Using the SAS® System to Evaluate Syphilis Control

Ray L. Ransom, Centers for Disease Control
Joann Schulte, Centers for Disease Control
Melinda L. Flock, Centers for Disease Control

The Centers for Disease Control (CDC) were invited to evaluate a syphilis control program in the state of Arkansas. The purpose of this paper is to discuss how this process was facilitated by the SAS system and not the results of the evaluation. This will be presented at another forum by the second author. Since syphilis is a sexually transmitted disease (STD), partner notification (PN) is one strategy used to control spread. Sexual partners of infected patients are identified as quickly as possible and offered counseling and treatment. Optimally, the partner(s) will be contacted while still incubating, receive treatment aborting the infection, and disease transmission is prevented.

Syphilis is an acute and chronic STD characterized by three stages. Primary syphilis is recognized by a lesion which occurs at the site of exposure about three weeks after inoculation and without specific treatment can last up to five weeks before healing spontaneously. Secondary syphilis occurs after the primary stage, usually about eight weeks after initial exposure, and may consist of a generalized secondary eruption of lesions. These symptoms may disappear spontaneously in weeks or last up to a year. A latency period may follow and has no obvious signs or symptoms. In the first few years, latency is sometimes interrupted by the recurrence of infectious lesions or in several years could result in complications with cardiovascular or central nervous system functions. Latency in women may result in prenatal infection which is often fatal by infancy. Latency is divided into 2 stages, early and late, for treatment purposes. Late latency occurs one year after inoculation. Treatment consists of benzathine penicillin therapy. Traditional patient interview periods are based on the stages of disease. Primary cases are asked to name their partners from the past 90 days. Secondary cases are asked about the past 180 days and early latent cases are asked about the past year. Usually, PN is not used for late latent cases. When a patient is diagnosed with syphilis in a state clinic they are referred to a disease interventions specialist (DIS) who will conduct the PN interview. It is up to these staff to contact the partners as soon as possible. The information received during the interviews plus additional demographic and clinic information on the case and contacts are recorded on a standardized questionnaire developed by CDC. All partner referrals are confidential and all personal identifiers are removed before the data are sent to the CDC.

When the CDC received the data from Arkansas, they were keystumped and stored on a mainframe disk pack in the MVS environment. Version 6.06 of the SAS® system was used to edit the data and assign value labels and formats. The elaborate merging capabilities of SAS software allowed for the creation of a SAS dataset containing one record for each contact and FIRST.byvariable and LAST.byvariable processing was used to extract only case information or to summarize characteristics of contacts by case. This SAS dataset was downloaded to Version 6.04 of the SAS system for the personal computer using the DOWNLOAD procedure and the SAS Micro-to-Host link where it could be used by other pc users.

The entire analysis of the data was performed using SAS software and SAS/GRAPH® software. Graphics were generated to show distribution of contact characteristics by case and contact stage of disease. These graphics were important in determining which contacts were being infected or infectious contacts were being treated. This provided much information about the amount of disease intervention occurring and suggested possible problem areas in the interviewing process. Plotting techniques were used to represent the time elapsed between case interview and contact treatment. This is an obvious qualifier of the efficiency of the program. FIRST.byvariable and LAST.byvariable processing was used, once again, to count contacts with certain characteristics by case to determine indices commonly used for syphilis control program evaluations. These indices are "contacts elicited", "contacts examined", "new contacts examined", "contacts treated", "new contacts treated" and "new infections treated". This tells us the number of named, examined, examined for the first time, treated, treated for the first time, and infected and treated contacts per patient respectively. Some contacts are given penicillin without determination of infection, thus necessitating the later index. These indices are summed and the averaged using the MEANS procedure. SAS for the MVS environment was used once again to analyze 32 years of national syphilis data to produce baseline indices for comparison. This data was output to a subset dataset and downloaded for graphical and tabular comparisons.

Another aspect of syphilis control includes the determination of source and spread of the infection. A computational graphic is located on the back of the interview questionnaire which allows the DIS to plot exposure and lesion information to attempt to determine source and spread. A SAS program was written on the pc to take this information from the questionnaire and produce a table which presents this information by case and contact groups for quick analysis. This is much faster than the "by hand" method and standardization allows for easier interpretation. Although it may not be feasible to have every clinic equipped with SAS, this
technique might facilitate learning more of the source and spread of syphilis at the CDC level.

Soon after the analysis was completed, there was an interest in doing a like evaluation of the syphilis control program in the city of Philadelphia. Since all syphilis data were reported on these standardized forms, common variables were available to allow for the use of the same programs used for Arkansas. Philadelphia's data were already computerized using dBase III Plus and the DBF procedure was used to convert these files to SAS datasets. Once this was done, all that was required for the analysis was running the concatenated programs. This analysis was completed in a matter of hours.

The flexibility of the SAS System allows for the use of data from various platforms and sophisticated analysis which can be standardized and recreated over and over again. The analysis found no limits with the SAS System and the programming will only be done once. The problems that these analyses sought to resolve are inherent to all syphilis control programs throughout the nation and such efficient evaluation techniques could result in much more progress towards the control of syphilis.

SAS is a registered trademark of SAS Institute Inc. in the USA and other countries.

Ray L. Ransom
CPS, DSTD/HIVP, CRB, VSS
mailstop E-02
Centers for Disease Control
1600 Clifton Road
Atlanta, GA 30333
(404) 639-2570