Received: Added at production | Revised: Added at production | Accepted: Added at production

DOI: xxx/xxxx

ARTICLE TYPE

Survival Model Analysis of Tuberculosis Treatment among Patients with Human Immunodeficiency Virus Coinfection

Rong Duan

[1]Biostatistics Department, University of California, Davis, CA, USA

Correspondence
Rong Duan, 186 Orchard Park Drive. Email: rduan@ucdavis.edu

Present Address
186 Orchard Park Drive, Davis, CA

Abstract

Tuberculosis (TB) with human immunodeficiency virus (HIV) coinfection is the highest clinical epidemiology and public health issue. This study aims to assess the survival rate of HIV infection among TB patients and the risk factors of death during the retro of directly observed treatment, short-course program. This study is conducted to compare the survivorship between TB/HIV patients for 8 months program duration. The Cox proportional-hazards regression model were used to establish the hazard ratio (HR) of death for each variable at baseline and estimate the risk factors effect among TB patients. The findings revealed the risk of death was significantly higher in HIV-positive TB patients. The probability of TB patients surviving is significantly decreased with some factors such as age, smoking, and alcohol use.

KEYWORDS:
Tuberculosis (TB); HIV; survival analysis; hazard ratio; Cox regression model

1 | INTRODUCTION

Tuberculosis (TB) and human immunodeficiency virus (HIV) are both active infectious diseases for morbidity and mortality in the world. TB is a universal endemic, with an estimated 9 million newly diagnosed cases and at least 1.5 million deaths by the year 2013. The global encumbrance of TB and HIV coinfection falls heavily on poor and low-income countries, including those of Sub-Saharan African countries. Unsuccessful treatment of TB/HIV patients was found associate with some com-morbidity diseases such as diabetes, liver diseases, renal failure, hepatitis, and silicosis. Also, TB/HIV-positive patients also require more treatment measures and procedures for HIV, such as trimethoprim-sulfamethoxazole (co-trimoxazole) prophylaxis and antiretroviral treatment (ART). These treatments should not be long to initiate after the start of the treatment of anti-TB.

This study explores the survival probability and hazard rate at a different time based on our data set. In addition, this study wants to figure out the influence of age, alcohol and smoking on survival probability, and in each stratified covariate level whether the HIV-positive group still has a significantly higher hazard rate. The aim of this TB with HIV coinfection studies is to minimize morbidity, mortality, default, relapse, and prevent DR-TB and outcomes of TB treatment.

2 | METHOD

2.1 | Study design and population

This was a retrospective cohort study conducted on record review of TB patients with HIV coinfection admitted for treatment. Our source population was all the TB patients registered for anti-TB treatment and those TB patients who started ART after tested positive to HIV. So we implement the same method for within each group to minimize the bias. The sample size was determined using stratified random sampling technique. HIV-positive and HIV-negative strata comprised 660 and 979 TB patients, respectively. TB patients of 450 each HIV positive and HIV negative were randomly selected from each strata. The death of patients is referred to as failures and survived or defaulted patients were censored.

2.2 | Data Analysis

The Kaplan–Meier curve and log-rank test were used to estimate survival probability between HIV-positive and negative TB patients. Stratified test is conducted in alcohol and smoking group to find whether the hazard rate of HIV-positive group is significantly different with HIV-negative group. To determine the hazard ratio (HR) of death for the baseline predictors, Cox proportional hazards regression model was applied. Cox-Snell residual method is used to check the overall fit of the Cox model. The assumption of proportional hazard of HIV status, age, alcohol and smoking, is checked by Schoenfeld residuals plots and the corresponding coefficient test.

2.3 | Data simulation

To simulate the data, this research conduct the following procedures. First, since it is balanced data and the number of patients in each negative-HIV and positive-HIV is 455, the HIV group covariate can be simply simulated by repeated indicators. For alcohol and smoking covariates, we have the proportions of each level in the HIV groups, so we can randomly sample the value within each HIV group with the probability specified in the paper. Age covariate is assumed from a normal distribution, mean and standard deviation is used to simulate the values. Second The original paper uses the Cox proportional hazard model to estimate the hazard ratio of each covariate. Thus, the time to event is assumed from Weibull distribution. Utilize the simulated values from the simulated data before, we can get the time by sampling from the Weibull distribution. Lastly, for the event indicator, we assume a random censor happened during the study time. If the event time is larger than the cencored time or the largest study time (900 days), then the case is censored, otherwise, the event happened. Because the parameter from the weibull distribution is unknown, some adjustments are needed with the time variable and censored indicator. Combined with the life table offered in the paper, the final data in this study is adjusted as close to the data from the life table as possible. Combining the procedure above, we can get the simulated data.

3 | RESULTS

3.1 | Descriptive statistics of the cohort

IIn this study, 910 TB patients (455 HIV negative and HIV positive each) were extracted from the TB treatment record used in the hospitals. From this study, the mean age and standard deviation is 38 and 5, respectively. Alcohol use in negative-HIV group is 65% and in positive-HIV group is 74%. Smoking was predominant among patients with 52.5% HIV-negative and 74.2% HIV-positive TB patients, respectively [Table 1].

3.2 | Time to death and survival probabilities of tuberculosis patients

It was observed that 229 deaths occurred throughout the treatment of TB. HIV-positive TB patients had the majority of death 146 (64%). Specifically, 83 (18.2%) of HIV negative and 146 (32.1%) of HIV-positive TB patients died and were treated in the analysis as event failure, whereas the remaining part of HIV-negative 372 (81.8%) and HIV-positive 309 (67.9%) TB patients were censored. The mean and median estimation of time death for the TB patients who did not survive during TB treatment was 246 and 204 days, respectively.

Tabella 1 Baseline information of TB/HIV co-infected patients

Characteristics	Negative-HIV		Positive-HIV	
	Death	Cencered	Death	Censored
HIV	83 (18.2%)	273(81.8%)	146 (32.1%)	309(67.9%)
Alcohol Use	63(21.4%)	232(78.6%)	120(35.9%)	214(64.1%)
Smoking	61(26.2%)	172(73.8%)	125(36.7%)	216(63.3%)

The survival probability of the HIV groups are compared with KM estimators[Figure 1]. It shows the survival probability of positive-HIV group is lower compared with negative-HIV group. And the hazard rate of positive-HIV group in Epanechnikov Kernel-Smoothed Hazard plot is consistently higher than the negative-HIV group [Figure 1]. The equality between these two groups are tested by log-rank test (test statistics = 26.11, df = 1), and the p-value (<.0001) indicates the positive-HIV and negative-HIV have a different effect to hazard rate of death. Stratified test for the equality of HIV groups is conducted by alcohol strata (log-rank test = 22.44, df = 1, P-value < 0.0001) and smoking strata (log-rank test = 13.77, df = 1, P-value = 0.0002).

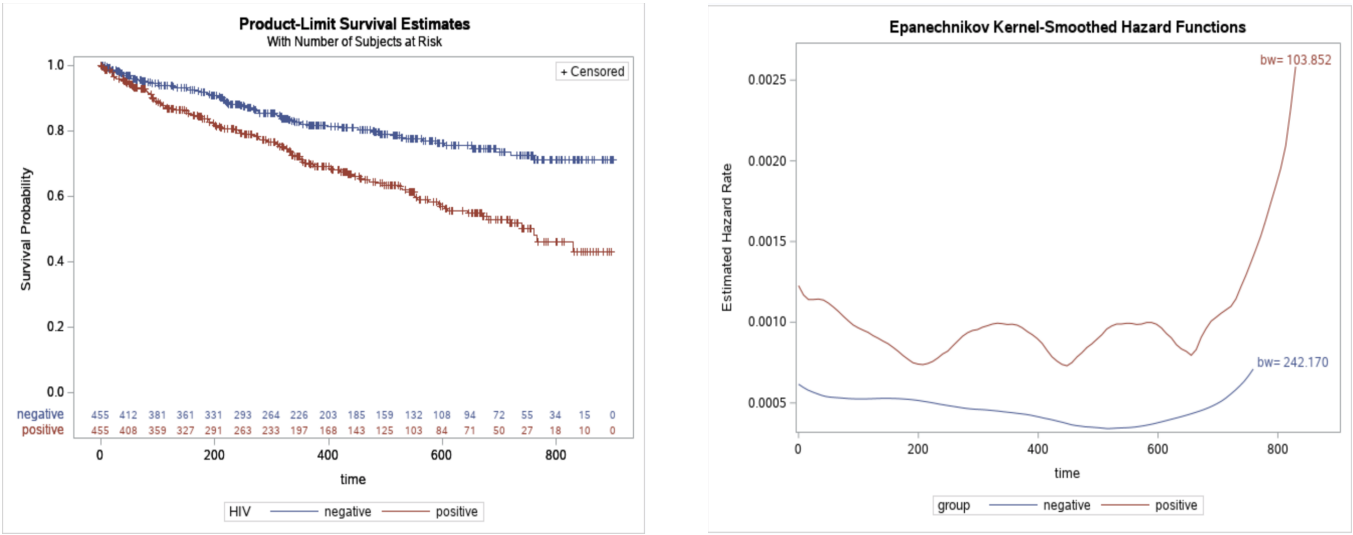


Figura 1 : KM survival plot (Left) and Smooth Hazard plot (Right)

3.3 | Risk of death factors of TB patients with HIV coinfection

Cox proportional hazard regression model is used to analyze the association between the survival of TB patients and the covariates. The overall fit is checked by Cox-Snell residual plot [Figure 2]. The plot follows the 45 degree line, it is an indication that the model fits well. Cox PH assumptions were checked by plotting Schoenfeld residual plot to test for independence between time and residual before fitting the covariates into the model. The result show that age, alcohol use, and HIV status are satisfied the PH assumption [Figure 3]. However, the PH assumption of smoking is violated, while in the original paper, the PH assumption of smoking is satisfied. Considering coincidence of the random sampling method and non-severe violation, we keep smoking in the Cox regression model.

The Cox model with the estimated coefficients [Table 2]and the arbitrary hazard function $h_0(t)$ listed as bellow:

$$h(t, \mathbf{Z}) = h_0(t) \exp(0.44 \times HIV(Positive) - 0.04 \times Age + 0.76 \times Alcohol + 1.00 \times Smoking),$$

The hazard ratio of positive-HIV to negative-HIV is 1.556, which means the hazard rate of patients in positive-HIV status is 55.6% higher than the negative-HIV status. The hazard ratio of age is 0.957, which means one additional year in age can lower

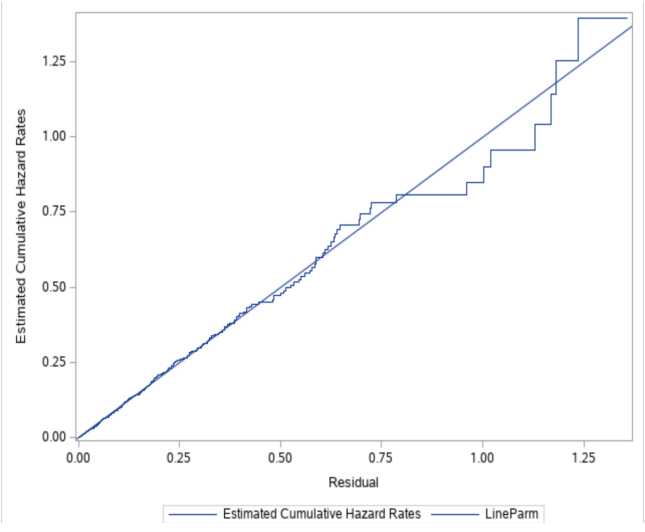
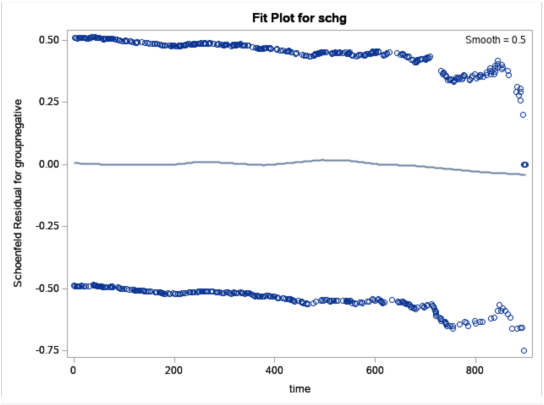
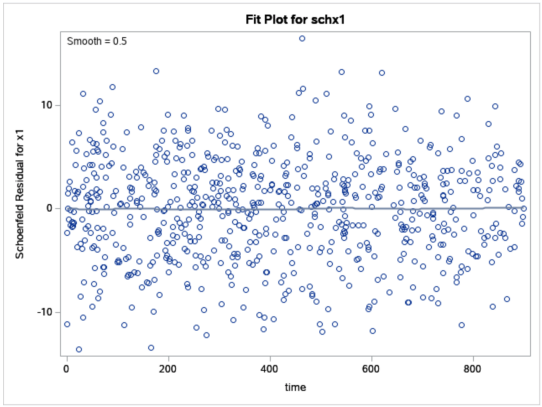


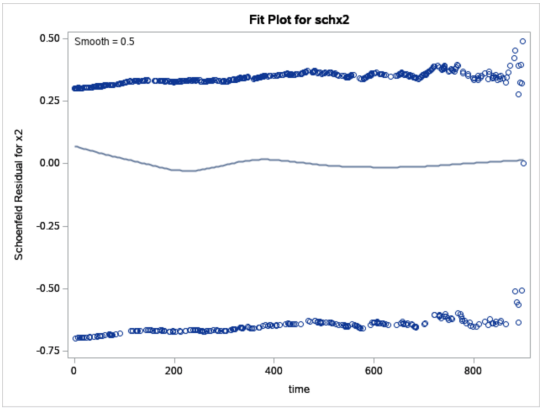
Figura 2 Cox-Snell residual plot



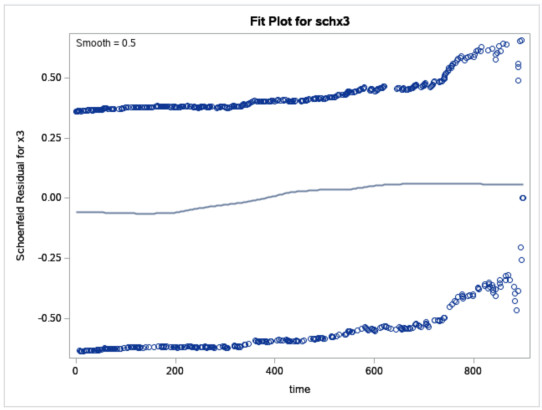
(a) HIV



(b) Age



(c) Alcohol



(d) Smoke

Figura 3 Schoenfeld residual plot

the hazard rate by 4.7%. The hazard ratio of alcohol is 2.137, which means the use of alcohol can increase the hazard rate by 113.7% compared with non-alcohol use. Smoke also indicates an increased hazard rate, with the hazard ratio of 2.724, which

Tabella 2 Cox PH model .

Covariate	DF	Estimated parameter	SE	P-value	HR
HIV	1	0.44191	0.1402	0.0016	1.556
Age	1	-0.0439	0.0135	0.0011	0.957
Alcohol	1	0.7595	0.1659	<.0001	2.137
Smoke	1	1.0020	0.1722	<.0001	2.724

indicates the hazard rate in the smoking group is 2.724 times of the hazard rate in the non-smoking group.

4 | DISCUSSION

TB is still one of the major causes of preventable death in the world. The TB prevalence has largely increased with HIV coinfection and rated as one of the top causes of overall maternal mortality. This study was used to assess the outcome of survival of TB patients with HIV coinfection during the treatment duration by evaluating the difference between the survival time of HIV negative and positive TB patients. Similarly, the study has explained the difference of the time to death with HIV coinfection. The result shows that positive-HIV can significantly increase the hazard rate of death with or without the adjustment of other covariates. Some covariates are found also highly related with death. Increasing in age can lower the hazard rate, while alcohol and smoking will higher the hazard rate.

Some other fitures are also found corelated with hazard rate. In the original paper, Weight is seen as predictors of death. through the Cox regression model. The risk of diabetes patients is also higher compared to patients without diabetes concomitant. This is due to weak immune system. Besides, pregnancy increased exponentially the risk of death among TB patients, and it is significantly more severe particularly when the diagnosis is confirmed late in pregnancy.

Some limitations exists in this simulated research. Because without the access of the original data, although the simulation result is similar to the results in the original research, the estimated hazard rate in each level is biased. Also, with the limited information, some other covariates can not be explored, and other hazard ratio is also biased due to lacked variables. Last, in the original research, some covariates are deleted because of violating PH assumption. Some remedy methods, such as stratified PH model can be applied in this situation. Thus, the result might be improved with additional information about patients.

5 | CONCLUTION

This study showed the difference in hazard rate of death between positive-HIV and negative-HIV status. The result is established by the survival probability and hazard rate between each HIV group. The positive-HIV lead to a higher risk of death with or without other covariates adjustment. Some features can also influene the hazard rate. Additional age lowers the hazard rate, while smoking and alcohol use increase the hazard rate.

APPENDIX

Simulation Codes (R)

```
library(MASS)

sim_cox<- function(N, beta, censor.right)
{
  # N = Total sample size
  # beta = PH coefficients
```

```
  # Randomization to Neigtive-HIV or Positive-HIV
  A <- c(rep(0,455),rep(1,455)) #Positive-indecator:1

  # Generate continuous covariate: AGE
  X1 <- rnorm(N,38,5)

  # Generate bivariate covariates: Gender, Smoking. mutually indepedent
  # Alcohol (use:1)
  X2 <- c(sample(x=c(0, 1), size=N/2, replace=TRUE, prob=c(0.35, 0.65)),
          sample(x=c(0, 1), size=N/2, replace=TRUE, prob=c(0.26, 0.74)))
  # Smoking (use:1)
  X3 <- c(sample(x=c(1, 0), size=N/2, replace=TRUE, prob=c(0.52, 0.48)),
          sample(x=c(1, 0), size=N/2, replace=TRUE, prob=c(0.74, 0.26)))

  # generate underlying event time
  T <- rweibull(n=N, shape=1, scale = 1000*exp(beta[1]*A+beta[2]*X1+beta[3]*X2+beta[4]*X3))

  # censoring times
  ctime <- runif(N, min=0, max=censor.right)

  # follow-up times and event indicators
  time <- pmin(T, ctime, censor.right)
  censor <- as.numeric(T>ctime | T>censor.right)

  # data set
  data.frame(id=1:N,
             group=A,
             x1=X1,
             x2=X2,
             x3=X3,
             time=time,
             censor=censor,
             status=1-censor)
}
mydata=sim_cox(N=910, beta=c(-0.45,0.0456,-0.63,-0.99), censor.right=900)
```

Data Analysis: SAS

```
*Import data;
proc import datafile='/folders/myfolders/222-lab/project/BST222-project2.csv'
out=TB dbms=csv replace;
datarow=2;
getnames=yes;
run;

*Add lable;
data TB(drop=var1);
set TB;
label group=HIV
x1=age
x2=alchol
x3=smoking
;
run;
*Convert the format of the dataset;
proc format;
value group 1='positive' 0='negative';
value ac 0='None' 1='Use';
value smoking 0='No' 1='Yes';
run;
```

```
data TB;
set TB;
format group group. x2 ac. x3 smoking.;
run;


*Test equality over hormon group;
proc lifetest data=TB plots=(survival(atrisk=0 to 900 by 50) hazard) notable ;
time time*censor(1);
strata group/test=(logrank TARONE PETO MODPETO
FLEMING(0,1)    );
run;


* PH model;
proc phreg data=TB plots(overlay)=(survival);
class group;
model time*censor(1) = group x1 x2 x3;
run;


* Cox-snell;
proc phreg data = TB;
  model time*censor(1) = group x1 x2 x3 ;
  output out = plot1_1 LOGSURV = logsurv1 /method = ch;  /*-logsurv is the cox-snell residual*/
run;
data plot1_1;
  set plot1_1;
  snell = -logsurv1;
  cons = 1;
run;
proc phreg data = plot1_1;
  model  snell*censor(1) = cons;
  output out = plot1_2 logsurv = logsurv2 /method = ch;
run;
data plot1_2;
  set plot1_2;
    cumhaz = - logsurv2;
run;
proc sort data = plot1_2;
 by snell;
run;
proc sgplot data = plot1_2;
step y=cumhaz x=snell /MARKERFILLATTRS=(color="red");
lineparm x=0 y=0 slope=1; /** intercept, slope **/
  label cumhaz = "Estimated Cumulative Hazard Rates";
  label snell = "Residual";
run;


* Graphical for PH assumption;
```

```
data TB1;
  set TB;
  cons = 1;
run;
* The baseline cumulative hazards are estimated using Breslow's estimator for each stratu;
* plot log(h) function;
proc phreg data = TB1 ;
class group/param=ref;
 model time*censor(1) = cons/rl ; * risk limit: provide CI for harzard ratio;
 strata group;
 output out = base_1 logsurv = ls /method = ch;
 run;

data base_1;
  set base_1 ;
    logH = log (-ls);
        if group=0 then logH1 = logH;
    if group=1 then logH2 = logH;
    proc sort;by time group;
    proc print;var time group logH logH1 logH2 ;
run;

proc sgplot data =base_1;
where logH ne .;
series x=time y=logH /group=group ;
run;

*Plot Log difference function;
proc sort data = base_1;
 by time;
run;
data base_2;
  set base_1;
  retain temp1 temp2;
   if logH1 ~= . then temp1 = logH1;
   if logH2 ~= . then temp2 = logH2;
   diff2v1 = temp2 - temp1;
run;
proc sgplot data =base_2;
step x=time y=diff2v1  ;
lineparm x=0 y=0 slope=0 /lineattrs=(color= red ); /** intercept, slope **/
run;


* Plot log(ht)/log(ht);
proc sort data = base_1;
by time;
run;
data base_3;
  set base_1;
  retain H1 H2;
```

```
  *retain its value from one iteration of the DATA step to the next;
  if group=0 then H1 = -ls;
  if group=1 then H2 = -ls;
   proc print;
   var time group ls H1 H2  ;
run;

proc sgplot data =base_3;
title "Anderson Plot";
series x=H1 y=H2 ;
lineparm x=0 y=0 slope=1.9/lineattrs=(color= red ); /** intercept, slope **/
run;

* Shoenfeld residual for time with grade;
proc phreg data = TB plots=survival ;
class group/param=ref;
 model time*censor(0) =group x1 x2 x3/rl ;
 output out=schoen
 ressch=schg schx1 schx2 schx3  ;
run;

proc loess data = schoen plots=FITPLOT;
model schg=time / smooth=(0.5);
run;
* Shoenfeld residual for group time correlation test;
* p-value is 0.7403 means the ph assumption is not violated;
proc sort data=schoen;by time;run;
data schoen;
set schoen;
retain temp 0;
temp=temp+1;
rank=temp;
proc print ;var group rank;
run;
proc corr data=schoen; var rank schg;run;


* Shoenfeld residual for time with age:x1;
* Both plot and p-value 0.9231 shows the age meet assumption;
proc loess data = schoen plots=FITPLOT;
model schx1=time / smooth=(0.5);
run;
proc corr data=schoen; var rank schx1;run;


* Shoenfeld residual for time with achol;
* Both plot and p-value 0.6353 shows the achol meet ph assumption;
proc loess data = schoen plots=FITPLOT;
model schx2=time / smooth=(0.5);
run;
proc corr data=schoen; var rank schx2;run;
```

```
* Shoenfeld residual for time with smoking;
* Both plot and p-value 0.0090 shows the smoking  does not meet ph assumption;
proc loess data = schoen plots=FITPLOT;
model schx3=time / smooth=(0.5);
run;
proc corr data=schoen; var rank schx3;run;


* time varying covaraite: smoking ---> significant p-value ;
data TB_2;
set TB;
 s_t=time*x3;
 run;

proc phreg data = TB_2 ;
class x3/param=ref;
 model time*censor(0) =x3 s_t/rl ; * risk limit: provide CI for harzard ratio;
 run;

* Graphical for PH assumption---smoking;
data TB1;
  set TB;
  cons = 1;
run;
* The baseline cumulative hazards are estimated using Breslow's estimator for each strata;
* plot log(h) function;
proc phreg data = TB1 ;
class x3/param=ref;
 model time*censor(1) = cons/rl ; * risk limit: provide CI for harzard ratio;
 strata x3 ;
 output out = base_1 logsurv = ls /method = ch;
 run;

data base_1;
  set base_1 ;
    logH = log (-ls);
        if x3=0 then logH1 = logH;
    if x3=1 then logH2 = logH;
    proc sort;by time x3;
run;

proc sgplot data =base_1;
where logH ne .;
series x=time y=logH /group=x3 ;
run;

*Plot Log difference function;
proc sort data = base_1;
 by time;
```

```
run;
data base_2;
  set base_1;
  retain temp1 temp2;
   if logH1 ~= . then temp1 = logH1;
   if logH2 ~= . then temp2 = logH2;
   diff2v1 = temp2 - temp1;
run;
proc sgplot data =base_2;
step x=time y=diff2v1  ;
lineparm x=0 y=0 slope=0 /lineattrs=(color= red ); /** intercept, slope **/
run;


* Plot log(ht)/log(ht);
proc sort data = base_1;
by time;
run;
data base_3;
  set base_1;
  retain H1 H2;
  if x3=0 then H1 = -ls;
  if x3=1 then H2 = -ls;
run;

proc sgplot data =base_3;
title "Anderson Plot";
series x=H1 y=H2 ;
lineparm x=0 y=0 slope=2.7/lineattrs=(color= red ); /** intercept, slope **/
run;
```