Project MUSE®
http://muse.jhu.edu

Epidemic Cycles in Agricultural Populations: A Cross-Cultural Study

L. MADRIGAL[1] AND T. KOERTVELYESSY[2]

Abstract A cross-cultural analysis of mortality patterns is of interest to biological anthropologists and genetic epidemiologists. In this paper, we examine four agricultural populations from Costa Rica, Hungary, and the United States in order to determine if they suffered from a cyclical distribution of epidemics. When possible, we look at the mortality time series of adults and children separately. Of the12 series, only 2 show significant epidemic cycles. Both are in the Hungarian groups and both affect subadults. Otherwise, the Costa Rica, U.S., and adult series of the Hungarian groups do not show any periodicity of mortality peaks. Our results indicate that epidemic cycles are not as ubiquitous in small agricultural groups as the literature would suggest.

In his book *Health and the Rise of Civilization*, Cohen (1989) notes the many factors that lead to an increase in infectious disease exposure once humans adopt a settled, agricultural lifestyle. Agricultural groups, characterized by sedentism, close proximity among potential hosts, and frequent exposure to human and animal fecal matter, show a rise in the frequency of infections. Some infections reach epidemic levels at which they kill a large number of hosts.

As Cohen (1989) states, a pattern of repeated epidemics was unlikely to have occurred in preagricultural, mobile human populations. Presumably, agricultural populations experience epidemic periodicities because after they suffer an epidemic, they accumulate a large enough number of potential hosts through the cycle's period. Scott et al. (1998) also indicate that malnutrition cycles in agricultural groups might be intrinsically correlated with mortality cycles: undernourished individuals are not able to withstand the insult of infectious diseases. The link between an agricultural lifestyle and the rise of epidemic cycles has been stressed even in works from the popular press (McNeill 1979; Karlen 1995). It should be noted that we are interested in the presence of multiyear cycles, instead of cycles within a year. The latter usually result from other causes, such as the seasonal change in pathogens associated with seasonal temperature and rain fluctuations (Madrigal 1994).

Cycles of epidemics have been detected in populations such as the Aland

[1]Department of Anthropology, University of South Florida, Tampa, FL 33620.
[2]Department of Anthropology, Ohio University, Athens, OH 45701.

Human Biology, June 2003, v. 75, no. 3, pp. 345–354.
Copyright © 2003 Wayne State University Press, Detroit, Michigan 48201-1309

KEY WORDS: EPIDEMIC CYCLES, HISTORICAL DEMOGRAPHY, SPECTRAL ANALYSIS

Islands (Mielke 1982; Mielke et al. 1984; Mielke and Pitkänen 1989); Finland (Waris 1991); Britain in general, and London in particular (Duncan et al. 1996, 1997; Schwartz and Marcus 1990); Venezuela (Bouma and Dye 1997); and Denmark (Olsen et al. 1988). Indeed, the identification of pathogen-specific cycles has been in some cases of primary importance to researchers (Scott and Duncan 1998, 2001). The periodicity of epidemics in populations located under dissimilar ecological conditions was the focus of a now-classic paper by Lin and Crawford (1983).

In this paper we evaluate the presence of cycles of epidemics in small agricultural populations in Costa Rica, Hungary, and the United States. Our purpose is to determine whether these groups suffer from epidemic periodicities. Although the raw data were collected differently (some as counts of deaths, some as mortality rates) in the Costa Rican, Hungarian, and U.S. populations, we think that such heterogeneity is a small price to pay to have a cross-cultural test of the hypothesis that agricultural populations suffer epidemic periodicities.

Materials and Methods

The Populations. The sizes and years of the settlements analyzed are shown in Table 1. The data are at http://www.cas.usf.edu/anthropology/faculty/madrigal.html

The Parish of San Miguel de Escazú. The parish of San Miguel de Escazú (Escazú, for short) is 15 km from San José, Costa Rica's capital. Since its foundation in 1799, the Parish has kept excellent vital event certificates. During the 1800s, Catholic priests were appointed civil servants, with the obligation of registering all vital events, even if non-Catholics were involved. In 1864, the settlement size was 2533 (Ministerio de Economia y Hacienda 1964). See Madrigal (2003) for a more in-depth description of the history of Escazú and the records available.

The data consist of the total number of deaths per year from 1851 to 1921. The data were broadly divided into two series, one containing "adults" and the other "children." The priest signing the document frequently made the distinction, but in many instances we have no way of knowing how old the child was. When age was recorded (in most cases after 1860), we follow the suggestion of Scott and Duncan (1998) and include under the category of "child" individuals who are less than 15 years of age.

Acsa and Miklosi, Hungary. Acsa and Miklosi are two villages located in western Hungary, south of Lake Balaton (see Koertvelyessy and Nettleship 1996). Miklosi is a German settlement and Acsa is a Hungarian one. This area of Hungary is often referred to as "Swabian Hungary" because of the large numbers of German settlements. Miklosi was formed by German immigrants at the invitation of local Hungarian landowners. According to the official census records (Sta-

Table 1. Population Sizes and Time Periods Analyzed

Population	*Population Size*	*Period Analyzed*
Escazú, Costa Rica	2533 in 1864	1851–1921
Acsa, Hungary	1149 in 1882	1800–1895
Miklosi, Hungary	679 in 1882	1800–1895
Hungarian settlement, Louisiana (USA)	488 in 1978	1903–1980

tisztikai Hivatal 1882), the sample sizes of Acsa and Miklosi were 1149 and 679, respectively, in 1882.

The Hungarian data consist of mortality rates based on the ecclesiastic records available from the Genealogical Society of Utah (Salt Lake City, Utah) for the period 1800–1895. The data are as follows: infant mortality rate (number of deaths less than 1 year in a given year divided by number of live births in that year multiplied by 1000), neonatal mortality rate (number of deaths less than 1 month in a given year divided by number of live births in that year multiplied by 1000), post-neonatal mortality rate (number of deaths aged 1 month to 1 year in any given year divided by number of live births in that year multiplied by 1000). Adult mortality consists of counts of deaths, where the gender of the deceased was noted.

Hungarian Settlement, Louisiana. The largest rural settlement of Hungarians in the United States is located around the village of Albany, Livingston Parish, Louisiana. This data set differs from the two previous ones in its time frame, which extends from 1904 to 1980. However, the subsistence mode was agrarian for most of the time period, though some individuals worked in a sawmill or other business ventures. Although this group has experienced considerable gene flow, it has remained culturally distinct until recently (see Koertvelyessy 1983). Even though the data are available by gender and age categories, the sample size is so small that all deaths that occurred in one year will be the only variable analyzed, without consideration of gender or age.

Statistical Methodology. Periodogram analysis is ideally suited for the purpose of determining the presence of epidemic cycles of various pathogens. This type of harmonic analysis estimates the percentage of the variance in the time series that is accounted for by each of a set of different cycles. Thus, it allows the researcher to partition how much of the mortality variance is due to cycles of each of the pathogens that affected a population. The negative aspect of periodogram analysis is its subjectivity: the graph of the periodogram is viewed, and the researcher decides which peaks are to be considered as major periodic components. However, the subjectivity of this process can be avoided with a Fisher kappa test, which tests the null hypothesis of no cycles in a manner similar to an analysis of variance: the sum of squares (SS) associated with a cycle (the intensity estimate of each cycle) is divided by the total sum of squares of the overall time series (Warner 1998; Fuller 1976).

Admittedly, the purpose of periodogram analysis is to explore the presence of mortality cycles, more than to test hypotheses, per se. For this reason, it is not uncommon to divide the series by gender, by age groups, and by time periods, and to perform a periodogram analysis for each one. Obviously, in this situation the probability of an overall type I error increases dramatically, even though Fisher's test is conservative. Although we acknowledge the value of exploratory data analysis, we think that it is best to be conservative with the overall type I error of this paper, since we wish to determine if epidemic cycles were present in Escazú for:

1. The entire series
2. Children
3. Adults.

For *each* of the Miklosi and Acsa series, we wish to determine if epidemic cycles were present for:

4. The neonatal series
5. The post-neonatal series
6. The infant series
7. The adult series.

For the Hungarian settlement we analyze only one series:

8. All deaths recorded in one year.

Therefore, we will adjust our significance value by using Bonferroni's correction, where $\alpha_{corrected} = \alpha/T$, where $\alpha = 0.05$ and T is the number of tests run (Morrison 1976). Given that we will run a total of 12 tests, our $\alpha_{corrected} = 0.05/12 = 0.0004$. To find out the exact p values associated with a Fisher's kappa statistic, a short SAS code was written, since the SAS program itself does not provide exact probabilities (LM will be happy to share this code with interested researchers). If a cycle's sum of squares (intensity estimate or SS) is found to be significant at this level, then it is worthwhile to perform a spectral analysis, which reduces the sampling error of the periodogram. However, the periodogram analysis should have first indicated a significant peak of intensity. This approach is more conservative than performing a periodogram and a spectrum analysis for every series.

A further problem that may arise with periodogram analysis is that of "leakage," which occurs when the length of the time series is not an integer multiple of the cycle length. In this case the sum of squares associated with a cycle length close to the real one that cannot be detected is significant. For example, if a series is 83 observations long, and a true cycle of 5 exists, it would not be detected because 83 is not an integer multiple of 5. Thus, the initial periodogram will indicate a peak close to 5. To solve this problem, Warner (1998) suggests that the initial indication of a peak be used as an approximation, and that the length of the time series be shortened so that it becomes a multiple integer of the suggested cycle (from 83 to 80, for example). Leakage is an obvious problem for the detection of mortality cycles, since the length(s) of the cycle is unknown at the beginning of the analysis.

Two assumptions must be met before any time series analysis is attempted: the series must be detrended, and it must have a stationary variance. If not, the peaks detected by the periodogram might actually reflect the violation of either of these assumptions (Jenkins 1979; McCleary and Hay 1980). Jenkins (1979) suggests that a mean*range plot is an excellent way to decide if the series has been rendered stationary with a constant variance. For all series analyzed, we followed his advice and inspected the mean*range plot, although we do not show it because of space considerations. If a trend or a nonstationary variance is detected, then the series is often differenced. First-order differencing consists of subtracting the first observation from the second, the second from the third, and so on. Second-order differencing consists of subtracting the first observation from the third, the second from the fourth, and so on. Even if the time series did not violate either assumption, we differenced them for the purpose of smoothing them, as suggested by Chatfield (1984) and Jenkins (1979).

Another common problem to all time-series analyses of mortality data is the presence of an extreme outlier due to a massive epidemic. In those cases, the outlier masks the presence of any cycles. This problem cannot be smoothed out by differencing or even twice-differencing. In this paper, we follow the suggestion of Lin and Crawford (1983), who "winsorized" the outlier by substituting it with the next highest datum. Winsorization is the substitution of the outlier by the next highest point in the series. It should be noted that winsorization might be needed for high as well as low mortality extremes. Warner (1998) notes that if there is more than one pronounced peak, then it might be better to take the log of the entire series.

The statistical package SAS 8.2 was used for all the analyses. Specifically, the whitest option was requested of Proc Spectra to obtain Fisher's kappa (SAS Institute 2001).

Results

Analysis of the Escazú Data. The Escazú data ($n = 71$) had to be winsorized to address the problem of an extreme outlier in 1856, due to a cholera epidemic. Although the time series did not violate the lack-of-trend assumption, it did suffer from variance-nonstationarity. The data were then first-differenced, and a stable variance around the mean was achieved. Although a plot of the periodogram indicates that there is a peak of cycles at 3.5 years, a Fisher test indicates that the peak is not significant ($g = 0.16$, $p = 0.09$).

An analysis by two broad age categories revealed that the children's time series suffered from many more fluctuations than did the total series. Indeed, second-order differencing was necessary to achieve variance-stationarity and a stable mean. A periodogram of the second difference revealed a peak associated with a period of 2.4 years. This peak was again not significant ($g = 0.18$, $p = 0.04$, ns at $p = 0.004$). In contrast with the children's time series, the adult data did not show any trend or nonstationarity. It was, however, differenced with the purpose of smoothing it (Chatfield 1984). The periodogram analysis suggested a nonsignificant peak at 3.5 years ($g = 0.15$, $p = 0.13$) and is shown in Figure 1.

To address the possibility of "leakage," and having determined that two results hinted to a 3-year cycle, the length of the entire series was shortened to 69 and the same (within a few rounding differences) periodogram intensities (or sums of squares) were obtained. The third result suggested a 2-year cycle, for which leakage was not a possibility, since the original series of 70 is divisible by 2.

Analysis of the Acsa Data. In analyzing the Acsa data ($n = 95$), our first approach was to model the adult series, which only required first-differencing. A periodogram analysis yielded no significant peaks ($g = 0.09$, $p = 0.5$). The infant mortality series, which subsumes within it the neonatal and post-neonatal series, was then analyzed. The infant series needed to be twice-differenced and was winsorized for one year. The periodogram did not yield any significant peak ($g = 0.10$, $p = 0.33$). The neonatal mortality series required only first-differencing and yielded no significant peaks ($g = 0.13$, $p = 0.07$). In contrast, the post-neonatal series (which was only first-differenced) yielded a significant peak at 2 years with $g = 0.23$ ($p = 0.0003$). Leakage is not a problem because $n = 96$, which is divisible by 2. We then followed the periodogram analysis with a power spectrum analysis, which reduces the sampling error of the periodogram. The spectrum confirmed that the period of the cycle was 2 years; its graph is shown in Figure 2.

Analysis of the Miklosi Data. For the Miklosi data ($n = 95$), we first modeled the adult series, which after winsorizing and first-differencing did not yield any significant peaks ($g = 0.08$, $p = 0.7$). The infant mortality series suffered from a very high peak in 1833 and from a low outlier in 1836. We winsorized these peaks with the next highest and lowest values, proceeded to take the first difference, and achieved a stationary variance. No trend was obvious. The periodogram analysis showed a significant peak at 2 years ($g = 0.19$, $p = 0.004$). When the components of the infant mortality series were independently analyzed, neither the neonatal ($g = 0.08$, $p = 0.7$) nor the post-neonatal series ($g = 0.14$, $p = 0.04$, ns at $p = 0.004$) showed a significant peak. We followed the significant periodogram analysis of the infant series with a power spectrum, which confirmed the previous cycle at 2 years. The power spectrum is shown in Figure 3. The other apparently large peak

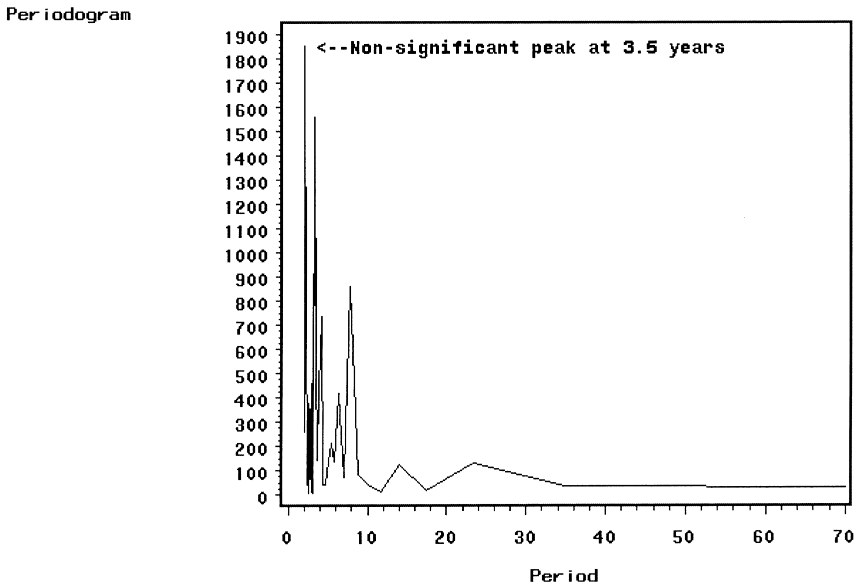

Figure 1. Periodogram of the adult series. Escazú.

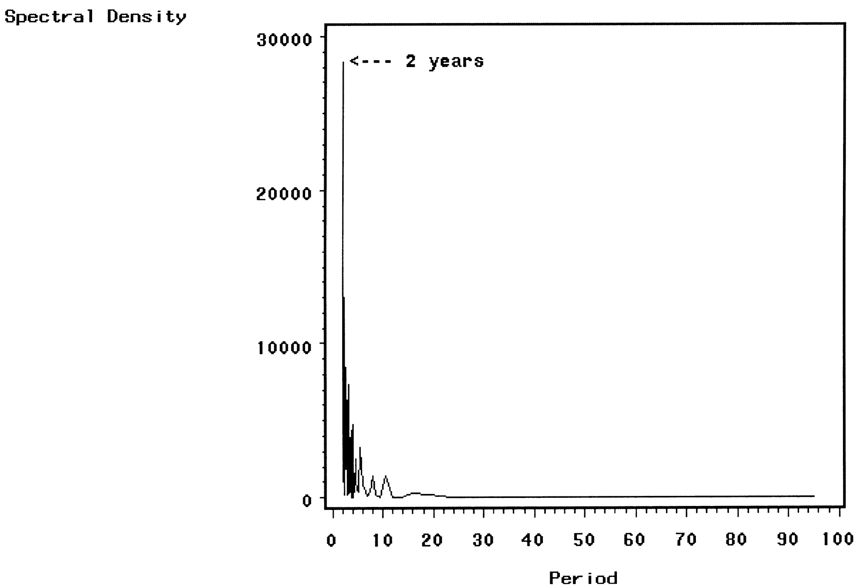

Figure 2. Spectral density of the neonatal mortality series. Acsa.

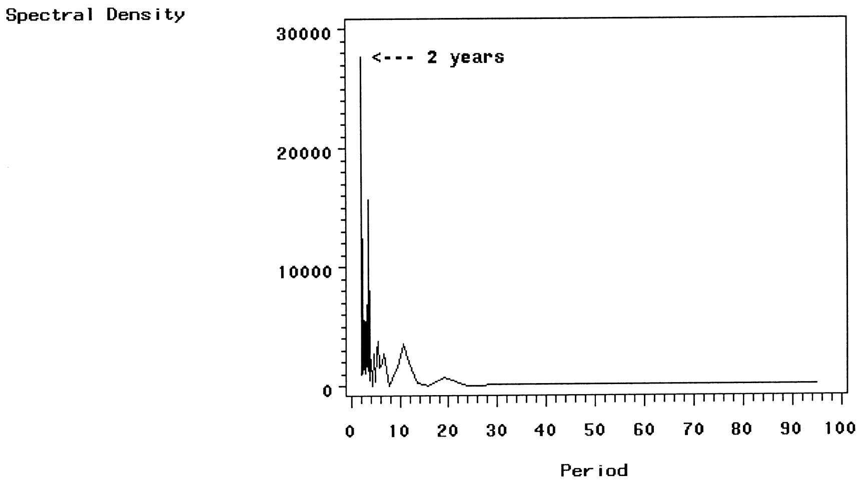

Figure 3. Spectral density of the infant mortality series. Miklosi.

did not reach significance. Since Warner (1998) suggests that it might be better to take the log of a series with extreme outliers, and since there were two in the infant series, we did another periodogram of the log of the differenced and winsorized series. We obtained a nonsignificant g statistic of 0.13 ($p = 0.07$).

Analysis of the Hungarian Settlement Data. The Hungarian settlement series was the most unstable of all the data sets we analyzed. It suffered from a trend and massive variance nonstationarity. Since it had no clear peaks, winsorization was not an option. Thus, we chose to follow Warner's (1998) suggestion to analyze the log of the raw data. The periodogram analysis of the log yielded a nonsignificant g of 0.11 ($p = 0.32$). We do not show the periodogram because, having computed it from the log, it is too "flat."

Conclusions

A cross-cultural analysis of mortality patterns is of interest to biological anthropologists, genetic epidemiologists, and public health officials. The overwhelming conclusion reached by our study is that mortality cycles are not as prevalent in agricultural populations as has been previously assumed. As opposed to Lin and Crawford's 1983 paper, in which mortality cycles were demonstrated in a variety of ecological settings, we find few cycles in four small agricultural populations in three entirely different regions: Costa Rica, Hungary, and Louisiana. We speculate that perhaps the population sizes of our groups are so small that they do not accumulate enough hosts with any regularity in order to enable the pathogen to be able to produce an epidemic.

Indeed, only the post-neonatal series in Acsa and the infant mortality series in Miklosi had significant peaks, both of 2 years. Given the small population size of these two Hungarian villages, we were surprised that the mortality cycles were so short. We expected that the longer time period was necessary for the population to accumulate enough susceptible individuals for it to suffer an epidemic. We think the suggestion of Mielke et al. (1987) that nonisolated populations experience short-term cycles applies here. That is, Miklosi and Acsa were at a virtual crossroads in Central Europe, and the young children of these villages were frequently exposed to new pathogens. This theory is similar to what Mielke et al. (1987) describe for the population of Finland, which was not isolated and experienced short-mortality cycles, in contrast to the Aland Islands, where the population was of low density, geographically isolated, and experienced longer cycles. This project supports the value of examining mortality time series of adults and subadults separately, and shows that epidemic cycles might be found in some subgroups, but not in others within the same settlement.

A review of the literature reveals that when the topic of epidemic cycles has been researched, the question has not been whether a population has a cyclical pattern of epidemics, but how long the cycle was. In other words, it has been assumed that epidemics had a cyclical pattern in agricultural groups. Upon realizing that we expected to reject the null hypothesis of no cycles for all the series, and that we are not alone in expecting agricultural groups to show epidemic cycles, we were reminded of what S.J. Gould refers to as the bias for "a good story" in scientific writing. In "Cordelia's Dilemma," Gould (1993) discusses the bias in scientific papers in which the null hypothesis is rejected, but in which a treatment effect is shown to exist. This of course, results in a biased view of the actual nature of the topic of interest: by concentrating so much on the "good story" of epidemic cycles, we have failed to see that agricultural groups in diverse ecological settings do not always follow a cyclic epidemic pattern. We would be interested in seeing other papers that demonstrate no cycles in similar groups. Perhaps what we have assumed to be the rule in small agricultural groups is not a rule at all.

Received 1 May 2002; revision received 7 March 2003.

Literature Cited

Bouma, M.J., and C. Dye. 1997. Cycles of malaria associated with El Niño in Venezuela. *J. Am. Med. Assn.* 278 (21):1772–1774.

Chatfield, C. 1984. *The Analysis of Time Series.* Bristol, UK: Chapman and Hall.

Cohen, M.N. 1989. *Health and the Rise of Civilization*. New Haven, CT: Yale University Press.

Duncan, C.J., S.R. Duncan, and S. Scott. 1996. Whooping cough epidemics in London, 1701–1812: Infection dynamics, seasonal forcing and the effects of malnutrition. *Proc. R. Soc. Lond. B. Biol. Sci.* 263:445–450.

Duncan, C.J., S.R. Duncan, and S. Scott. 1997. The dynamics of measles epidemics. *Theor. Popul. Biol.* 52:155–163.

Fuller, W.A. 1976. *Introduction to Statistical Time Series.* New York, NY: John Wiley.

Gould, S.J. 1993. Cordelia's dilemma. *Natural History* 2:10–18.

Jenkins, G.M. 1979. *Practical Experiences with Modelling and Forecasting Time Series.* UK: Titus Wilson Ltd.

Karlen, A. 1995. *Man and Microbes*. New York, NY: G.P. Putnam's Sons.

Koertvelyessy, T.A. 1983. Demography and evolution in an immigrant ethnic community: Hungarian settlement, Louisiana, USA. *J. Biosoc. Sci.* 15:223–236.

Koertvelyessy, T.A., and M.T. Nettleship. 1996. Ethnicity and mating structure in Southwestern Hungary. *Rivista di Antropologia (Roma)* 74:45–53.

Lin, P.M., and M.H. Crawford. 1983. A comparison of mortality patterns in human populations residing under diverse ecological conditions: A time series analysis. *Hum. Biol.* 55:35–62.

Madrigal, L. 1994. Mortality seasonality in Escazú, Costa Rica: 1851–1921. *Hum. Biol*. 66:433–452.

Madrigal, L. 2003. The use of archives in the study of microevolution: The case of Escazú, Costa Rica. In *Human Biology in the Archives Symposium*, A. Herring and A. Swedlund, eds. Cambridge, UK: Cambridge University Press. In press.

McCleary, R., and R.A. Hay. 1980. *Applied Time Series Analysis for the Social Sciences*. Beverly Hills, CA: Sage Publications.

McNeill, W.H. 1979. *Plagues and Peoples*. Garden City, NY: Anchor Press.

Mielke, J.H. 1982. Population movements and genetic microdifferentiation in Aland, Finland. *Collegium Antropol.* 6(1):19–38.

Mielke, J.H., L. Jorde, P.G. Trapp et al. 1984. Historical epidemiology of smallpox in Aland, Finland: 1751–1890. *Demography* 21(3):271–295.

Mielke, J.H., and K.J. Pitkänen. 1989. War demography: The impact of the 1808–09 war on the civilian population of Aland, Finland. *Eur. J. Population* 5:373–398.

Mielke, J. H., K.J. Pitkänen, L.B. Jorde et al. 1987. Demographic patterns in the Aland Islands, Finland, 1750–1900. *Yearbook of Population Research in Finland* XXV:57–74.

Morrison, Donald F. 1976. *Multivariate Statistical Methods*. 2nd ed. New York, NY: McGraw-Hill.

Olsen, L.F., G.L. Truty, and W.M. Schaffer. 1988. Oscillations and chaos in epidemics: A nonlinear dynamic study of six childhood diseases in Copenhagen, Denmark. *Theor. Popul. Biol.* 33:344–370.

SAS Institute Inc. 2001. Cary, NC.

Schwartz, J., and A. Marcus. 1990. Mortality and air pollution in London: A time series analysis. *Am. J. Epidemiol.* 131(1):185–194.

Scott, S., and C.J. Duncan. 1998. *Human Demography and Disease*. Cambridge, UK: Cambridge University Press.

Scott, S., and C.J. Duncan. 2001. *Biology of Plagues: Evidence from Historical Populations*. Cambridge, UK: Cambridge University Press.

Scott, S., S.R. Duncan, and C.J. Duncan. 1998. The interacting effects of prices and weather on population cycles in a preindustrial community. *J. Biosoc. Sci.* 30:15–32.

Statisztikai Hivatal. 1882. *A magyar korona orszagaiban az 1881. Ev elejen vegrehajtott nepszamlalas eredmenyei.* Budapest: Stephaneum Nyomda. [Statistical Office. 1882. *Results of the census conducted in the beginning of 1881 in the countries of the Hungarian Crown.* Budapest: Stephaneum Printing-house].

Waris, M. 1991. Pattern of respiratory syncytial virus epidemics in Finland: Two-year cycles with alternating prevalence of groups A and B. *J. Infect. Dis.* 163:464–469.

Warner, R.M. 1998. *Spectral Analysis of Time-Series Data*. New York, NY: The Guilford Press.