ABSTRACT
A SAS program was designed to monitor HIV seroconversion (i.e., the transition from HIV seronegative to HIV seropositive status, indicating a new infection) in a prospective cohort of HIV discordant (one partner HIV+, the other HIV-) cohabiting couples in Lusaka, Zambia. Couples are followed up every three months. The investigator was interested in computing seroconversion rates during specified time intervals: (1) during each 3-month follow-up interval, (2) during progressively longer periods (e.g., 0-6 months, 0-9 months, etc.) and 3) for specified combinations of these same three month intervals (e.g., from month 6 to month 12). The program consists of three macros implementing the three calculations described above. Total person months and number of seroconvertors are calculated by sex and for both sexes combined among at-risk (i.e., HIV-) subjects. Seroconversion rates are expressed as the number of seroconvertors per 100 person-years of follow-up. Confidence intervals for seroconversion rates are calculated using the GAMINV function in SAS to determine the upper and lower limits for the number of seroconvertors (which is assumed to follow a Poisson distribution), then divided by the total person months of follow-up. This program can be easily modified to calculate incidence rates of other outcomes and can accommodate variable follow-up intervals.

INTRODUCTION
The Zambia-UAB HIV (Human Immunodeficiency Virus) Research Project is a prospective follow-up study of HIV discordant (one partner HIV+, the other HIV-) cohabiting heterosexual couples in Lusaka, Zambia. 1 The objectives of this study are 1) to determine whether voluntary HIV testing and counseling increases condom use and decreases the incidence of HIV in initially HIV- partners, and 2) to assess self-reported sexual behavior (including condom use) and biological markers of exposure in these couples. 1-3 To identify and recruit eligible couples, the project established a center for voluntary HIV counseling and confidential serological testing for cohabiting couples in Lusaka, Zambia. Eligibility criteria included being engaged in a sexual relationship and cohabiting for at least six months. 1 Age at recruitment must have been <= 48 years for a woman and <= 65 years for a man. 1 Eligible recruits had to reside in Lusaka, Zambia, be free of overt psychiatric disturbance and willing to complete required procedures. Assessment of serologic HIV status was made at entry and during follow up by employing a rapid screening assay followed by a confirmatory assay when indicated (i.e., to confirm an initially positive test, or to confirm the seronegative status of the partner of a seropositive subject). 1

When HIV seroconversion (i.e., the transition from HIV seronegative to HIV seropositive status) was detected, DNA sequencing was used to confirm within-couple transmission, i.e., that the virus transmitted to the seroconvertor was the virus hosted by the HIV+ partner. 1 Follow-up data were available for 1011 eligible couples. The earliest baseline visit date was on February 6, 1995 and the last follow-up visit attended was on November 30, 1998. The maximum length of follow-up for HIV- subjects for these analyses was 45 months; 107 HIV seroconversions occurred during the study period of which 94 had DNA sequencing data available. 1 Within-couple transmission was confirmed for 82 (87%) of 94 seroconverting couples with DNA sequencing data, while the other 12 HIV seroconvertors were assumed to have acquired HIV from an HIV+ partner who was not a member of the discordant couple. 1 The 13 HIV seroconvertors with missing DNA sequencing data were assumed to be epidemiologically linked for the purpose of this analysis. 1 A SAS program was designed to monitor seroconversion and the transmission of HIV within the couples. Two SAS data step procedures calculate HIV seroconversion and transmission rates for the entire follow-up period of the cohort. The investigator was also interested in computing seroconversion and transmission rates during specified time intervals: 1) during each 3-month follow-up interval, 2) during progressively longer periods (e.g, 0-6 months, 0-9 months, etc.) and 3) for specified combinations of these same three month intervals (e.g., from month 6 to month 12). The program consists of three macros implementing the three calculations described above. Total person months and number of seroconvertors are calculated by sex and for both sexes combined among at-risk (i.e., HIV-) subjects. Seroconversion rates are expressed as the number of HIV seroconvertors per 100 person-years of follow-up. Transmission rates are expressed as the number of HIV seroconvertors per 100 person-years of follow-up after eliminating seroconvertors in whom within-couple transmission could not be confirmed from the numerator of the rates. Confidence intervals for seroconversion rates are calculated using the GAMINV function in SAS to determine the upper and lower limits for the number of seroconvertors (which is assumed to follow a Poisson distribution), then divided by the total person months of follow-up. This program can be easily modified to calculate incidence rates of other outcomes and can accommodate variable follow-up intervals.

COHORT DATA SET
For all segments of this program which calculate seroconversion or transmission rates, the input data set to the initial DATA step is the cohort data set with one observation for each HIV- member of a couple. There are 5 variables in the data set (all date variables are SAS date values):
PROC SORT DATA=newdata.serrate OUT=serrate;
BY numsuj; RUN;

DATA convrate;
SET serrate END=last;
IF _N_=1 THEN DO;
NUMCONV=0; TOTPDAY=0; TFEMPDAY=0;
NMALCONV=0; TMALPDAY=0;
PDAY=0; FEMPDAY=0; MALPDAY=0;
END;

*** All subjects;
PDAY=LVDATE-FSTDATE;
IF PDAY=0 THEN DELETE;
TOTPDAY = TOMPDAY + PDAY;

IF STDYSTAT IN (1,2) THEN NUMCONV = NUMCONV + 1;

*** Female subjects;
IF CJ='+-' THEN FEMPDAY=LVDATE-FSTDATE;
TFEMPDAY = TFEMPDAY + FEMPDAY;

IF STDYSTAT IN (1,2) AND CJ='+-' THEN NFEMCONV= NFEMCONV + 1;

*** Male subjects;
IF CJ='+-' THEN MALPDAY=LVDATE-FSTDATE;
TMALPDAY = TMALPDAY + MALPDAY;

IF STDYSTAT IN (1,2) AND CJ='+-' THEN NMALCONV = NMALCONV + 1;

IF LAST THEN DO;
CONVRATE=(NUMCONV/(TOTPDAY/365.25)) * 100;
FEMCRATE= (NFEMCONV/(TFEMPDAY/365.25)) * 100;
MALCRATE = (NMALCONV/(TMALPDAY/365.25)) * 100;
OUTPUT;
END;

RETAIN NUMCONV NFEMCONV NMALCONV TOTPDAY TFEMPDAY TMALPDAY;

PROC PRINT DATA=convrate NOOBS;
VAR NUMCONV TOTPDAY TFEMPDAY TMALPDAY;
TITLE1 'Seroconversion rates expressed per 100 person years of follow-up';
RUN;

The second data step calculates 95% confidence intervals for the seroconversion rates. The program employs the known statistical relation between the cumulative distribution of a Poisson random variable and the Chi-square distribution, which in turn is a member of the gamma family of statistical distributions. In this case, the GAMINV(a,b) function in SAS returns the value of the mean of the Poisson variate that marks the cumulative probability “a” (in this program, 2.5%) of observing at most k-1 events when b=k observed events; or the cumulative probability “a” (in this program, 97.5%) of observing at least k events when b=k+1. Thus, the SAS statement “LO=GAMINV(0.025,5);” returns the lower 95% confidence limit for 5 observed events, whereas the statement “HI=GAMINV(0.975,6);” returns the corresponding upper 95% confidence limit. Next, the two boundaries computed using the procedure described above are divided by the total person days of follow-up (which is
assumed to be a constant with no random variability, divided by 365.25 and multiplied by 100 to obtain the 95% confidence boundaries of the seroconversion rate (per 100 person-years).

** Checks for subjects with follow-up time rounding to macro the nearest 3 month-visit value for the length of follow-up in numbers. The program creates the variable LASTVMON as integer values for the consecutive 3 month follow-up visit NUMMON is repeatedly multiplied by 3 generating 3 month from 1 to 17. Within this macro, the macro variable numeric integer value which for this cohort ranges in value statement with the %EVAL function converts NUMMON to a parameter NUMMON as a string value. The %LET macro program statement reads the value of the macro subjects in the cohort. When the macro is invoked the The first macro was designed to calculate seroconversion RATES FOR 3-MONTH FOLLOW-UP INTERVALS if (stdystat = 1 then numconv = numconv + 1; if (stdystat = 1 and cj='+=' then nfemconv = nfemconv + 1; if (stdystat = 1 and cj='-=') then nmalconv = nmalconv + 1; RATES FOR 3-MONTH FOLLOW-UP INTERVALS The first macro was designed to calculate seroconversion rates during each 3 month follow-up interval for at-risk HIV-subjects in the cohort. When the macro is invoked the macro program statement reads the value of the macro parameter NUMMON as a string value. The %LET statement with the %EVAL function converts NUMMON to a numeric integer value which for this cohort ranges in value from 1 to 17. Within this macro, the macro variable NUMMON is repeatedly multiplied by 3 generating 3 month integer values for the consecutive 3 month follow-up visit numbers. The program creates the variable LASTVMON as the nearest 3 month-visit value for the length of follow-up in months using the ROUND function. PMON is the number of person-months accumulated for that subject within the 3 month follow-up interval of interest. If the value of LASTVMON is greater than NUMMON multiplied by 3 then the value of PMON equals 3. When LASTVMON equals NUMMON multiplied by 3 then the person months value for these subjects is the total length of follow-up in months minus the integer value for the prior 3 month visit. This provides a more accurate accumulation of follow-up time for subjects whose end of follow-up date may have been less than or greater than the 3 month visit for which we are calculating seroconversion rates since the prior 3 month follow-up visit. The cumulated variables output from this data set for total person months are TOTPMON for all subjects, TFEMPMON for females and TMALPMON for males. Seroconvertors are counted when STDYSTAT equals 1 or 2 and when the 3 month visit when seroconversion occurred (LASTVMON) equals the 3 month visit interval selected for the rate calculations (NUMMON multiplied by 3). Cumulated # of seroconvertors have the same names as before; total number of seroconvertors: NUMCONV, total female seroconvertors: NFEMCONV and total male seroconvertors: NMALCONV. When processing the last record of the input data set, seroconversion rates for the 3-month visit interval are computed by dividing total seroconvertors by total person months divided by 12 and multiplying by 100. The 3-month interval seroconversion rates are expressed as the number of seroconvertors per 100 person-years of follow-up. Three month visit interval seroconversion rates for males and females are also computed. All variables for rates, total # seroconvertors and total # person days are output to a temporary data set. Ninety-five percent confidence intervals for the seroconversion rates are calculated as before except total person months are divided by 12 because we are no longer using a measurement in person-days. The form delimiter option separates 3 month interval rates by asterisks on the printed output page instead of generating a new page for each new 3 month interval. Modifications for transmission rates are similar to the changes noted previously.

** data conf; set convrate; ** all seroconvertors; lo=gaminv(.025,numconv); hi=gaminv(.975,numconv+1); *** # seroconvertors+1; rloi=(lo/(totpday/365.25))*100; rhii=(hi/(totpday/365.25))*100; *** male seroconvertors; mlo=gaminv(.025,nmalconv); mhi=gaminv(.975,nmalconv+1); *** # seroconvertors+1; mrloi=(mlo/(tmalpday/365.25))*100; mrhii=(mhi/(tmalpday/365.25))*100; *** female seroconvertors; flo=gaminv(.025,nfemconv); fhi=gaminv(.975,nfemconv+1); *** # seroconvertors+1; frloi=(flo/(tfempdny/365.25))*100; frhii=(fhi/(tfempdny/365.25))*100; run; proc print data=conf noobs; var lo hi rlo rhi flo fli mlo mroi mhi; title1 'Confidence intervals for # seroconvertors and seroconversion rate'; title2 'Both sexes, male and female seroconvertors'; title3 'Zambia discordant couple study'; run;
*** All subjects;
if lastvmon >= %eval(&nummon*3) then do;
if lastvmon>=%eval(&nummon*3) then pmom=3;
else if lastvmon=%eval(&nummon*3)
then pmom=
( ((lvdate-fstdate)/30.3)-%eval((&nummon*3)-3));
end;
totpmom = totpmon + pmom;
if stdystat in (1,2) and lastvmon = %eval(&nummon*3)
then numconv = numconv + 1;

*** Female subjects;
if cj='+' and lastvmon >= %eval(&nummon*3) then do;
if lastvmon>=%eval(&nummon*3) then fempmon=3;
else if lastvmon=%eval(&nummon*3)
then fempmon=
( ((lvdate-fstdate)/30.3)-%eval((&nummon*3)-3));
end;
tfempmon = tfempmon + fempmon;
if stdystat in (1,2) and cj='+' 
and lastvmon = %eval(&nummon*3)
then nfemconv = nfemconv + 1;

*** Male subjects;
if cj='+' and lastvmon >= %eval(&nummon*3) then do;
if lastvmon>=%eval(&nummon*3) then malpmon=3;
else if lastvmon=%eval(&nummon*3)
then malpmon=
( ((lvdate-fstdate)/30.3)-%eval((&nummon*3)-3));
end;
tmalpmon = tmalpmon + malpmon;
if stdystat in (1,2) and cj='+' 
and lastvmon = %eval(&nummon*3)
then nmalconv = nmalconv + 1;
if last then do;
convrate = (numconv/(totpmom/12))*100;
femcrate = (nfemconv/(tfempmon/12))*100;
malconrate = (nmalconv/(tmalpmon/12))*100;
%let titmom=%eval(&nummon*3);
output;
end;
return numconv totpmon nfemconv nmalconv
tfempmon tmalpmon ;
run;

proc print data=conf3mos noobs;
var lo hi rloi rhii flo fhi frloi frhii mlo mhi mrloi mrhii;
proc print data=conf3mos noobs;
var lo hi rloi rhii flo fhi frloi frhii mlo mhi mrloi mrhii;
if _N_=1 then do;
nnumconv=0; totpmon=0; nfemconv=0; tfempmon=0;
nmalconv=0; tmalpmon=0;
end;
%macro convr(nummon);
data conv3mon;
set conv3mos;
* all seroconvertors;
lo=gaminv(.025,numconv);
hi=gaminv(.975,numconv+1); *** # seroconvertors+1;
frlo=(lo/(totpmom/12))*100;
frhi=(hi/(totpmom/12))*100;
* male seroconvertors;
mlo=gaminv(.025,nmalconv);
run;
%macro convr(nummon);
data conv3mon;
set conv3mos;
* all seroconvertors;
lo=gaminv(.025,numconv);
hi=gaminv(.975,numconv+1); *** # seroconvertors+1;
frlo=(lo/(totpmom/12))*100;
frhi=(hi/(totpmom/12))*100;
* male seroconvertors;
mlo=gaminv(.025,nmalconv);

RATES FOR PROGRESSIVELY LONGER INTERVALS
The second macro was designed to calculate seroconversion rates during progressively longer periods (e.g. 0-3 months, 0-9 months, 0-12 months, etc.) for at-risk HIV- subjects in the cohort. Changes to the existing macro to achieve this task are described below. PMON is the number of person-months accumulated for each subject from baseline up to the 3 month follow-up visit of interest. An iterative DO loop is used to accumulate additional follow-up time to the value of PMON with successive 3 month intervals. The loop always begins with 1 and ends with the value of NUMMON. If the value of LASTMON is greater than the index variable i multiplied by 3 then the value added to PMON equals 3. Otherwise when LASTMON equals i multiplied by 3 then the amount of follow-up time added to PMON in months for these subjects is the total length of follow-up in months minus the integer value for the prior 3 month visit. Seroconvertors are counted when STDYSTAT equals 1 or 2 and when the 3 month visit when seroconversion occurred (LASTVMON) is between 0 and the 3 month visit interval selected for the rate calculations (NUMMON multiplied by 3).
*** All subjects:
do i=1 to &nummon;
  if lastvmon >= i*3 then do;
    if lastvmon > i*3 then pmon=pmon + 3;
    else if lastvmon = i*3 then
      pmon=pmon + (((lvdate-fstdate)/30.3)-(i*3)-3));
  end;
  end;
totpmon = totpmon + pmon;
if stdystat in (1,2) and 0 < lastvmon <= %eval(&nummon*3)
  then numconv = numconv + 1;
*** Female subjects:
do i=1 to &nummon;
  if cj='+' and lastvmon >= i*3 then do;
    if lastvmon > i*3 then fempmon=fempmon + fempmon + 3;
    else if lastvmon = i*3 then
      fempmon=fempmon + (((lvdate-fstdate)/30.3)-(i*3)-3));
  end;
  end;
tfempmon = tfempmon + fempmon;
if stdystat in (1,2) and cj='+' and 0 < lastvmon <= %eval(&nummon*3)
  then nfemconv = nfemconv + 1;
*** Male subjects:
do i=1 to &nummon;
  if cj='-' and lastvmon >= i*3 then do;
    if lastvmon > i*3 then malpmon=malpmon + 3;
    else if lastvmon = i*3 then
      malpmon=malpmon + (((lvdate-fstdate)/30.3)-(i*3)-3));
  end;
  end;
tmalpmon = tmalpmon + malpmon;
if stdystat in (1,2) and cj='-' and 0 < lastvmon <= %eval(&nummon*3)
  then nmalconv = nmalconv + 1;
  if last then do;
    convrate = (numconv/(totpmon/12))*100;
    femcrate = (nfemconv/(tfempmon/12))*100;
    malcrate = (nmalconv/(tmalpmon/12))*100;
    %let titlmon= %eval(&nummon*3);* # seroconvertors+1;
    output;
  end;
  end;
retain numconv totpmon nfemconv nmalconv
tfempmon tmalpmon ; run;
proc print data=conv3mon noobs;
  var numconv totpmon convrate femcrate malcrate;
title1 "Seroconversion rates expressed per 100 person years of follow up";
title2 "for up to and including &titlmon months";
title3 "*"; title4 "*"; run;

rates for specific combinations of 3-month follow-up intervals
The third macro was designed to calculate seroconversion rates for specified combinations of these same 3 month follow-up intervals (e.g. from month 6 to 12) for at-risk HIV subjects in the cohort. Specific modifications to the existing macro for this purpose are explained in detail below. During invocation of the macro, the macro program statement reads the value of the macro parameters NUMMON and ENDMON as string values. The %LET statement with the %EVAL function converts NUMMON and ENDMON to numeric integer values which for this cohort ranges in value from 1 to 17. ENDMON can only be a number equal to or greater than the value of NUMMON. PMON is the number of person-months accumulated for each subject for selected visit month intervals. A DO loop iterates for values of i, the index variable, beginning with the macro variable NUMMON and ending with the macro variable ENDMON. Within this DO loop, follow-up time is added to the value of PMON with successive 3 month intervals. The index variable multiplied by 3 is compared to the value of the variable LASTVMON which is the nearest 3 month visit value for the subject's entire length of follow-up time in months using the ROUND function. If the value of LASTVMON is greater than the index variable i multiplied by 3 then the value added to PMON equals 3. Otherwise when LASTVMON equals i multiplied by 3 then the amount of follow-up time added to PMON in months for these subjects is the total length of follow-up in months minus the integer value for the prior 3 month visit. Seroconvertors are counted when STDYSTAT equals 1 or 2 and when the 3 month visit when seroconversion occurred (LASTVMON) is between the macro variable NUMMON multiplied by 3 and the macro variable ENDMON multiplied by 3.
CONCLUSION

Computing incidence rates, such as the seroconversion and transmission rates that are the epidemiologic measures of interest in this study, is an important programming aspect in the analysis of all follow-up studies in which new outcomes occur during an extended observation period. SAS provided an excellent work environment for data entry, data manipulations and other statistical analyses of predictors of seroconversion in this study, but did not provide specific support for the computation of incidence rates. Given the versatility of the SAS programming language and the availability of advanced statistical function, however, it was relatively easy to write the program presented in this paper which includes both rate calculation and statistical inference about rate estimates. This program can be easily modified to calculate incidence rates of other outcomes and can accommodate variable follow-up intervals. The program described in this paper addresses the important issue of accommodating variable follow-up intervals. The program can be easily modified to calculate incidence rates of other outcomes and can accommodate variable follow-up intervals. The program presented in this paper addresses the important issue of accommodating variable follow-up intervals.

Computing incidence rates, such as the seroconversion and transmission rates that are the epidemiologic measures of interest in this study, is an important programming aspect in the analysis of all follow-up studies in which new outcomes occur during an extended observation period. SAS provided an excellent work environment for data entry, data manipulations and other statistical analyses of predictors of seroconversion in this study, but did not provide specific support for the computation of incidence rates. Given the versatility of the SAS programming language and the availability of advanced statistical function, however, it was relatively easy to write the program presented in this paper which includes both rate calculation and statistical inference about rate estimates. This program can be easily modified to calculate incidence rates of other outcomes and can accommodate variable follow-up intervals. The program described in this paper addresses the important issue of accommodating variable follow-up intervals.

Computing incidence rates, such as the seroconversion and transmission rates that are the epidemiologic measures of interest in this study, is an important programming aspect in the analysis of all follow-up studies in which new outcomes occur during an extended observation period. SAS provided an excellent work environment for data entry, data manipulations and other statistical analyses of predictors of seroconversion in this study, but did not provide specific support for the computation of incidence rates. Given the versatility of the SAS programming language and the availability of advanced statistical function, however, it was relatively easy to write the program presented in this paper which includes both rate calculation and statistical inference about rate estimates. This program can be easily modified to calculate incidence rates of other outcomes and can accommodate variable follow-up intervals. The program described in this paper addresses the important issue of accommodating variable follow-up intervals.
monitoring change in seroconversion or transmission rates for variable lengths of follow-up time since the induction of intervention programs consisting of HIV testing and counseling to encourage safer sex in HIV discordant couples.

REFERENCES


SAS and SAS/STAT software are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are registered trademarks or trademarks of their respective companies.

AUTHOR CONTACT INFORMATION

Ilene Brill
Department of Epidemiology and International Health
University of Alabama at Birmingham
Mortimer Jordan Hall, Room 108
1530 3rd Avenue South
Birmingham, AL 35294-2010
Work Phone: (205) 975-5359
Fax: (205) 975-2435
Email: ibrill@ms.soph.uab.edu