Investigating Risk Factors Associated with Syphilis Rate in the United States
Based on ARIMA and ARCH/GARCH Time Series Models

Chau-Kuang Chen, Ed.D., Camilia Comeaux, MSPH, Vanisha Brown, Ph.D., Juanita Buford, Ed.D.,
and Karmen Williams, MSPH

Meharry Medical College,
Nashville, Tennessee,
USA

ABSTRACT
The purpose of this study was to establish the relationship between U.S. syphilis rate and its risk factors using Autoregressive Integrated Moving Average (ARIMA), Autoregressive Conditional Heteroskedasticity (ARCH), and Generalized Autoregressive Conditional Heteroskedasticity (GARCH) time series models. The U.S. syphilis incidence was collected in years 1957-2009. Risk factors included U.S. per capita income, unemployment rate, poverty rate, percentage of the population who had less than nine years of education, and per capita alcohol consumption. The aggregated statistics were retrieved from national longitudinal databases through the Centers for Disease Control and Prevention (CDC). The study findings indicated that the most suitable ARCH/GARCH model contained four significant risk factors: poverty rate ($\alpha_1 = 1.893$ and $p = 0.001$), alcohol consumption ($\alpha_2 = 13.584$ and $p = 0.000$), unemployment rate ($\alpha_3 = -0.513$ and $p = 0.039$), and the three-year lagged syphilis rate ($\varphi_3 = -0.458$ and $p = 0.001$). ARCH/GARCH model exhibited superiority to the ARIMA model, because of the accuracy with a smaller mean absolute percentage error (MAPE) and the findings of risk factors were consistent with the literature reviews.

INTRODUCTION
The U.S. syphilis incidence decreased; yet, there were still multiple spikes observed throughout several periods within a 10-year span (Kilmarx et al., 1997). In 1990, the syphilis rates peaked with 20.3 per 100,000 primary and secondary (P&S) syphilis cases (Kilmarx et al., 1997). In 2000, the syphilis rate was 11.52 per 100,000 cases; while in 2008, 15.40 per 100,000 cases were reported (CDC, 2011). Research indicates that the increasing number of syphilis cases may be due to any combination of economic stressors, behavioral choices, and environmental conditions (Holtgrave & Crosby, 2003; Hogben & Leichliter, 2008). The frequent changes in syphilis rate represented a growing concern not only to medical providers, but also policy-makers and community populations nationwide. Therefore, constructing a statistical model to describe the relationship between syphilis rate and its risk factors may help policy makers facilitate the implementation of interventions effectively.

The reduction of the national syphilis rate was not only the results of medical technology, but also the intervention of public health policy that occurred throughout the years. Specially, the deeper understand...
and linkage concerning one’s environment and social conditions as it relates to adverse health outcomes also contributed to the reduction in syphilis rate.

The impact of syphilis on individuals is still a concern due to the lack of diagnosis and need for treatment to prevent transmission. A 1997 national study on socio-demographic factors and the variation in syphilis rate found the 10-year mean incidence of syphilis per 100,000 persons ranged from 0 to 140 in 896 counties nationwide (Kilmarx et. al, 1997). The time frame of 1957-2009 was important to the scope of this study because the pivotal economic, social, and environmental shifts in history that could possibly explain the various peaks and valleys observed. Additionally, the case definition of sexually transmitted diseases (STDs) and the national emphasis on STD prevention in the 1990s, signified the importance of the selection of syphilis as the STD of primary interest (Kilmarx et al., 1997).

History of Syphilis

Historically, syphilis has been a national concern that has been the focus of several initiatives and interventions aimed at decreasing incidence cases and eliminating the spread of the disease. Throughout the years there have been shifts and changes in syphilis incidence rates across the United States. During the 1940s, syphilis rates were widely distributed in the U.S. Although the incidence peaked during the 1940s, subsequent aggressive public health interventions that involved penicillin, case finding, and contact tracing led to a significant decreases in the incidence of P&S syphilis during the following decade (Nakashima, Rolfs, Flock, Kilmarx, & Greenspan, 1996). Since the late 1950s syphilis rates had shown recurrent peaks and troughs in approximately 10-year cycles (Nakashima, Rolfs, Flock, Kilmarx, & Greenspan, 1996). As a result of various public health interventions between 1947 and 1956 incidence cases decreased from 66.9 to 3.9 cases per 100,000 persons (Nakashima, Rolfs, Flock, Kilmarx, & Greenspan, 1996).

The following events caused spikes in recorded syphilis data: the sexual revolution in the 1970s; the HIV epidemic in men who had sex with men (MSM) in the 1980s and 2002; and the HIV epidemic in heterosexual men and women in the 1990s (Grassly, 2005). From 1986 to 1990, the U.S. observed an 85% increase in the incidence of primary and secondary syphilis (P&S) (CDC, 1999). Syphilis is a sexually transmitted disease (STD) that is caused by the bacterium *Treponema pallidum*, identifiable by blood testing, and curable with a dose of penicillin. During the late 1990s, syphilis resurfaced as a major public health problem (Bofill, 1996; Thomas, Clark, Robinson, Monnett, Kilmarx, & Peterman, 1999). However, it was not until 1999 the Surgeon General announced “The National Plan to Eliminate Syphilis from the United States,” that a dramatic decrease in syphilis incidence rates occurred (CDC, 1999). This plan aimed to reduce the annual number of syphilis cases to less than 1000 by 2005 (Biddlecom, 2004). The plan called for enhanced prevention efforts in areas with high rates of syphilis or high potential for re-emergence of syphilis. After the launch of the syphilis elimination plan, the rate of P&S syphilis in the United States declined in 1999 and 2000 (CDC, 2006).

Research suggests the resurgence of syphilis paralleled the rise of social problems such as drug/alcohol abuse, poverty, and unemployment (Bofill, 1996). Social problems were also correlated with environmental conditions and consequently social behaviors. More research suggested that syphilis rates continued to remain high in some urban and rural areas where minority populations suffered from poverty, lack of access to health care, and breakdown of stable community and personal relationships (CDC, 2006; MMWR, 1985).
Risk Factors

Previous research investigated the changes in behavior and sexual practices that might account for the variation in syphilis rates over time (Kilmarx et al., 1997; Nakashima, Rolfs, Flock, Kilmarx, & Greenspan, 1996; Thomas, Clark, Robinson, Monnett, Kilmarx, & Peterman, 1999; Holtgrave & Crosby, 2003; Hogben & Leichliter, 2008). Studies report that, “The distribution and trends of syphilis in the population are influenced by several determinants including: biologic factors; sexual behaviors; biomedical factors; availability of and access to healthcare; healthcare seeking behaviors; public health efforts to prevent and control syphilis; population factors; and sociocultural factors” (Kilmarx et al., 1997; Nakashima, Rolfs, Flock, Kilmarx, & Greenspan, 1996). The hypothesis that the above mentioned determinants impact varying syphilis rates over the years, proposes ideas for change in social, economic and environmental conditions in which the population lives.

Research by Fenton, Breban, Vardavas, et. al. (2008) suggested syphilis outbreaks can be explained by: changes in sexual behavior, the gay liberation movement in the 1970s, the HIV epidemic, the sexual revolution, and changes in the intensity of syphilis prevention programs. Past research studies have focused on correlations between race and syphilis (Holtgrave & Crosby, 2003). In 2006, Farley proposed a model for 13 predictor variables to explain elevated rates of STDs in African-Americans. Not only did they look at historical and current racism, and racial segregation; they looked at other variables including: drug and alcohol use; loss of low-skill jobs; joblessness; long-term concentrated joblessness; and poor education (Farley, 2006). Societal changes/movements suggest that there are other factors then racial and regional demographics that contribute to volatile syphilis rates over the years. Therefore, in this study we chose to investigate what social and economic factors influence behaviors and the environment that contribute to fluctuating syphilis incidence rates nationwide.

Research has found direct and indirect associations between factors such as: social capital, poverty, education, income inequality, and alcohol use that contribute to syphilis rates (Semaan, Sternber, Zaidi, & Aral, 2007; Toomey, Moran, Rafferty, & Beckett, 1993). Knowledge of these economic and environmental factors that have contributed to syphilis rates of the past may be helpful in developing intervention procedures to alter future societal shifts that potentially impact syphilis incidence rates.

SIGNIFICANCE OF THE STUDY

The purpose of this study was to investigate risk factors that contributed to the U.S. syphilis rate using the ARIMA and ARCH/GARCH time-series models. Utilizing these two types of techniques, this study aided understanding two focus areas: (1) of the two time-series methods, which better described the relationship between the risk factors and volatile syphilis rate; and (2) how did the spikes of the past affect the present, and how the present risk factors, in turn, will affect the future incidence of syphilis (Chen, 2008; Engle, 2001).

In a national study in 1997, researchers found that socio-demographic characteristics accounted for 71% of the variation in syphilis rates between counties (Kilmarx et al., 1997). Furthermore, Nakashima, Rolfs, Flock, Kilmarx, & Greenspan (1996) highlighted the key understanding of syphilis disparities within various U.S. communities included defining correlation between syphilis and ecologic indicators. Ecologic indicators encompass economic and environmental conditions measurable by social disparities (Holtgrave & Crosby, 2003; Hogben & Leichliter, 2008). Therefore, past research encouraged the selection of U.S. per capita income, U.S. unemployment rate, U.S poverty rate, percentage of the population who had less than nine years education, and per capita alcohol consumption as risk factors to be investigated in this study.
This study aimed to investigate fluctuations in syphilis incidence rates over time through the use of the three time series models and the explanation of how socially- and economically-influenced factors can be altered at multiple levels based on recommendations from predictive models.

**METHODOLOGY**

**ARIMA**

Time series analysis techniques are statistical methods generally used for forecasting. The Box-Jenkins Autoregressive Integrated Moving Average (ARIMA) Model is one forecasting technique that involves the three stages of the modeling process: model identification, parameter estimation, and diagnostic checking (Chen, 2008). The iteration of these stages is essential for finding the most appropriate model. ARIMA modeling is operational under the assumption that the data series is stationarity, and that it requires a minimum of forty-five data points (Chen, 2008).

**ARCH/GARCH**

Although typically used in the financial and economic arena for risk assessment and market changes, the techniques of ARCH/GARCH can be useful for social science and health assessment (Engle, 2001). ARCH/GARCH is another time series method in which the average error between the best fit line and the actual data are used to develop a model for forecasting (Engle, 2001). This type of modeling is particularly useful when variances of the residuals are not constant over time. The uncertainty of syphilis data points over time presents difficulties when predicting future data; therefore, the use of ARCH/GARCH modeling may be an essential tool for measuring volatile values.

**Study Design**

The syphilis incidence data utilized in this quasi-experimental study was collected from the years of 1957 to 2009. Typically in a quasi-experimental design there is no control group; trends or effects are estimated between the before and after changes or “shocks” in the data series. The use of time-series techniques, such as ARIMA and ARCH/GARCH, on volatile syphilis rates over time was chosen to construct a reasonable and workable model that best described the relationship between syphilis and risk factors and ultimately, yielded a statistically sound model for future forecasting. The STATA 12.0 software package was used to construct the ARCH/GARCH models, while the SPSS 19.0 software package was utilized to build the ARIMA model.

**Risk Factors and their Measure Scales**

The data, collected at a national level, were pre-coded as aggregated counts according to uniform specifications, and passed through several statistical quality checks by the CDC Wonder data system (CDC, 2011). The syphilis rate and its risk factors were retrieved from U.S. longitudinal databases through the Centers for Disease Control and Prevention (CDC, 2011).

Table 1 depicts the risk factors used for data analysis: U.S. poverty rate, percentage of U.S. population with less than nine years of education, U.S. per capita alcohol consumption, U.S. per capita income, and U.S. unemployment rate. The risk factors for syphilis were collected from the following longitudinal databases: Poverty rate was collected from the U.S. Census Bureau database; per capita alcohol
consumption rate was collected from the National Institute on Alcohol Abuse & Alcoholism; per capita income was collected from the Bureau of Economic Analysis; and the unemployment rate and percentage of the population who had less than nine years education were collected from the Bureau of Labor Statistics (Table 1) (CDC, 2011). Therefore, no unique identifiers were associated with the data collected.

Table 1.
Risk Factors for U.S. Syphilis Rate

<table>
<thead>
<tr>
<th>Risk Factors</th>
<th>Descriptions</th>
<th>Measurements</th>
<th>Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>U.S. Poverty Rate</td>
<td>Percentage of people who were in poverty in a calendar year</td>
<td>Using the sum of family incomes over the year divided by the sum of poverty thresholds.</td>
<td>U.S. Census Bureau database</td>
</tr>
<tr>
<td>U.S. Per Capita Income</td>
<td>The amount of income distribution by the population of that area.</td>
<td>Dividing the personal income of the residents by the population of that area as of July 1 for the reference year.</td>
<td>Bureau of Economic Analysis</td>
</tr>
<tr>
<td>U.S Per Capita Alcohol Consumption</td>
<td>The amount of liters of pure alcohol consumed by adult population</td>
<td>The sum of alcohol production and imports divided by the adult population</td>
<td>National Institute on Alcohol Abuse &amp; Alcoholism</td>
</tr>
<tr>
<td>Percentage of U.S. Population with Less Than 9 Years of Education</td>
<td>Highest level of education that the individual has completed</td>
<td>Number of responses recorded from adults of the household divided by number of survey participants</td>
<td>Bureau of Labor Statistics</td>
</tr>
<tr>
<td>U.S. Unemployment Rate</td>
<td>The prevalence of unemployment</td>
<td>Number of unemployed individuals divided by all individuals currently in the labor force.</td>
<td>Bureau of Labor Statistics</td>
</tr>
</tbody>
</table>

Model Construction

ARIMA

The ARIMA methodology was implemented to construct the suitable model and to identify significant risk factors in predicting syphilis rates over time. Longitudinal data with a minimum of 45 to 60 data points may need to obtain accurate forecasts (Chen, 2008). This study contained 53 data points to examine a trend. Below is the three-step procedure to construct ARIMA model: (1) Model Identification; (2) Parameter Estimation; and (3) Diagnostic Checking (Chen, 2008).

ARCH/GARCH

The most suitable fitting ARCH/GARCH models were constructed when the ARIMA model was identified. Testing the ARCH/GARCH effects involves the following six steps:
ARCH (q) model.

1) State the null and alternative hypotheses:
   \[ H_0 : \alpha_i = 0 \text{ for all } i = 1, \ldots, q \text{ in the absence of ARCH components.} \]
   Meaning, that for all of the estimated coefficients of the risk factors, there are no significant
difference in the absence of ARCH/GARCH components.
   \[ H_a : \text{at least one of the estimated } \alpha_i \text{ of the risk factors is significant in the presence of ARCH components.} \]

2) Estimate the best fitting AR(q) model by using the LeGrange Multiplier Test.
   \[ y_t = \alpha_0 + \alpha_1 y_{t-1} + \ldots + \alpha_q y_{t-q} + \epsilon_t \]
   \[ \text{where } t = q+1, \ldots, T \]
   By definition, LeGrange Multipliers provide a mathematical sequence for determining the
maximum and minimum of a certain function in the presence of various constraints (Engle, 2001).

3) Obtain equations for the mean and volatility.
   \[ y_t = x_i \hat{\beta} + \epsilon_t \] \hspace{1cm} \text{(conditional mean)}
   \[ \sigma_t^2 = \gamma_0 + \gamma_1 \epsilon_{t-1}^2 + \gamma_2 \epsilon_{t-2}^2 + \ldots + \gamma_m \epsilon_{t-m}^2 \] \hspace{1cm} \text{(conditional variance)}
   where \( \epsilon \) denotes the error term, \( t \) is the sample size, and \( \gamma_0, \gamma_1, \ldots, \gamma_m \) parameters (Engle, 2001).

4) Obtain Ordinary Least Squares (OLS) residuals (\( \epsilon \)) then square residuals (ie \( \epsilon_{t-1}^2, \ldots, \epsilon_q \)). The
   OLS method estimates unknown parameters by minimizing the squared distance between
   observed and predicted values.
   \[ \epsilon_t = \sigma_t z_t, \quad \hat{\epsilon}_t = \mu + e_t \] \hspace{1cm} \text{(Ordinary Least Squares Formulas)}

5) In an additional regression, regress the \( \hat{\epsilon}_t^2 \) on a constant and the first set of residuals, \( \hat{\epsilon}_{t-1}, \ldots, \epsilon_q \)
   (Engle, 2001).
   \[ \hat{\epsilon}_t^2 = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 \]

6) Accept or reject hypothesis based on chi-squared test. This step determines if
   the model is the most appropriate fit for the relationship between syphilis
   rate and the risk factors (Engle, 2001).
   Accept \( H_0 \) if \[ TR^2 < \chi^2(q) \]
   Reject \( H_0 \) if \[ TR^2 > \chi^2(q) \]
   \[ T \text{ is the number of observations in the additional regression and } R^2 \text{ is the additional regression that}
   \text{contains a constant.} \]

Meaning, in a sample of \( T \) residuals under the null hypothesis of no ARCH errors, the test
statistic \( TR^2 \) follows \( \chi^2 \) distribution with \( q \) degrees of freedom. If \( TR^2 \) is greater than the Chi-square critical
value, we reject the null hypothesis and conclude there is an ARCH effect in the ARMA model. If \( TR^2 \) is
smaller than the Chi-square critical value, we do not reject the null hypothesis.

RESULTS

The recorded syphilis rates over the 53-year period in the United States are shown in Figure 1. As displayed
in this figure, there is a generally significant declining trend observed from about 73 per 100,000 reported
cases in 1957 to about 11 per 100,000 reported cases in 1999. Yet, there were still several spikes of
increases and decreases throughout the 53-year time span. This phenomenon reflects multiple events,
initiatives and changes in national syphilis incidence.
There is particular interest in the significant spike in the early 1990s, where rates increased from 28 per 100,000 to about 55 per 100,000. This increase in syphilis cases could be paralleled with the literature on the rise of HIV/AIDS infections in heterosexual and homosexual individuals. The introduction of the idea of “coinfection” may have increased transmission rates during this period (Bartlett et al., 2008). These two STDs were linked because syphilis typically produced genital sores which provided a vehicle for HIV/AIDS transmission. Additionally, a gradual increase was observed starting around 2005 (see Figure 1). The decade between 1977 and 1987 and the five-years between 2000 and 2005 showed relative stability in the number of syphilis cases reported about 30 per 100,000 and 11 per 100,000, respectively. The stability observed in the early 2000s could be a direct reflection of the national policy initiative for syphilis incidence reduction and the CDC’s focus of syphilis prevention in 1999 (CDC, 1999).

ARIMA (p, d, q) Models

The baseline model of the ARIMA methodology started with the ARIMA (0, 0, 0) in which the orders of autoregressive (p), differencing (d), and moving average (q) are set to zero. The syphilis rate series and all fiverisk factors (U.S Poverty Rate, U.S. Per Capita Income, U.S. Per Capita Alcohol Consumption, Percentage of Population with Less Than Nine Years of Education, and U.S. Unemployment Rate) were entered in the model equations simultaneously.

The first step in developing the ARIMA model was to investigate whether or not the syphilis rate series was stationary, because of the basic assumption and requirement for ARIMA is stationary (i.e., detrended) data. Figure 1 displays the 1957 to 2009 syphilis rate series as a non-stationarity pattern, i.e., a chain of rapid decrease, steady increase, and declining trends. The declining trend observed in Figure 1, which began in the upper left-hand corner and ended in the bottom right-hand corner, further supported the need for differencing. Additionally, in Figure 2 the ACF shows nonrandom pattern (i.e., one or more of the autocorrelations are different from zero). Therefore, the first order of differencing, ARIMA (0, 1, 0) model, was implemented to achieve the stationarity requirement. In other words, the syphilis rate series ARIMA (0, 0, 0) model was transformed to the new syphilis rate series with the first order of differencing, ARIMA (0, 1, 0) model.
In comparing Figure 2 and Figure 3, a higher level of stationarity was achieved as shown by ACF values of the residual series more frequently approaching zero, as a result of the first differencing. In Figure 3, the non-zero values of the ACF are seen in lags 1, 4 and 5; while, the zero values of ACF are seen between lag 6 and lag 24 for the ARIMA (0, 1, 0) model. Conversely, the non-zero values of the ACF are seen in lags 1 and 2; while, the zero values of ACF in Figure 2 appear between lag 3 and lag 24 for the baseline model, ARIMA (0, 0, 0). Therefore, it was concluded that there is no need to consider a higher order of differencing because the first order of differencing has achieved sufficient stationarity for the syphilis rate series.

**Figure 3.**
**ACF and PACF Graphs for ARIMA (0, 1, 0)**

The residual ACF and PACF plots for ARIMA (0, 1, 0) in Figure 3, the ARIMA (4, 1, 0) model is suggested for the syphilis rate series, which represents a fourth order of the autoregressive \( p = 4 \) with a first degree of differencing \( d = 1 \) and a zero order of moving average \( q = 0 \). This fact is because ACF exhibits a pattern of exponential or sine-wave decay and PACF cuts off sharply at lag 4. Meaning the higher-order
autocorrelations are effectively explained by lag 4, data from the four past years significantly impacted the event of the current year’s rates. As depicted in Table 2, the resulting model, ARIMA (4, 1, 0), contains two significant risk factors, Alcohol consumption rate ($\beta_1 = -11.258$ and $p = 0.034$) and AR (4), the four-year lagged syphilis rate ($\phi_1 = -0.477$ and $p = 0.005$).

Table 2.
ARIMA Model for U.S. Syphilis Rate Generated by SPSS Software

<table>
<thead>
<tr>
<th>ARIMA (4, 1, 0)</th>
<th>Estimate</th>
<th>SE</th>
<th>t</th>
<th>P Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syphilis</td>
<td>constant</td>
<td>52.726</td>
<td>23.532</td>
<td>2.241</td>
</tr>
<tr>
<td>AR</td>
<td>Lag 1</td>
<td>0.280</td>
<td>0.151</td>
<td>1.855</td>
</tr>
<tr>
<td></td>
<td>Lag 2</td>
<td>0.255</td>
<td>0.171</td>
<td>1.493</td>
</tr>
<tr>
<td></td>
<td>Lag 3</td>
<td>0.022</td>
<td>0.173</td>
<td>0.128</td>
</tr>
<tr>
<td></td>
<td>Lag 4</td>
<td>-0.477</td>
<td>0.159</td>
<td>-2.997</td>
</tr>
<tr>
<td>Poverty</td>
<td>Lag 0</td>
<td>-0.611</td>
<td>0.424</td>
<td>-1.439</td>
</tr>
<tr>
<td>Alco_com</td>
<td>Lag 0</td>
<td>-11.258</td>
<td>5.126</td>
<td>-2.196</td>
</tr>
<tr>
<td>Unemploy</td>
<td>Lag 0</td>
<td>0.408</td>
<td>0.553</td>
<td>0.737</td>
</tr>
<tr>
<td>Pc_incom</td>
<td>Lag 0</td>
<td>0.000</td>
<td>0.000</td>
<td>-1.332</td>
</tr>
<tr>
<td>P_le9_ed</td>
<td>Lag 0</td>
<td>-0.045</td>
<td>0.172</td>
<td>-0.262</td>
</tr>
</tbody>
</table>

Notes: *$p < .05$

In the final step of the ARIMA modeling process, the Ljung-Box Chi-square test was used to assess the model appropriateness. The Ljung-Box Chi-square test result indicates that ARIMA (4, 1, 0) model is not appropriate because the null hypothesis—the appropriateness of ARIMA model—is rejected.

ARCH/GARCH models

In the STATA software package, ARCH (1) and GARCH (1) models were built by using ARIMA (3,1,0) model because of the production of significant results. The ARIMA (4,1,0) model identified in the previous analysis was not inputted into the ARCH/GARCH analysis because the data did not converge; therefore, the next modeling step down was utilized. As shown in Table 3, the ARCH(1)/GARCH(1), contained the significant risk factors, Poverty Rate ($\alpha_1 = 1.893$ and $p = 0.001$), Alcohol Consumption ($\alpha_1 = 13.584$ and $p = 0.000$) and Unemployment Rate ($\alpha_1 = -0.513$ and $p = 0.039$).

Table 3.
ARCH/GARCH Models for U.S. Syphilis Rate Generated by STATA Software

<table>
<thead>
<tr>
<th>ARCH (1) / GARCH (1) Model</th>
<th>Estimate</th>
<th>SE</th>
<th>z</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Syphilis</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Poverty</td>
<td>1.893</td>
<td>0.548</td>
<td>3.45</td>
<td>0.001**</td>
</tr>
<tr>
<td>Alco_com</td>
<td>13.584</td>
<td>2.387</td>
<td>5.69</td>
<td>0.000***</td>
</tr>
</tbody>
</table>

ARCH (1) / GARCH (1) Model
Notes: *p < .05; **p < .01; ***p < .001

The Ljung-Box Chi-square test result indicates that this model is appropriate because it shows statistically significant risk factors in the presence of ARCH/GARCH effects in the model and future syphilis rate predictions should be performed based on the resulting model.

DISCUSSION

Summary

The important findings can be derived based on the relationship between volatile syphilis rate and its risk factors. As discussed in the literature review, syphilis is a disease transmissible by sexual contact and pregnancy. As shown in Figure 1, a sharp increase was observed in syphilis rate in the early 1990s. Also, researchers discussed the emergence of the HIV/AIDS epidemic; and the crack/cocaine and prostitution epidemic during the same timeframe. Therefore, these two critical events might suggest a significant component of STD transmission as well as syphilis volatility (Biddlecom, 2004; Guinn et al., 1995).

In the early 1990s, the surprise and shock of the HIV/AIDS epidemic coexisted with its twin brother, syphilis. This trend presented major concerns not only for the men who have sex with men (MSM) community but also heterosexual individuals engaging in risky sexual behaviors (Fichtenberg, Jennings, Glass, & Ellen 2010; Heffelfinger, Stuart, Berman, & Weinstock, 2007). Risky sexual behavior and improper condom use were also observed during the crack/cocaine epidemic in urban areas (Guinn et al., 1995; Respiso,Frieyro, Rivas-Ruiz,& De Troya, 2010). In urban centers the prostitution business was utilized as a method of obtaining crack/cocaine narcotics (Guinn et al., 1995). Improper condom use during these risky encounters continued the spread of disease throughout multiple sex partners (Respiso,Frieyro, Rivas-Ruiz,& De Troya, 2010). The combined effects of decreased risky behavior amongst individuals or groups; and an increase of government-sponsored STD education on transmission and preventive methods might have contributed to the decline observed in syphilis the years after.

Table 3 illustrates the comparison of the ARIMA and ARCH/GARCH models. The ARIMA (4, 1, 0) model, contained two significant risk factors: Alcohol consumption rate ($\beta_1 = -11.258$ and $p = 0.034$) and AR (4), the four-year lagged syphilis rate ($\phi_1=-0.477$ and $p = 0.005$). Conversely, the most suitable ARCH/GARCH model consisted of the three significant risk factors: U.S. poverty rate ($\alpha_1=1.893$ and $p = 0.039$).
alcohol consumption ($\alpha_2 = 13.584$ and $p = 0.000$) and unemployment rate ($\alpha_3 = -0.513$ and $p = 0.039$). These significant risk factors in each model were consistent with findings from literature reviews on the perpetuation of syphilis transmission.

**Table 3.**
**Comparison between ARIMA and ARCH/GARCH Models**

<table>
<thead>
<tr>
<th>Model Name (Model Type)</th>
<th>Model A: ARIMA (4,1,0) (First Order of Autoregressive with First Differencing)</th>
<th>Model B: ARCH(1)/GARCH(1) (ARIMA (3,1,0) foundational input)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Risk Factors Entered in the Equation Simultaneously</td>
<td>All 5 risk factors</td>
<td>All 5 risk factors</td>
</tr>
<tr>
<td>Chi Square Test for Model Appropriateness</td>
<td>$\chi^2 = 13.038$ with $p = 0.524$</td>
<td>$\chi^2 = 74.62$ with $p = 0.000$</td>
</tr>
</tbody>
</table>
| Significant Risk Factors Contained | AR (4) $p = 0.005^{**}$
U.S. Alcohol Consumption $p = 0.034^{*}$ | AR (3) $p = 0.001^{***}$
U.S. Poverty Rate, $p = 0.001^{**}$
U.S. Alcohol Consumption $p = 0.000^{***}$
U.S. Unemployment Rate $p = 0.039^{*}$ |
| Remarks | Best ARCH/GARCH Model |

Notes: $^*p < .05$; $^{**}p < .01$; $^{***}p < .001$

The structures of these two models were somewhat similar, in that, the autoregressive terms suggest that they both have the ability to construct predictive models with longitudinal data sets. Using the ARIMA and ARCH/GARCH time series models, the goal of this study was to investigate risk factors that contributed to U.S. syphilis rate.

The ARCH/GARCH model was found to satisfy the three elements of having the most suitable model: (1) It had the most number of risk factors either positively or negatively associated with volatile syphilis rates over the 53-year period; (2) The Chi-squared test supported the findings that ARCH/GARCH model was capable of performing accurate prediction.

The ARCH/GARCH model found that spikes or variations in syphilis rates over a 53-year span could be attributable to risk factors such as alcohol consumption, unemployment rate, and poverty rate. One can conclude that elevated levels of alcohol consumption can influence the engagement of risky sexual behavior while impaired. Additionally, low unemployment rates might encourage individuals to engage in risky
sexual behaviors or seek risky environments that influence negative behavior. High poverty rates could be inferred to correlate to communities unable to secure and access adequate education and prevention measures, such as condoms, due to lack of financial resources or convenience. These risk factors could further be correlated with the influx of crime and substance abuse rates over time periods.

This study demonstrated that the use of the Social-Ecological Model could aid in understanding the connection and influence multiple levels of the society and environment have on syphilis transmission rates and health outcomes. Essentially, the success of reducing future syphilis rates, specifically in the risk factors identified in this research, are to encourage a change in a “top-down” effect. This means to impact national policy-level influence to promote changes in legislative involvement and accountability; development of media support tools for awareness and lobbying; increased institutional research and practice; provision of facilities and opportunities for population networks to seek and maintain assistance; and the development of support tools and awareness education for individuals.

CONCLUSION
This study demonstrated the usefulness of ARIMA and ARCH/GARCH modeling tools. It highlighted the history of syphilis in the United States to establish the nonlinear function between syphilis rate and its risk factors. There is historical reasoning supporting the selection of syphilis as the STD of interest. National research has not only classified the disease as an important health issue to be addressed, but also plausible explanations have been suggested to prevent perpetuation of transmission. Although there is not much research on the use of time series analysis specifically with syphilis and other related public health issues, there is a need to understand the impact of risk factors on syphilis trends and changes. Therefore, ARIMA and ARCH/GARCH modeling provide an avenue for detecting the trend, determining the risk factors contributing to fluctuation.

In addition to contributing to the knowledge-based in the field of Public Health field on time-series analysis and related syphilis rates, there are major strengths of the ARIMA and ARCH/GARCH models. First, the structural approaches in both models were to construct candidate models, eliminate inappropriate ones, and ultimately retain the most suitable model. This approach avoids the arbitrary choice of a specific model that may not be the best fit. Also, a side-by-side comparison was made between both ARIMA and ARCH/GARCH methods to illustrate each models structure and demonstrate the capability of explaining relationship among variables. Second, the modeling process enables the researcher to identify a subset of key risk factors from a pool of potential risk factors. The ARIMA model yielded one risk factor and one four-year lagged syphilis rate while the most suitable ARCH/GARCH model generated three risk factors and only one-year lagged syphilis rate series. These risk factors were significantly associated with syphilis rate and were consistent with the research findings from the literature.

Finally, this study confirmed the applicability of the social ecological perspective as a theoretical basis for implementing change. There was a need to understand the impact of risk factors on syphilis trends. Therefore, the findings from the ARIMA and ARCH/GARCH models could contribute to the field of Public Health by (1) providing an avenue for visualizing trends in medical diseases, and (2) determining the significant risk factors in contributing to the uncertainty of the adverse health outcome. Therefore, the use of these time series models may enhance the innovation of preventive measures in combination with enforcement of prevention strategies for both health professionals and policymakers.
Implications to Public Health

ARCH/GARCH models are trend recognition tools that allow researchers to use specific longitudinal dataset to create the workable model. They were typically used to assess risks and market changes overtime to predict adverse health outcomes. Engle (2001) described the usefulness of these models in terms of their ability to factor in major shocks in volatility shifts over time; therefore, in public health, social, economic and ecological factors attributed to these shocks can be predicted for health planning purposes.

The research findings might highlight the need for the implementation of intervention programs for high risk populations. The programs might address how to reduce risky behaviors associated with rising syphilis incidence rates and ways to address strategies for altering the role that socioeconomic status, education, and environmentally-influenced status play in trends of syphilis.

REFERENCES


