2A and Table 1), it has remained relatively constant and at low levels (<50%) among the non-homeless (Figure 2B and Table 1). At the same time, the proportion of the homeless TB cases reported as HIV+ has increased either due to improved reporting, rise in the prevalence of HIV infection, or both (Figure S1A and Table 1), but declined for the non-homeless (Figure S1B and Table 1).

A similar trend is observed for the reported contacts of the TB cases. While the HIV status reporting has improved among the reported homeless contacts (Figure 3A), it has worsened among the non-homeless (Figure 3C). The same is true when contacts are stratified according to the PPD skin test results. The HIV status reporting has generally improved for the PPD- homeless contacts, particularly after 1999 (Figure 3B), but the converse is observed for the PPD- non-homeless contacts (Figure 3D). Similar trends are observed for PPD+ contacts in the homeless and non-homeless populations (Figure S2). While it is unclear why the two populations manifest such contrasting behavior, these findings, along with those in Figure 1, suggest that HIV is likely a complicating factor in controlling the TB transmission among the homeless, and that the HIV related data for the homeless have come to more reliably capture the underlying population risk behavior and disease frequency among the homeless than that for the non-homeless.

Association and Synchrony between HIV and TB

While the existence of a periodic component is not visibly detectable from the time series plots of the TB cases and their HIV+/PPD- contacts (Figure 4A), wavelet coherence analysis of the two reveals otherwise. Figure 4B is the wavelet coherence plot of the TB cases and their HIV+/PPD- contacts over the 13-year period in the data examined at various periodic scales (3-36 months on the y axis). The coherence between the two time series appears to be strongest at the 27-31 months periodic band and almost consistently significant over the entire 13-year period (see patches of significant signals indicated by black lines along the stretch of red band). This suggests that the two time series are synchronized and locked in phase approximately every 27-31 months. We performed

similar analysis for the non-homeless TB cases and their HIV+/PPD- contacts to demonstrate the dichotomy in the dynamics of the two populations. **Figure 5** is the coherence plot of the non-homeless TB cases and their HIV+/PPD- contacts, where no evidence of consistently significant coherence at a periodic component is observed between the two time series.

To quantify the synchrony and temporal lag between the two time series, phase evolution of each time series was computed from its wavelet transform at the 27-31 months periodic scale (see Methods and Appendix A). **Figure 4C** demonstrates the resulting phase evolution of the homeless TB cases and their HIV+/PPD- contacts over the 13-year period in the data. The phase difference of the two time series was then computed by **Equation (3)** in Appendix A and is plotted in **Figure 4D**. The phase difference (in months) represents the lag time between the waves of the two time series at the 27-31 months periodic component where they appear to be highly correlated and synchronized. The mean lag time between the TB cases and their HIV+/PPD- contacts is 5.4 months for the entire 13-year period. As can be seen in **Figure 3B**, the four-year period from January 2002 to December 2005 is the longest period in the SFDPH TBCP data with the least amount of missing HIV related information for the PPD- contacts. Specifically, there are only two PPD- contacts with unknown HIV status: one in November 2002 and one in July 2003. The mean lag time between the homeless TB cases and their HIV+/PPD- contacts for this four-year period, where HIV related information is most complete, was estimated at 4.4 months with waves of TB infection following those of the HIV+/PPD-.

Missing Data Imputation: Verification of Robustness

To investigate the effect of missing HIV related data on the statistical significance of association and synchrony between the TB cases and their HIV+/PPD- contacts, a randomized procedure was applied to the data as follows. Each PPD- contact with missing HIV status in the SFDPH TBCP data was randomly assigned a status (HIV+ or HIV-), and coherence analysis was performed. This experiment was repeated 1000 times and the average coherence was recorded. The result is demonstrated in [Figure 6](#), where it can be clearly seen that the key periodic component of [Figure 4](#) is preserved. In spite of randomized missing data imputation, the two time series manifest high coherence at the 27-31 periodic scale with the average lag time of 5.2 months for the entire 13-year period, a finding quite similar to that of the original data with missing HIV data ([Figure 4](#)). To verify that the observed robustness of the results is unique to the homeless and not an outcome of chance, we applied the same randomization procedure to the TB surveillance data for the non-homeless. The result is illustrated in [Figure 7](#), where there is no evidence of coherence at any periodic scale over the entire period. While this may suggest that HIV infection is likely not a significant factor in sustaining the TB transmission among the San Francisco non-homeless, together these results with those of previous section suggest that the TB surveillance data, in spite of missing HIV related information, is sufficiently representative of the underlying disease frequency among the homeless.

Discussion

Using extensive TB surveillance data on the San Francisco homeless, we found that the evolution of TB cases and their HIV+/PPD- contacts are significantly interdependent, synchronized at about every 2.5 years and separated by the mean lag time of 5.4 months.

What generates the observed association and synchrony in the two time series? The SFDPH TBCP begins active contact tracing soon after the detection of new TB cases, which can take a few months before all contacts are traced and their risk factors are identified. Thus, it would seem reasonable to suggest that the co-evolution of TB cases and their HIV+/PPD- contacts is likely a confounding effect of contact tracing. If contact tracing is to explain the joint dynamics between the TB cases and their HIV+/PPD- contacts, then surely similar dynamics must exist between the TB cases and their PPD- contacts, albeit with a shorter lag time since it may take longer to identify the HIV status of all contacts. To investigate, we applied similar wavelet coherence and phase analyses to the data on the homeless TB cases and their PPD- contacts, the same subpopulation from which data on HIV+ individuals were extracted. Figure S4 illustrates both time series (Figure S4-A) and the outcome of their coherence analysis (Figure S4-B). As can be clearly seen, there is no evidence of stable coherence and synchrony over time between the TB cases and their PPD- contacts. Thus, contact tracing alone is not sufficient to rationalize the observed joint dynamics between the TB cases and their HIV+/PPD- contacts.

What other processes can potentially explain these trends? Although it is well established that HIV plays a key role in the sustenance and amplification of TB transmission in high

risk populations, the epidemiology of their interaction is not well understood (17). Clearly, any significant epidemiological interaction must manifest itself, among other mathematical representations, through high degree of coherence and synchrony at some periodic scale between TB and HIV within a population (e.g., homeless). Although we demonstrated significant coherence and synchrony between the homeless TB cases and their HIV+/PPD- contacts, comparable independent HIV data on the homeless population through HIV/AIDS surveillance are required to provide sound evidence of such an interaction. Thus, while we demonstrated that the co-dependency between the evolution of TB cases and their HIV+/PPD- contacts is significant, unique to the homeless, not an outcome of chance, and contact tracing cannot fully account for these trends, it remains an open problem as to what epidemiological mechanisms generate and govern the joint dynamics of HIV and TB among the homeless.

Thus, a limitation of this study is in the absence of independent data on homelessness, irrespective of the HIV or TB status, and HIV related data on the homeless population, irrespective of their TB status. Recall that the primary role of the SFDPH TBCP is to conduct active surveillance with respect to TB, and not HIV, in San Francisco. As a result we sought to obtain HIV data independently collected on the homeless population through HIV/AIDS surveillance. However, the extent and completeness of the HIV-related data through the TB surveillance were not matched by similar data obtained from direct HIV surveillance by the AIDS office in San Francisco. This is likely because active HIV/AIDS surveillance is mostly targeted at other high risk populations for HIV transmission, such as men having sex with men, and not at the indigent populations with

poor access to health care and equally or more severely affected by HIV and AIDS. Similarly, we were not able to obtain data on homelessness, including the size and time of influx of new individuals into the homeless population. Joint analyses of these data coupled with those from TB surveillance should greatly enhance understanding of the multi-scale epidemiological interaction between TB and HIV, leading to adoption of more effective control measures and intervention policies.

Two important findings of this study are worthy of note. First, while the prevalence, or seroprevalence, of HIV infection among the TB infected homeless population in San Francisco remains unknown, using the SFPDH TBSP data we demonstrated that it must be much higher than that of the general population of the United States, exclusive of California (18), that of the San Francisco general population (19), and even that of the San Francisco homeless (6, 7). Second, we demonstrated that the HIV status reporting has improved markedly among the San Francisco homeless compared to the non-homeless (Table 1 and Figures 2, 3, S1, S2), and the U.S. general population, exclusive of California (18). The observed coherence and synchrony together with these two findings further substantiate the premise that HIV is likely a key factor in the sustenance of TB transmission among the San Francisco homeless.

Finally, the findings of this study have clear implications for public health policy. First, the voluntary nature of testing and reporting of HIV status presents a challenge to the surveillance and control of HIV and TB infection in high risk populations. Our study highlights the importance of HIV related data for investigating the association between HIV and TB infection in high risk populations, and reconfirms the need for simplifying

the current procedures for obtaining written consent for HIV testing (33). Second, high prevalence of HIV infection among the San Francisco homeless, together with the observed association and synchrony between the TB cases and their HIV+/PPD- contacts, underscore the need to revisit the current reductionist approach to public health surveillance. Traditionally, public health departments are divided into distinct disease surveillance programs. However, separate surveillance programs cannot effectively address problems that arise due to interactions between multiple diseases in vulnerable populations. The strong interdependency in the epidemiology of TB cases and their HIV+ contacts among the San Francisco homeless, where the dynamics of the two were shown to be significantly intertwined, embodies this assertion. This puts forward the notion that to effectively reduce the TB transmission among the San Francisco homeless, the TB surveillance efforts would have to be coupled with those of HIV.

Appendix A

Wavelet Time Series Analysis

Wavelet transform achieves good localization in both time and frequency by decomposing the signal into high- and low-frequency components. This is accomplished by contracting or dilating the analysis window, which is a complex or real function (wavelet) (20-24). Each log transformed and standardized time series (see Methods) is wavelet transformed:

$$W_s(a, \tau) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} s(t) \psi^* \left(\frac{t - \tau}{a} \right) dt \quad (1)$$

Here $s(t)$ is the original time series, parameters a and τ represent the scale (or period) and time shift respectively, $\psi(t)$ is the wavelet chosen for the study, and $\psi^*(t)$ is the complex conjugate form of the wavelet. We chose the Morlet wavelet defined as $\psi(t) = \pi^{-1/4} \exp(-2i\omega_0 t) \exp(-t^2/2)$ for our study, which is a complex exponential. The local wavelet power spectrum of a signal is then $\|W_s(a, \tau)\|^2$ where ‘ $\| \ \|$ ’ denotes the magnitude of the wavelet coefficients evaluated at the scale a , and the time shift τ . The statistical association or correlation between the spectra of two time series can be quantified using the wavelet coherence (24):

$$R_{s_1, s_2}(a, \tau) = \frac{\|W_{s_1, s_2}(a, \tau)\|}{\|W_{s_1, s_1}(a, \tau)\|^{1/2} \|W_{s_2, s_2}(a, \tau)\|^{1/2}} \quad (2)$$

The numerator $W_{s_1, s_2}(a, \tau) = W_{s_1}(a, \tau) W_{s_2}^*(a, \tau)$ represents the cross-spectrum between two time series s_1, s_2 . Thus the wavelet coherence is the cross-spectrum normalized by the

product of the local power spectrum of each time series. Because the Morlet wavelet is a complex function, and thus the resulting wavelet coefficients $W_s(a, \tau)$ are complex numbers, the phase of a wavelet transformed time series can be written as the inverse tangent of the ratio of the imaginary part over the real part of $W_s(a, \tau)$ (25). Analogously, the phase difference between two time series Φ_{s_1, s_2} can be interpreted as the ratio of the imaginary part of the cross-spectrum of the two time series, $W_{s_1, s_2}(a, \tau)$, over the real part of $W_{s_1, s_2}(a, \tau)$. Thus, the phase difference indicates the statistical tendency for the two signals to be phase locked:

$$\Phi_{s_1, s_2}(a, \tau) = \arctan \frac{\operatorname{Im} \left(W_{s_1, s_2}(a, \tau) \right)}{\operatorname{Re} \left(W_{s_1, s_2}(a, \tau) \right)} \quad (3)$$

In **Equation (3)**, $\operatorname{Im} \left(W_{s_1, s_2}(a, \tau) \right)$ is the imaginary part, and $\operatorname{Re} \left(W_{s_1, s_2}(a, \tau) \right)$ is the real part of the cross-spectrum between the two series. At and near the beginning and end of a time series, there is tendency for loss in statistical power, a phenomenon referred to as the edge effects (26). The cone of influence, depicted as a parabola in **Figures 4B, 5, 6, and 7**, captures this effect, where the statistical power in the area outside the parabola should be interpreted with caution. Both time series were padded with zeros up to the next highest power of two to reduce the bias due to the edge effects.

To determine whether the observed association between the two time series is significant (alternate hypothesis) or random and due to chance (null hypothesis), its statistical

significance must be assessed. In this study, we used bootstrap methods (27) similar to those used in (28, 29). For each time series, a control series was constructed by permuting the original series while maintaining the statistical properties of the series. Then each similarly constructed control series underwent wavelet analysis and corresponding values were recorded. This procedure was repeated 400 times. Finally, the results of the original series were compared against the 95th quantile of the resulting distributions.

References and Notes

1. San Francisco Department of Public (SFDPH), "Profile of Tuberculosis in San Francisco 2007, Epidemiology Report" (2007).
2. World Health Organization (WHO), "Changing History (World Health Report 2004)," Geneva, Switzerland (2004).
3. World Health Organization (WHO), "Global Tuberculosis Control: Surveillance, Planning, Financing (World Health Report 2006)," Geneva, Switzerland (2006).
4. Institute of Medicine (IOM), "Emerging Infections: Microbial Threats to Health in the United States". Washington, D.C.: National Academy Press (1992).
5. M. R. Jasmer *et al.*, *Ann. Intern. Med.* **130**, 971 (1999).
6. A. R. Zolopa *et al.*, *JAMA* **272**(6), 455 (1994).
7. M. J. Robertson *et al.*, *Am. J. Public Health* **94** (7), 1207 (2004).
8. A. Bermejo, H. Veeken, A. Berra, *AIDS* **6**, 1203 (1992).
9. C. Dye, G. P. Garnett, K. Sellman, B. G. Williams, *The Lancet* **352**, 1886 (1998).
10. M. Badri, D. Wilson, R. Wood, *The Lancet* **359**, 2059 (2002).
11. E. L. Corbett *et al.*, *Arch Intern. Med.* **163**, 1009 (2003).
12. B. G. Williams, C. Dye, *Science* **301**, 1535 (2003).
13. S. D. Lawn, M. Badri, R. Wood, *AIDS* **19**, 2109 (2005).
14. S. D. Lawn, R. Wood, *AIDS* **20**, 1787 (2006).
15. R. Wood *et al.*, *Am. J. Respir. Crit. Care Med.* **175**, 87 (2007).
16. E. L. Corbett *et al.*, *PLoS Med.* **4**(1) ; e22. doi:10.1371/journal.pmed.0040022 (2007).
17. B. G. Williams, D. Maher, *Am. J. Respir. Crit. Care Med.* **175**, 6 (2007).
18. Centers for Disease Control and Prevention (CDC), Morbidity and Mortality Weekly Report (MMWR) **56**, 42 (2007).
19. K. DeRiemer, L. M. Kawamura, P. C. Hopewell, C. L. Daley, *Am. J. Respir. Crit. Care Med.* **176**, 936 (2007).
20. D. B. Percival, A. T. Walden, *Wavelet Methods for Time Series Analysis* (Cambridge University Press, Cambridge, 2000).
21. I. Daubechies, *Ten Lectures on Wavelets* (Society for Industrial and Applied Mathematics, Philadelphia, 1992).

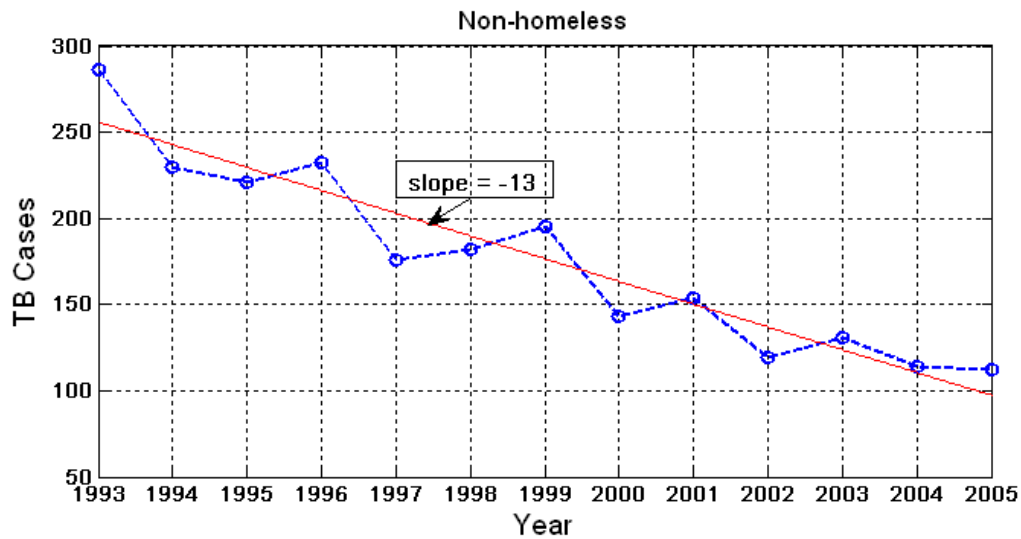
22. 4. C. Torrence, G. P. Compo, *Bull. Am. Meteorol. Soc. Lond.* **79**, 61-78 (1998).
23. 5. G. P. Nason, R. von Sachs, *Phil. Trans. R. Soc. Lond. A.* **357**, 2511-2526 (1999).
24. J. R. Chatfield, *The analysis of Time Series: An Introduction* (Chapman and Hall, London, UK, 1989).
25. M. Le Van Quyen *et al.*, *J. Neurosci. Methods* **111**, 83-98 (2001). doi:10.1016/S0165-0270(01)00372-7.
26. A. S. Pikovsky *et al.*, *Synchronization : A Universal Concept in Nonlinear Sciences* (Cambridge University Press, Cambridge, UK, 2001).
27. B. Efron, R. J. Tibshirani, *An Introduction to the Bootstrap* (Chapman and Hall/CRC, London, UK, 1993).
28. B. T. Grenfell, O. N. Bjornstad, J. Kappey, *Nature* **414**, 716 (2001).
29. B. Cazelles, M. Chavez, G. C. de Magny, J.-F. Guegan, S. Hales, *J. R. Soc. Interface* (2006). doi :10.1098/rsif.2007.0212.
30. C. Viboud, O. N. Bjornstad, D. L. Smith, L. Simonsen, M. A. Miller, B. T. Grenfell, *Science* **312**, 447 (2006).
31. G. C. de Magny, J.-F. Guegan, M. Petit, B. Cazelles, *BMC Infectious Diseases* **7**(20) (2007). doi :10.1186/1471-2334-7-20.
32. L. Yang *et al.*, *PLoS ONE* **3**(1) : e1399. doi :10.1371/journal.pone.0001399 (2008).
33. N. M. Zetola, C. G. Grijalva, S. Gertler *et al.* *PLoS ONE* **3** (7), e2591 (2008).

Acknowledgements

The authors would like to thank Alexander Yeh and Jennifer Grinsdale for their assistance with data management.

FIGURE 1 (A-B)

A



B

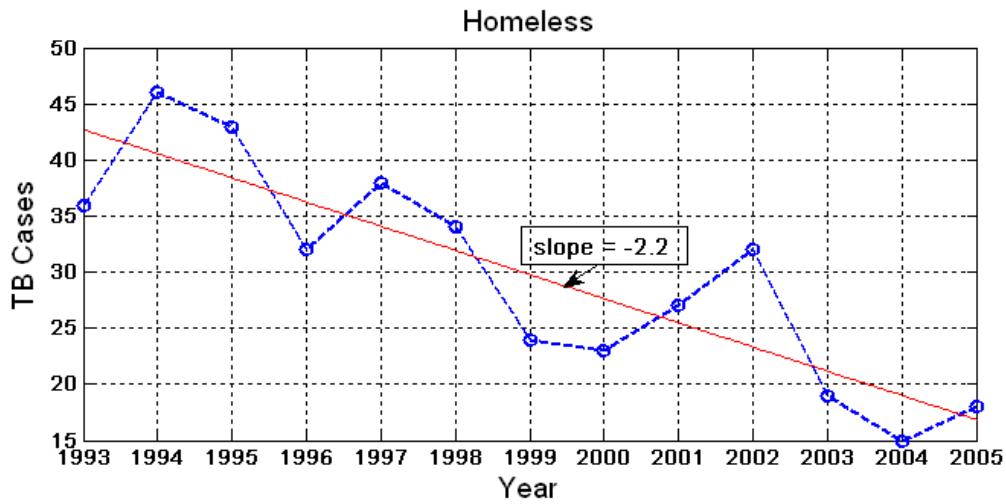
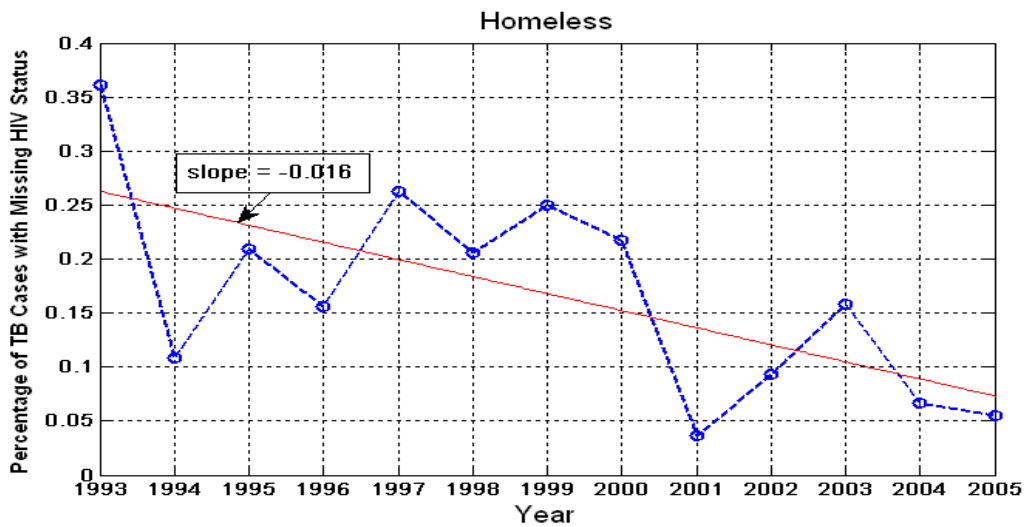


Figure 1. Annual number of TB cases in San Francisco, January 1993-December 2005. **(A)** Non-homeless: general population, exclusive of homeless. **(B)** Homeless. The rate of decline over time in the number of TB cases is much faster for the non-homeless than that of the homeless. Red line: linear fit.

FIGURE 2 (A-B)

A



B

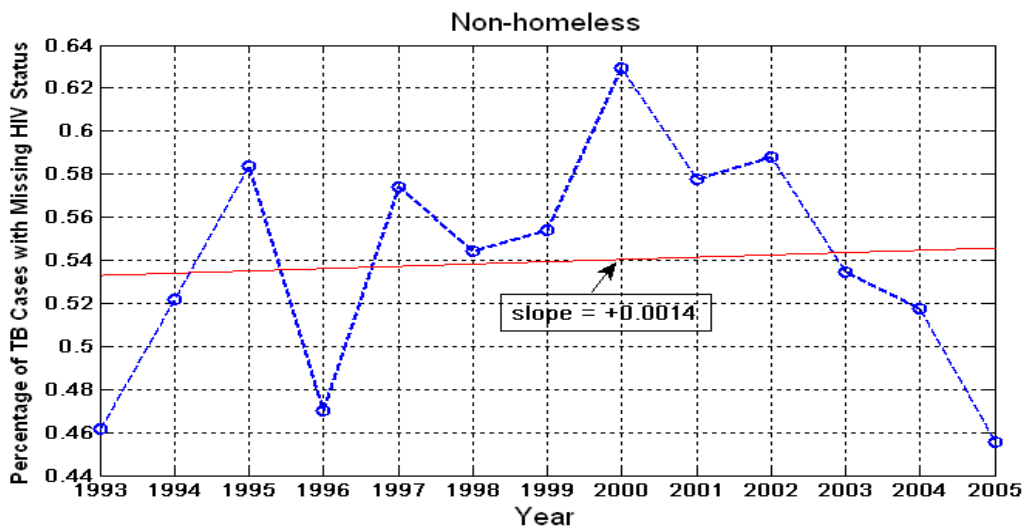
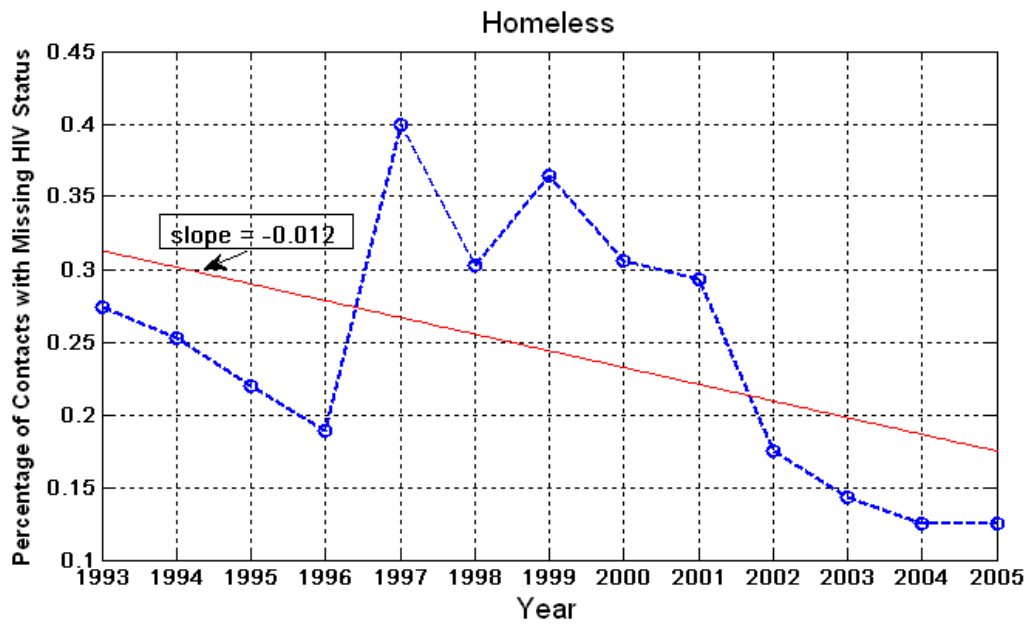


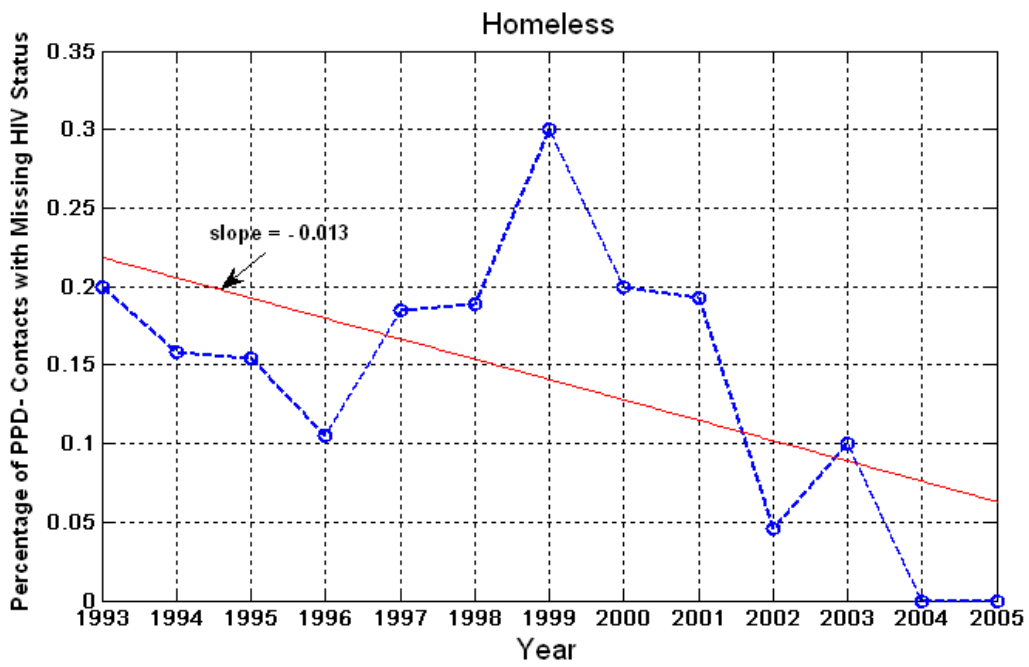
Figure 2. Annual percentage of TB cases with missing HIV status, January 1993-December 2005. **(A)** Homeless. **(B)** Non-homeless. While the reporting of the HIV status has improved over time among the homeless TB cases (negative slope), it has remained more or less constant among the non-homeless TB cases (small positive slope), with a visible improvement after 2002. Red line: linear fit.

FIGURE 3 (A-D)

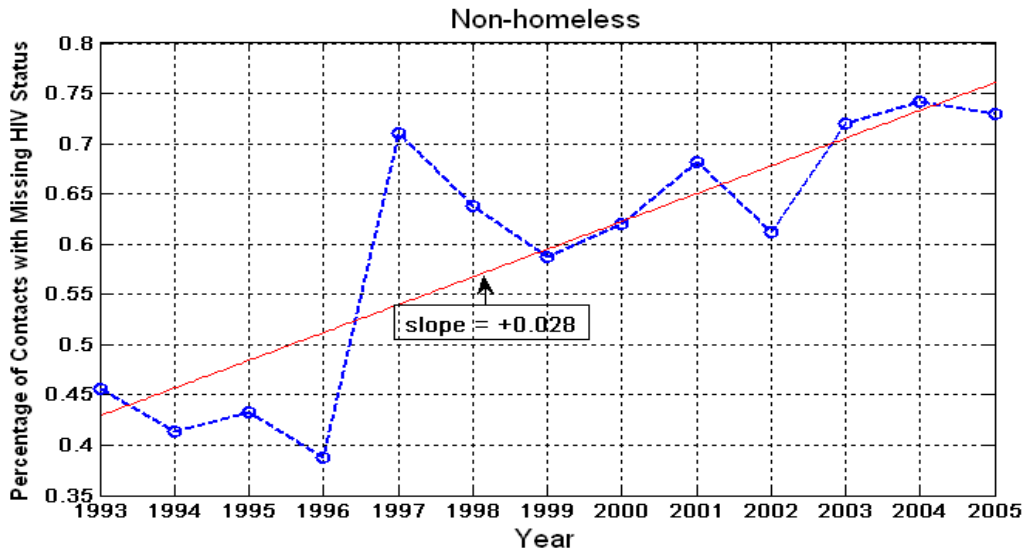
A



B



C



D

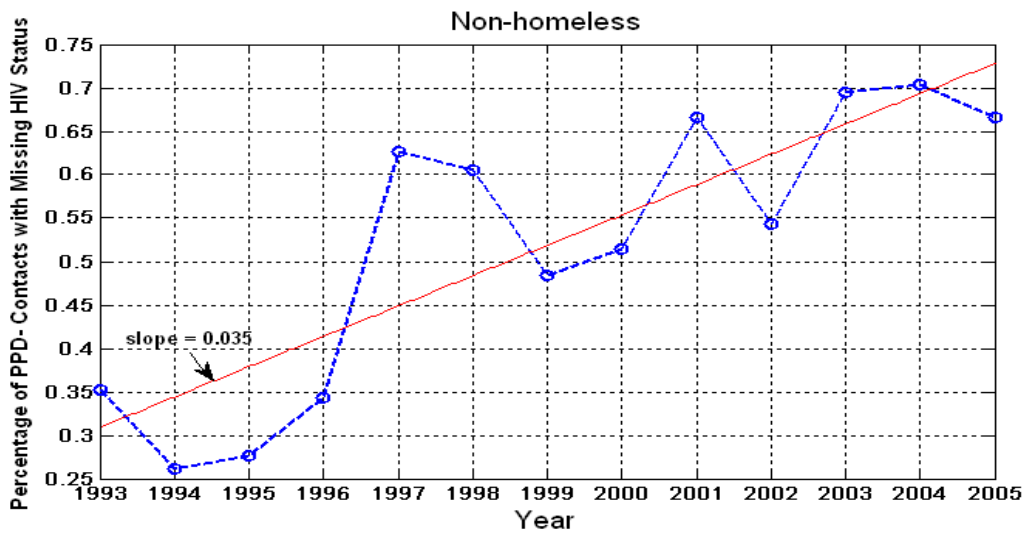
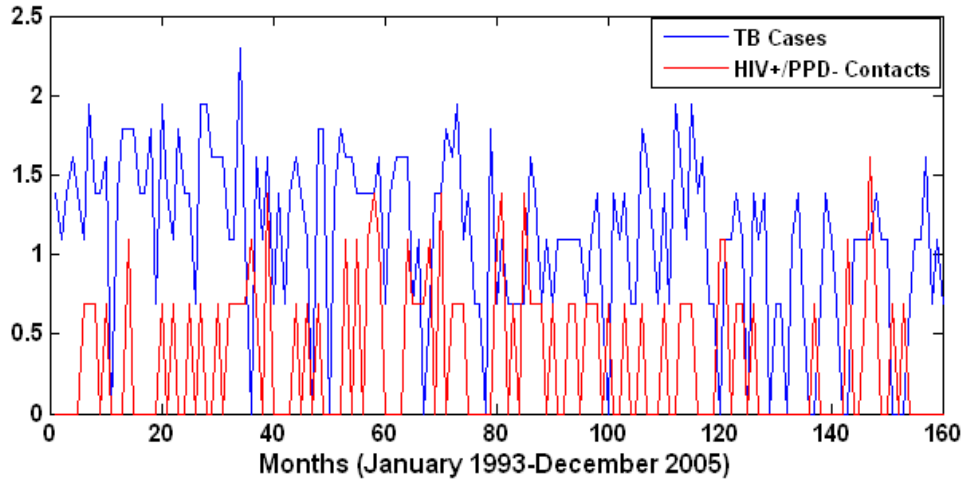


Figure 3. Annual percentage of all contacts and PPD- contacts with missing HIV status, January 1993-December 2005. **(A and B)** Homeless. **(C and D)** Non-homeless. In **A** and **C**, contacts include both PPD+ and PPD- individuals. While the reporting of the HIV status has improved over time among all contacts and PPD- contacts in the homeless

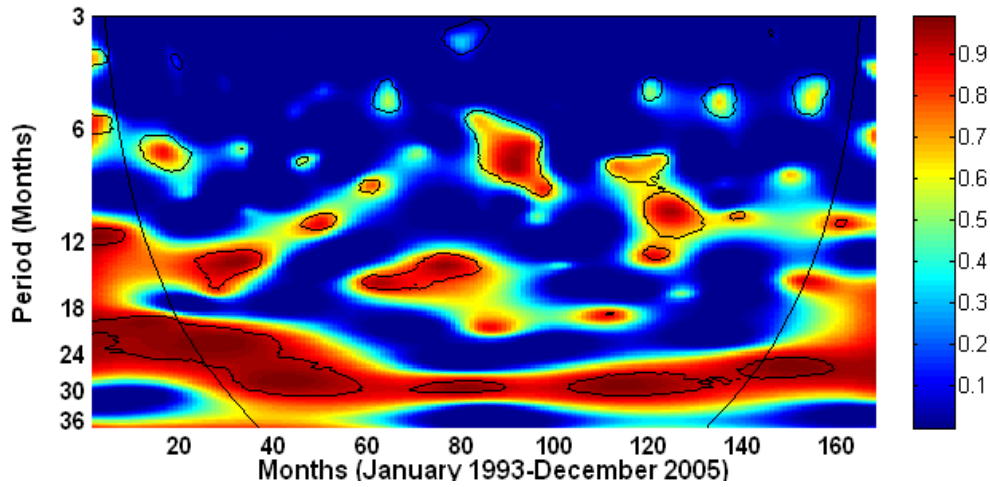
population (negative slopes), it has worsened among the non-homeless (positive slopes). Red line: linear fit.

FIGURE 4 (A-D)

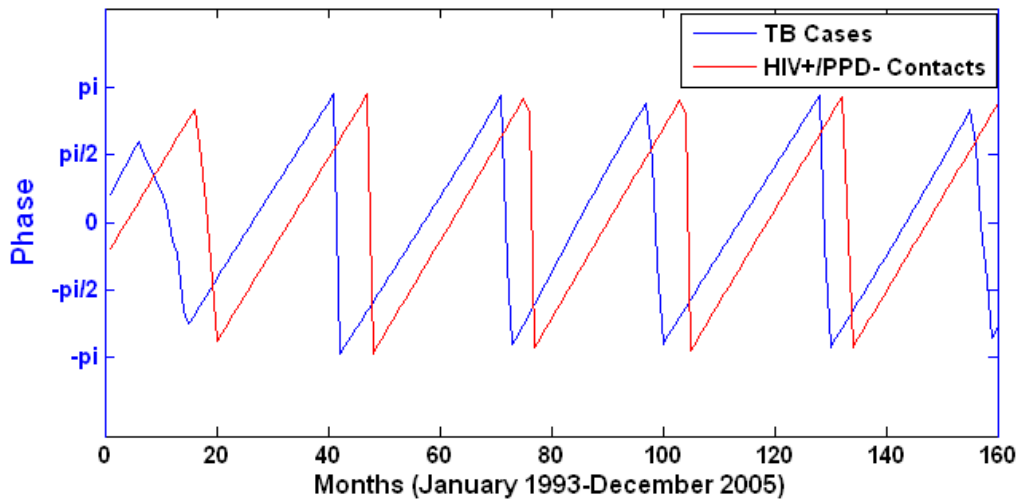
A



B



C



D

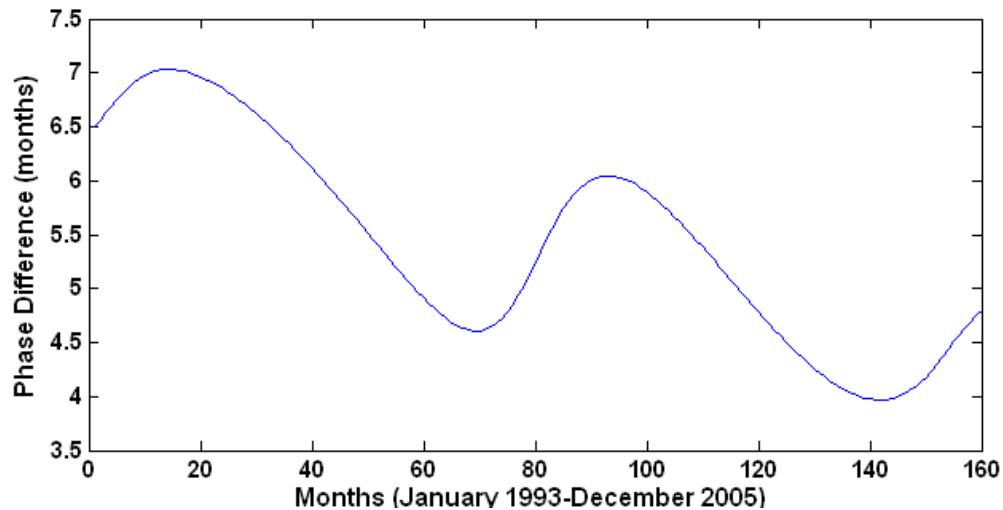


Figure 4. Wavelet time series analysis of the homeless TB cases and their HIV+/ PPD-contacts, January 1993-December 2005. Both time series were log-transformed and standardized (mean centered with unit variance) before wavelet analyses were performed. **(A)** Log-transformed time series of the TB cases and their HIV+/PPD-contacts. **(B)** Wavelet coherence between the homeless TB cases and their HIV+/PPD-

contacts. One periodic component at the 27-31 months is significantly present with the $\alpha=5\%$ significant level (black lines), computed based on 400 bootstrapped series. The colors code for power values from dark blue, representing low values, to dark red, representing high values; the superimposed parabola is the cone of influence which measures the extent of edge effects. **(C)** Wavelet phase evolution of the homeless TB cases and their HIV+/PPD- contacts computed at the at the 27-31 months periodic band in radians. **(D)** Wavelet phase difference of the phases of TB cases and their HIV+/PPD- contacts computed at the 27-31 months periodic band. Waves of TB cases and their HIV+ contacts are separated with a mean lag time of 5.4 months for the entire 13-year period, and with a mean lag time of 4.4 months for the four-year period from January 2002 to December 2005 when the HIV related data is most complete.

FIGURE 5

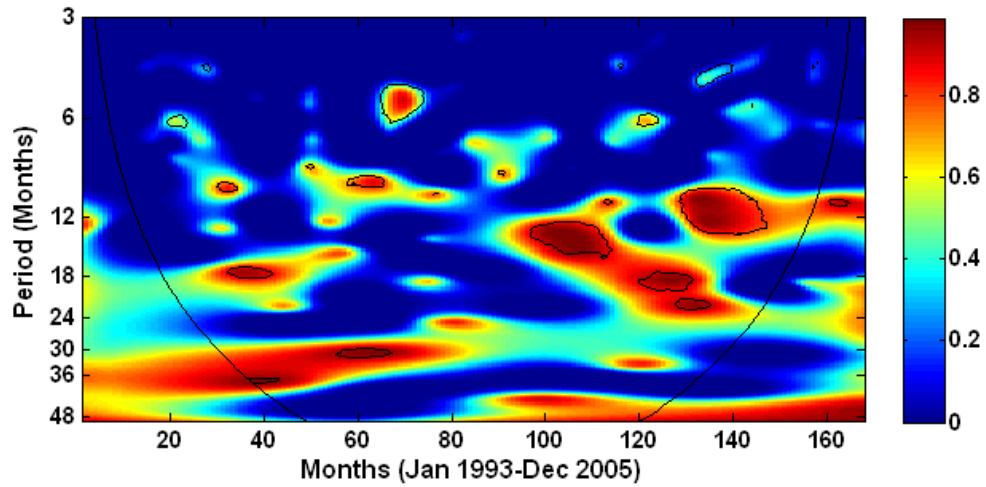


Figure 5. Wavelet coherence between the non-homeless TB cases and their HIV+/PPD- contacts, January 1993-December 2005. Except for patches of high coherence at various periodic scales appearing transiently in time, there is no evidence of consistently significant coherence and synchrony.

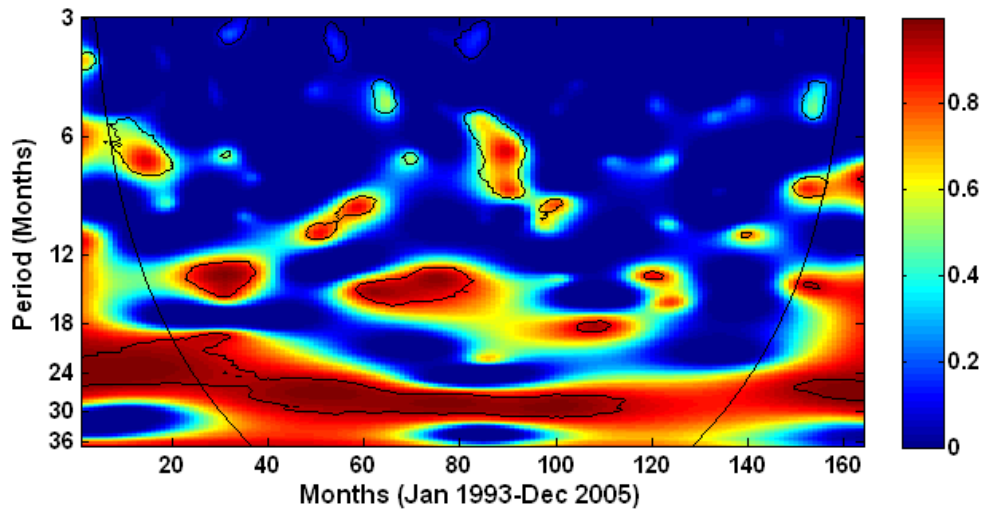
FIGURE 6

Figure 6. Robustness of coherence and synchrony between the homeless TB cases and their HIV+/PPD- contacts

Wavelet coherence analysis was performed, where each PPD- contact with missing HIV status was randomly assigned a status (HIV+ or HIV-). The experiment was repeated 1000 times and the average coherence was recorded. The data spans the 13-year period from January 1993 to December 2005. Both time series were log-transformed and standardized before wavelet analysis was performed. The two time series manifest high coherence at the same 27-31 months periodic band, as in Figure 4B, with the lag time of 5.2 months. The lag time for the period from January 2002 to December 2005 is 4.8 months.

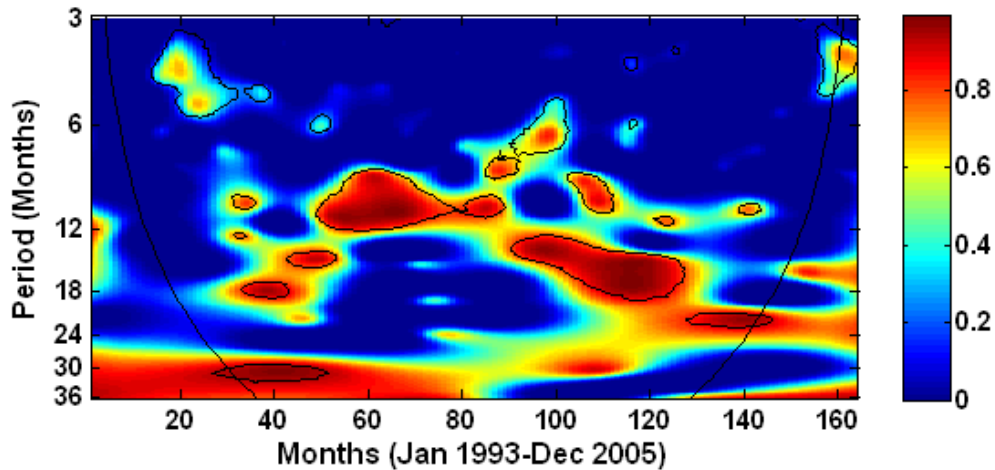
FIGURE 7

Figure 7. Verification of robustness for TB and HIV coherence and synchrony. Wavelet coherence analysis was performed between the non-homeless TB cases and their HIV+/PPD- contacts where each PPD- contact with missing HIV status was randomly assigned a status (HIV+ or HIV-). The experiment was repeated 10000 times and the average coherence was recorded. The data spans the period from January 1993 to December 2005. Both time series were log-transformed and standardized before wavelet analyses were performed. As in Figure 5, there is no evidence of a consistently significant periodic component in spite of missing HIV data imputation.

Table 1. Number of TB cases, HIV+ TB cases, and TB cases with missing HIV status, in the San Francisco homeless and non-homeless populations. Over time, HIV status reporting has improved among the homeless while it has remained at low levels, and almost constant, among the non-homeless (third column under each population category). At the same time, the percentage of HIV+ cases has increased among the homeless, but it has slightly declined among the non-homeless (second column under each population category). To compare against the CDC estimates for 2005 (18), the related numbers from the SFDPH TBCP data are also provided for that year.

	Homeless			Non-homeless		
	TB cases	HIV+	Unknown HIV Status	TB cases	HIV+	Unknown HIV Status
1993-2005	387	144 (37%)	69 (17%)	2329	233 (10%)	1269 (54%)
1996-2005	262	95 (36%)	42 (16%)	1579	122 (7.7%)	859 (54%)
1999-2005	158	68 (43%)	20 (12%)	982	63 (6.4%)	548 (55.8%)
2002-2005	84	36 (43%)	8 (9.5%)	478	33 (7%)	250 (53%)
2005	18	8 (44%)	1 (5%)	113	5 (4%)	54 (48%)