Developing a Surrogate Endpoint for AIDS Clinical Trials

Chamindi Kavindya Samarasekara[*] and Marina Roshini Sooriyarachchi

Department of Statistics, Faculty of Science, University of Colombo, Colombo 3, Sri Lanka

*Corresponding Author Email: chamindik@yahoo.com

http://orcid.org/0000-0002-7805-918X

ABSTRACT

When it comes to the process of developing new treatments, the choice of an endpoint is very crucial because this endpoint will be used to assess the effects of the treatments. However the most sensitive and clinically relevant endpoint which is called the 'true endpoint' is difficult to use in a clinical trial because the measurement of the true endpoint can be costly and difficult to measure. In such cases the most feasible solution is to replace the true endpoint by another endpoint termed 'surrogate endpoint' which can be measured earlier and frequently.CD4 and viral loads are used in majority of AIDS clinical trials as surrogate endpoints, however, no surrogate endpoint has yet been shown to be suitable in forecasting the effectiveness of anti-HIV treatments. As a solution, the current study is intended on developing a surrogate endpoint for AIDS based on a combination of variables. This study consists of 16 variables measured in 1151 HIV infected patients. From descriptive statistics, variables CD4 cell count and Karnofsky score were identified as potential candidates for surrogate. However a model with a combination of variables named score consisting of CD4, Karnofsky score and age yielded positive results in the log rank test and conventional statistics. Validation of the scoring model using Prentice's criteria fulfilled all four criteria of Prentice and the model was also successful in identifying the difference between the two treatments. When a comparison was made between CD4 cell count and the combined variable model as possible surrogate endpoints for AIDS, the combined variable model proved to be successful in almost every aspect. Also these results surpassed the results in past similar studies.

Keywords: True endpoint, Surrogate endpoint, score, CD4, Karnofsky score, Prentice's criteria

1. Introduction

Selecting a good surrogate endpoint for Acquired Immune Deficiency Syndrome (AIDS) which can assess the efficacy and the reliability of new drugs and treatments, is one of the major challenges that researchers have been facing throughout the years. There is no clear evidence that current surrogate endpoints CD4 cell count and HIV-1 RNA (viral load) can be reliably used to predict the effectiveness of new treatments. Since no permanent cure or vaccine has been found for Human Immunodeficiency Virus (HIV), there is an increasing pressure from the general public to approve new drugs to the market as quickly as possible, which are based on surrogate endpoints. Therefore, in this research the main focus is on developing a new surrogate endpoint for AIDS, which is a combination of many predictive factors of AIDS with the hope of accelerating the process of new drug development for AIDS. Therefore, the prime objective of this research is to develop a surrogate endpoint to be used in AIDS clinical trials, which can reliably assess the efficacy of the existing treatments.

Although many studies have been done, to develop a surrogate endpoint to be used in AIDS clinical trials, up to now a standard surrogate endpoint for AIDS, which can be used in any instance has not been found. The surrogate endpoints that have been developed up to date are mostly based on HIV-1 RNA and CD4 cell count, where these two variables have been taken separately and so may need complex statistical methods to evaluate their importance. More importantly the majority of surrogate endpoints that have been developed up to now haven't taken the patient's age into consideration, whereas the rate of HIV virus development highly depends on patient's age. This is the significance of this research, as in this research a surrogate endpoint will be developed considering three variables including the patient's age. As we are aiming to develop a composite surrogate endpoint, the statistical analysis is much simpler for the medical community to understand. However, it is important to note that when deriving the combined variable surrogate endpoint, principal component analysis technique was used in order to avoid multicollinearity because the variables in this study are highly correlated.

The data for the study is from a double-blind, placebo-controlled trial that compared the three-drug regimen of indinavir (IDV), open label zidovudine (ZDV) or stavudine (d4T) and lamivudine (3TC) with the two-drug regimen of zidovudine or stavudine and lamivudine in HIV infected patients (Hammer et al., 1997).

The data set is a public domain data set and is also available at the following Wiley's FTP site. http://www.umass.edu/statdata/statdata/data/actg320.txt. The

data set consists of 16 variables measured on 1151 HIV infected patients in the United States and Puerto Rico. Patients were entitled for the study if they had no more than 200 CD4 cells per cubic millimeter and minimum three months of prior zidovudine treatment. The length of the follow up study was 375 days and the time unit used was number of days. Randomization was stratified by CD4 cell count at the time of screening. The main outcome measure was time to AIDS describing event or death.

2. Methodology

In this study descriptive analysis tools such as Kaplan-Meier plots, sensitivity, specificity, attributable proportion were used to analyze the variables and the relationships among variables. Univariate statistics like the log rank test was used to decide whether each variable can be considered as a possible surrogate or not. Molenberghs, Burzykowski and Buyse (2005) gave a framework which summarizes the relationship between the surrogate endpoint and the true endpoint. Table 1 recaps the relationships that surrogate endpoints (S) can have with the true clinical endpoints (T).

Table 1 : Relationship of surrogate endpoint with the clinical endpoint

	T good	T poor	Total
S good	a	b	a + b
S poor	c	d	c + d
Total	a + c	b + d	N

Where T is the true endpoint, S is the surrogate endpoint.
a = number of patients where both S and T provide good disease characterization.
b = number of patients where S is good, but T provides poor disease characterization.
c = number of patients where S is poor, but T provides good disease characterization.
d = number of patients where both S and T provide poor disease characterization.
According to Molenberghs, Burzykowski and Buyse (2005) sensitivity, specificity, relative risk and attributable proportion of the surrogate endpoint for the clinical endpoint can be defined as follows.
Sensitivity (SE) of the surrogate endpoint for the clinical endpoint is defined as
$$SE = a / a + c \qquad (1)$$
Specificity (SP) of the surrogate endpoint for the clinical endpoint is defined as
$$SP = d / b + d \qquad (2)$$

For the surrogate to be useful sensitivity has to be numerically close to 1 and specificity should not be a value which is too low.

The relative risk (RR) is defined as

$$RR = a(c+d)/c(a+b) \qquad (3)$$

The attributable proportion (AP) as

$$AP = SE/(1-1/RR) \qquad (4)$$

For a surrogate endpoint to be a successful one, AP has to be numerically close to 1. Therefore the attributable proportion is considered to be a useful measure to assess the relationship among the surrogate endpoint and the true clinical endpoint (Molenberghs, Burzykowski and Buyse, 2005).

Prentice (1989) defines a surrogate endpoint as "a response variable for which a test of the null hypothesis of no relationship to the treatment groups under comparison is also a valid test of the corresponding null hypothesis based on the true endpoint" (Prentice, 1989). Prentice's definition can be written as an equation as follows using notations -:

$$f(S|Z) = f(S) \Leftrightarrow f(T|Z) = f(T) \qquad (5)$$

Here T is the true endpoint, S is the surrogate end point and Z is the treatment. In the above equation f(X) depicts the probability distribution of random variable X and f(X|Z) depicts the probability distribution of X conditional on the value of Z.

An operational criterion was set by Prentice in order to check whether the triplet (T, S, Z) satisfies the above definition. The four operational criteria can be symbolically written in the following format.

$$f(S|Z) \neq f(S) \qquad (6)$$

In simple words the requirements for the above operational criterion is that the treatment should have significant impact on the surrogate endpoint.

$$f(T|Z) \neq f(T) \qquad (7)$$

The above operational criterion means that treatment should have a significant impact on the true endpoint.

$$f(T|S) \neq f(T) \qquad (8)$$

The meaning of the above operational criterion is that the surrogate endpoint should have a significant impact on the true endpoint.

$$f(T|S,Z) = f(T|S) \qquad (9)$$

The final operational criterion means that the full impact of treatment on the true clinical endpoint is captured by the surrogate endpoint.

Multivariate analysis of variance (MANOVA) is a generalized method of univariate analysis of variance (ANOVA) and is a way for comparing multivariate sample means. As a multivariate technique, it is used when there are two or more dependent variables, and is followed by significance tests containing individual

dependent variables separately. This uses the variance-covariance among variables in testing the statistical significance of the mean differences (French et al.).

The most common statistics are summaries built on the eigenvalues or roots λ_p of the A matrix and they are as follows.

- Pillai-M. S. Bartlett trace,

$$\Lambda_{Pillai} = \sum_{1...p}(\lambda_p/(1 + \lambda_p)) = \text{tr}((I + A)^{-1})$$

(10)

- Lawley-Hotelling trace,

$$\Lambda_{LH} = \sum_{1...p}(\lambda_p) = \text{tr}(A)$$

(11)

- Roy's greatest root

$$\Lambda_{Roy} = max_p(\lambda_p) = \|A\|_\infty$$

(12)

(French et al.)

Principal component analysis (PCA) is a multivariate procedure aimed at reducing the dimensionality of the multivariate data while accounting for as many of the variation in the original data set. This technique is useful when the variables in the data set are highly correlated. Principal components try to transform the original variables to a set of new variables that are linear combination of the variable in the dataset, which are uncorrelated with each other, and ordered according to the amount of variation of the original variables that they describe.

Eigenvalue analysis is the mathematical technique used in PCA. The eigenvalues and eigenvectors of the square symmetric matrix are solved with cross products and sums of squares. The eigenvector linked with the largest eigenvalue takes the same direction as the first principal component and the eigenvector connected with the second largest eigenvalue decides the direction of the second principal component.

Let Σ be the covariance matrix associated with the random vector $X' = [X_1, X_2, ...X_n]$. Let Σ have the eigenvalue-eigenvector pairs (λ_1, e_1), (λ_2, e_2),.... (λ_p, e_p) where $\lambda_1 \geq \lambda_2 \geq .... \geq \lambda_p \geq 0$. The ith principal component is given by,

$$Y_i = e_i^T X = e_{1i}X_1 + e_{2i}X_2 + ...+ e_{3i}X_3, \quad i=1, 2 ...p \qquad (13)$$

With these choices

$$\text{Variance}(Y_i) = e_i^T \Sigma e_i = \lambda_i \qquad i=1,2......p \qquad (14)$$

$$\text{Cov}(Y_i, Y_k) = e_i^T \Sigma e_k = 0 \qquad i \neq k \qquad (15)$$

Source: (Johnson & Wichern, 2003)

The Cox proportional hazard model is a widely used and applied method in survival analysis.

The proportional hazards assumption denotes the point that the hazard functions are multiplicatively connected. This fact can be assessed using log cumulative hazard curves.

Let covariates $X_1, X_2,…,X_k$ and the hazard of a patient with covariate values $x_1, x_2,…,x_k$ at time t_k be given by,

$$h_i(t) = h_0(t) \; [\exp(\beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + … + \beta_k X_{ik})] \tag{16}$$

$h_i(t)$ – hazard case for the ith rate at time t

$h_0(t)$ – baseline hazard at time t

The regression coefficients $\beta_0, \beta_1,…,\beta_k$ need to be estimated and these coefficients are independent of time. Therefore, the property of proportional hazards holds (Collett, 2003).

3. Results and Discussion

It was clear from the log rank test results that variables CD4 cell count, Karnofsky score and age should be analyzed further to decide on the surrogate variables. From the Kaplan-Maier plots drawn for Karnofsky score and CD4 it can be concluded that when Karnofsky score and CD4 is high, AIDS patients are doing better. Therefore, based on the descriptive statistics calculated it can be concluded that on its own CD4 cell count is the best to use as a surrogate, followed by Karnofsky score.

Sensitivity values of CD4 and Karnofsky score are 0.7 and 0.43 respectively. Specificity values of CD4 and Karnofsky score are 0.65 and 0.84 respectively. Attributable Proportion values of CD4 and Karnofsky score are 0.95 and 0.62 respectively.

In this study there are two treatment groups under consideration. "txgrp 1" is the treatment group with Zidovudine+ Lamivudine. "txgrp 2" is the treatment group with Zidovudine + Lamivudine+ Indinavir. Since txgrp 2 includes Indinavir it is considered to be the treated group and the txgrp 1 is considered to be the control group

.

Figure 1gives the Kaplan-Meier plot for variable "txgrp" (treatment group).
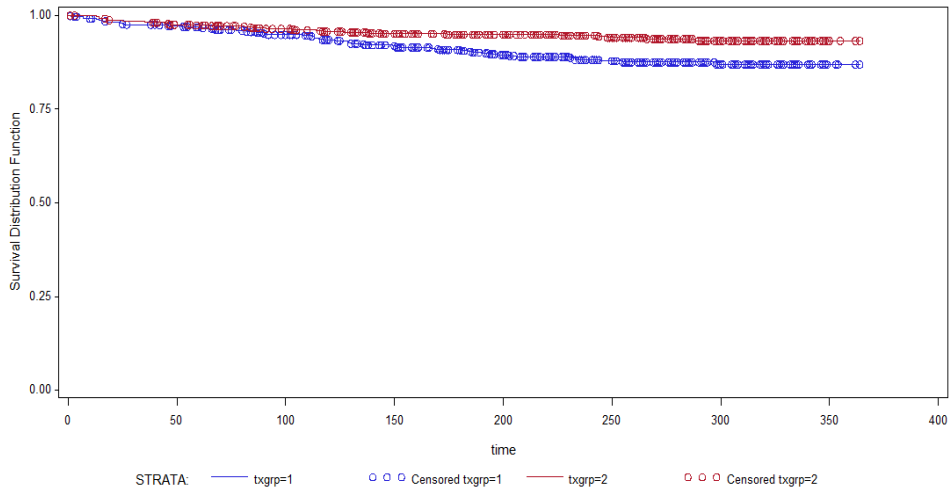
Figure 1 : Kaplan-Meier plot for txgrp

Figure 1 indicates a clear separation of two curves which means there is a clear distinction between two treatment groups with respect to survival. According to figure 1 when txgrp is equal to 1 it indicates a much lower survival. When txgrp is equal to 2 survival is higher, that is survival seems to be better in patients. Therefore, txgrp 2 (treated group) seems to be better. Here the time unit is indicated by number of days.

Earlier it was identified that variables CD4 and Karnofsky score can be taken as possible candidates for surrogate. However, in order to become a successful surrogate these variables should identify the differences between the two treatments. Therefore, tests were conducted to determine whether these two variables identify the difference between the two treatments individually using the PHREG procedure in SAS (a model using time dependent explanatory variables). The response variable here is the censoring indicator. The p-values based on Chi Square distribution corresponding to CD4 and txgrp are less than 0.0001 and 0.0022 respectively. Both variables (CD4 and txgrp) are highly significant. Therefore, it was concluded that CD4 successfully identifies the difference between the two treatments.

The p-values based on Chi Square distribution corresponding to Karnofsky score and txgrp are less than 0.0001 and 0.0010 respectively. Both variables (Karnofsky

score and txgrp) are highly significant. Therefore, it was concluded that Karnofsky score successfully highlights the difference between the two treatments.

Also it is important to note that individually all three variables CD4, Karnofsky score and txgrp are highly significant. An important past study of the same topic by O'Brien et al. (1996) used a combination of changes in Plasma HIV-1 RNA, CD4 and Lymphocyte Counts as a surrogate endpoint for an AIDS trial. Thus it is very important to compare our results with the results of O'Brien et al. (1996). Comparison shows that our p-values corresponding to CD4, Karnofsky score and treatment are more significant than the p values of the univariate analysis done by O'Brien et al. (1996) with the three variables HIV-1 RNA, CD4+ count, Lymphocyte Counts and treatment.

Since two variables were able to identify the difference between the two treatment groups they can be taken as potential candidates for surrogate. However it is important to note that individual tests were not carried out to the variable age(age at enrollment / treatment)because age alone cannot be taken as a surrogate endpoint since age does not change with the treatment where as variables CD4 cell count and Karnofsky score change with the treatment and individually they can be taken as surrogate endpoints. Althogh age increases with time, the length of the follow up study is only 375 days which is not very long. Therefore it was decided to consider the variable age when conducting the study.

In order to improve the procedure, it was decided to try out a combination of the three variables CD4, Karnofsky score and age (combined variable) as a surrogate rather than taking one variable alone as a surrogate. Therefore, initially, a logistic model was applied for the three variables CD4, Karnofsky score and age to check whether these are significant or not. According to the results all three variables are significant at the 5% level of significance. Probability values of CD4 and Karnofsky score are less than 0.0001 and the probability value of age is 0.0384. Therefore, it was decided to go with a combined variable model. To get the weights for the combined variable model Principal Component Analysis technique was applied. 1st Principal component explains 40% of the variation. 1st and 2nd Principal components explain 74% of the variation.

The coefficients of the 1st Principal component for Karnofsky score is 0.7362, for CD4 is0.5719 and for age is –0.362.Therefore the equation for the combined variable surrogate endpoint named 'score' using the 1^{st}principal component is as follows.

score=(0.572*cd4+0.736*karnofsky score-0.362*age)

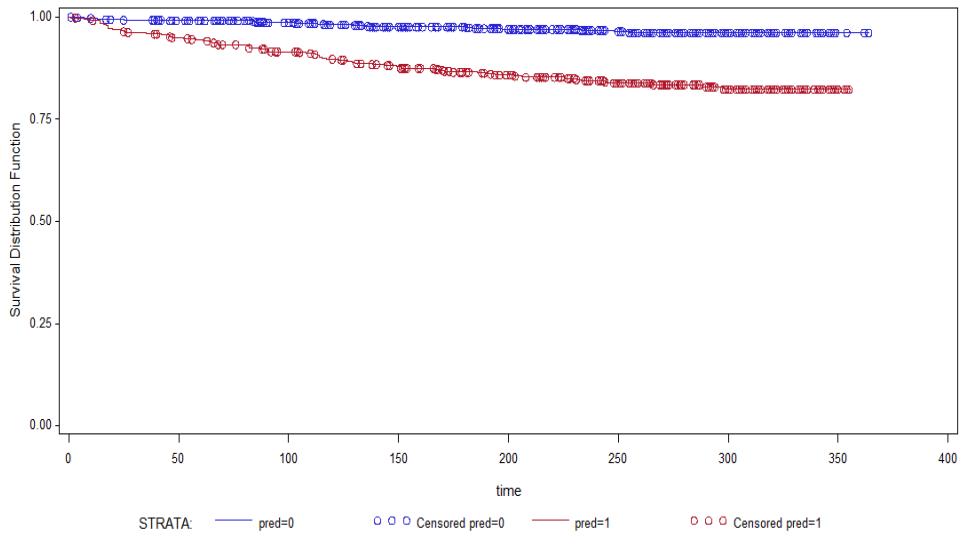A Kaplan-Maier plot was drawn for the combined variable model to analyze the survival behaviour.

Figure 2 : Kaplan-Meier plot for combined variable model

Based on the Kaplan-Maier plot indicated by Figure 2 it can be said that when score is higher it falls under the better category and when score is lower it falls under the worse category. Here the time unit is indicated by number of days.

Sensitivity for score was 0.78, specificity was 0.62 and the attributable proportion was 0.98 and when the descriptive statistical results of the score are compared with the results for the CD4 which was found out to be the best surrogate endpoint in the preliminary analysis, score produces better results in sensitivity and attributable proportion compared to CD4 which means that score is better than CD4.Also the attributable proportion value of score is closer to 1 than CD4 indicating that the relationship between the surrogate endpoint and the true endpoint is much stronger when it comes to score.

However, in order to classify the combined variable model as a successful surrogate endpoint, the model should identify the differences between the two treatments. Therefore a test was conducted to determine whether the combined variable model identifies the difference between the two treatments using the PHREG procedure in SAS (a model using time dependent explanatory variables). The response variable here is the censoring indicator. The p-values based on Chi Square distribution corresponding to score and txgrp are less than 0.0001 and

0.0021 respectively. Both score and txgrp are highly significant. Therefore, it was concluded that the combined variable model successfully identifies the difference between the two treatments. Also it is important to note that individually score is highly significant and it is more significant than the p values of the multivariate analysis done by O'Brien et al. (1996) with the three variables HIV-1 RNA, CD4+ count, Lymphocyte Counts and treatment.

However, the combined variable model with 1st two principal components which explains 74% of the variation cannot be taken as a good surrogate endpoint because with different cutoff values either the descriptive statistics do not give good results or the model does not highlight the difference between the two treatments. Therefore, it was decided to go ahead with the combined variable model or the score model based on the 1st principal component.

Then it was decided to test whether the combined variable model satisfies the four Prentice's criteria. When validating using the Prentice's criterion, the true endpoint (T) was considered to be the survival time. The logarithms of the two endpoints were considered when deriving the result. Therefore, the log of score and the log of time were fitted using a generalized linear model (GLM) on both variables assuming these are normally distributed. Here the censoring indicator corresponding to time is ignored (Molenberghs, Burzykowski& Buyse, 2005). A MANOVA test was done for the variables taking log of score and the log of time with respect to txgrp.

According to the results all four statistics (Wilk's Lambda, Pillai's Trace, Hotelling – Lawley Trace and Roy's Greatest Root) are significant and give the same probability value 0.0369 since the data set is balanced. However the main attention was paid to Hotelling –Lawley Trace statistic which is significant at the 5% level of significance which means that txgrp is significant. Since the treatment or the txgrp is significant in the above multivariate model it was concluded that the first two Prentice's criteria are satisfied for the newly developed surrogate endpoint score.

In order to satisfy the third criteria, it was decided to fit a model between surrogate endpoint (S) score and true endpoint (T) survival time and then to show that the surrogate endpoint, that is the score is significant in the model. However, since the survival time is not normally distributed, a Cox model or a parametric model should be fitted to model the relationship between the surrogate endpoint and the true endpoint.

Since the Cox model is used in a majority of biomedical studies, it was decided to go ahead with the Cox model and to use the PHREG procedure in SAS.

However, in order to check the Cox model's validity, the Cox-Snell residual plot was also plotted. The Cox-Snell residuals need to come from a unit exponential for the Cox model to be valid. That is the Log Negative Log Survival or the LLS

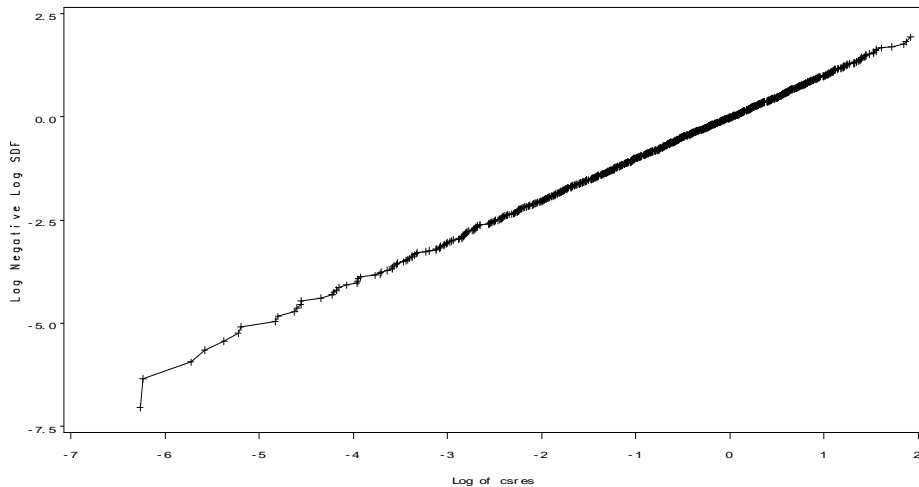plot of the Cox-Snell residuals have to be a straight line with unit slope and zero intercept.

Figure 3 : Cox-Snell residual plot for Prentice's criteria 3 in the score model

According to Figure 3 the plot is very linear and the LLS plot of the Cox-Snell residuals is a straight line with unit slope and zero intercept. This satisfies our model. Therefore, the Cox model is valid to model the relationship between surrogate endpoint score and the true endpoint and the proportional hazard assumption is well satisfied for the model.

The p-value associated with the score is 0.0079 indicating that the surrogate endpoint score is highly significant in the model. Therefore, it was concluded that Prentice's third criteria is satisfied by the score model.

In order to satisfy the fourth criterion, it was decided to fit a model between surrogate endpoint (S) and true endpoint (T) where treatment (Z) is also included and then to show that in the presence of surrogate endpoint score, treatment or the txgrp is no longer significant. To show this, a Cox model or a parametric model should be fitted to model the relationship between the surrogate endpoint and the true endpoint where treatment is also included. Since it was found out that with the parametric model, txgrp is also significant in the presence of the score and the Cox model was also used in the verification of Prentice's criteria 3 for the score, it was decided to go ahead with the Cox model. However, in order to check the Cox model's validity, the Cox-Snell residual plot was also plotted.
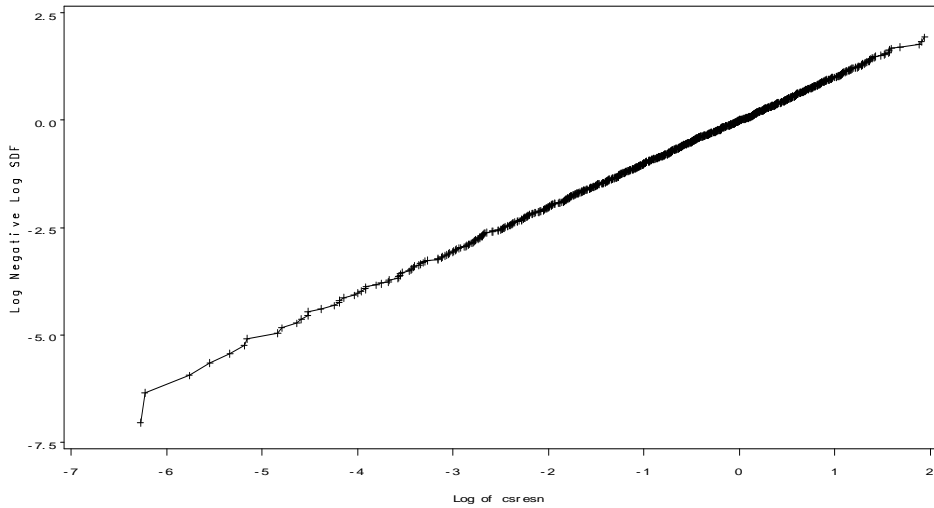
Figure 4 : Cox-Snell residual plot for Prentice's criteria 4 in the score model

According to Figure 4 the plot is very linear and the LLS plot of the Cox-Snell residuals is a straight line with unit slope and zero intercept. This is well satisfied for our model. Therefore, the Cox model is valid to model the surrogate endpoint score and the true endpoint where treatment is also included and the proportional hazard assumptions are well satisfied for the model.

The p-value for the score is 0.0081 and the p-value for txgrp is 0.2340 which means that score is highly significant and the txgrp is not significant. Therefore Prentice's fourth criterion is satisfied since txgrp is not significant in the presence of newly developed surrogate endpoint score. Since all four Prentice's criteria are satisfied, the use of the combined variable model can be justified as a new surrogate endpoint for AIDS.

In the same manner all four Prentice's criteria are also satisfied for the surrogate endpoint CD4 cell count. However, it is important to note that the newly developed surrogate endpoint score is better than CD4 with respect to all four criteria because score produces better results in all 4 criteria compared to CD4. The p-values obtained for score are more significant than the p-values obtained for CD4 cell count. In addition, when it comes to the 4th criterion treatment group is much less significant in the presence of score whereas in CD4 although treatment group is insignificant it is not as less significant in the score. Results are summarized in Table 2.

Table 2 : Comparison of p-values of score and CD4

Prentice's Criteria	score	CD4
1 and 2	0.0369	0.0438
3	0.0079	0.0335
4	0.0081	0.0334

Therefore, it is important to note that the newly developed surrogate endpoint score is better than CD4 alone with respect to all four criteria.

Although it is mentioned that HIV-1 RNA (viral load) as another surrogate endpoint for AIDS, a study by Lagakos and Hoth (1992) raise the concern about the limitations of viral load as a surrogate endpoint in AIDS (Lagakos &Hoth,1992). Therefore no direct comparison was done with the viral load and proposed score.

Although score produces better results in every test there is a limitation in the score model since it explains only 40% of variation in the data. It was unable to develop a good surrogate endpoint based on first two principal components, which explains 74% of the variation of the data. The major reason for this is the data set analyzed is not big enough. It would have been preferred to analyze a larger data set with HIV patients from many countries than this because this data set consists only 1151 HIV infected patients in the United States and Puerto Rico. Also, since there was only one observation for treatment group 3 and two observations for treatment group 4, those observations were removed from the study. It would have been preferred to have a data set where there are many patients getting treatment group 3 and 4, so that a better surrogate endpoint could have been developed reflecting the variations in treatment group 3 and 4 as well. If these points are adjusted in the data set a better surrogate endpoint could have been obtained and there might

 be a chance to get the first two principal components instead of only the 1st principal component, to develop the surrogate endpoint since the data set represents much variation.

Currently HIV-1 RNA and CD4 cell count is used in AIDS clinical trials as a surrogate endpoint. It would have been preferred to have the variable HIV-1 RNA in the data set that was analyzed so that a better surrogate endpoint could have been developed since currently HIV-1 RNA is regarded as a surrogate endpoint for AIDS.

4. Conclusion

Descriptive and univariate statistics suggested that score can be taken as a possible candidate for surrogate. All four Prentice's criteria were satisfied for both surrogate endpoints score and CD4. On the whole, the developed surrogate endpoint score was well validated using Prentice's criteria and gave accurate predictions about the two treatment groups being considered. Also our suggested surrogate endpoint score is better than that of previous work on a similar topic by O'Brien et. al. (1996). Therefore, score can be used as a surrogate endpoint for AIDS in future clinical trials. Apart from that score produces better outcomes in descriptive statistics compared to CD4 as indicated in Table 3.

Table 3: summary statistics for score & CD4

Statistic	score	CD4
Sensitivity	0.78	0.70
Specificity	0.62	0.65
Attributable Proportion	0.98	0.95

Also score is better than CD4 with respect to all four Prentice's criteria. Therefore, by considering all these facts, it can be concluded that the newly developed surrogate endpoint score is better than CD4 to be used in AIDS clinical trials.

Apart from this the developed surrogate endpoint score and txgrp are highly significant and they are more significant than the p-values of the multivariate analysis done by O'Brien et al. (1996) with the three variables HIV-1 RNA, CD4+ count, Lymphocyte Counts and treatment. Therefore, score successfully identifies the difference between the two treatments than the three variables HIV-1 RNA, CD4+ count, Lymphocyte Counts.

Acknowledgements

References

1. Agresti, A. (2007). *An Introduction to Categorical Data Analysis.* New Jersey: John Wiley and Sons, Inc. DOI:10.1002/0470114754

2. Burzykowski, T., Molenberghs, G. and Buyse, M. (2005).The Evaluation of Surrogate Endpoints.Springer.DOI:10.1007/b138566

3. Buyse, M. (2000). The validation of surrogate endpoints in meta-analyses of randomized experiments. *Biostatistics, 1*(1), 49–67.DOI:10.1093/biostatistics/1.1.49

4. Buyse, M., & Molenberghs, G. (1998, September). Criteria for the Validation of Surrogate Endpoints in Randomized Experiments. *Biometrics, 54*(3), 1014-1029.DOI:10.2307/2533853

5. Collett, D. (2003). Modelling Survival Data in Medical Research. New York: Chapman and Hall/CRC.

6. Ensor, H., Lee, R. J., Sudlow, C., & Weir, C. J. (2015, February). Statistical approaches for evaluating surrogate outcomes in clinical trials: a systematic review. *Journal of Biopharmaceutical Statistics*, 1-56

7. Fleming, T. R. (1994). Surrogate Markers in AIDS and Cancer Trials.*Statisticsin Medicine*,1423-1435.DOI:10.1002/sim.4780131318

8. Fleming, T. R., & Powers, J. H. (2012, March). Biomarkers and Surrogate Endpoints in Clinical Trials. *Statistics in Medicine, 31*, 2973–2984. DOI:10.1002/sim.5403

9. Gruttola, V. D., Fleming, T., Lin, D. Y and Coombs, R. (1997, February). Validating Surrogate Markers: Are We Being Naive? The Journal of Infectious Diseases, 175(2), 237-246. DOI:10.1093/infdis/175.2.237

10. Hosmer, D.W. and Lemeshow, S. (2000). Applied Logistic regression. United States of America:John Wiley and Sons, Inc. DOI:10.1002/0471722146

11. Johnson, R.A., and Wichern, D.W. (2003). *Applied multivariate Statistical analysis* (3rded.). New Delhi, India: Prentice-Hall of India.

12. Lalkhen, A.G and McCluskey, A.(2008). Clinical Tests : Sensitivity and Specificity.

13. O'Brien, W. A. (1996). Changes in Plasma HIV-1 RNA and CD4+ Lymphocyte Counts and the Risk of Progression to AIDS. *The New England Journal of Medicine,* 426-431. DOI:10.1056/NEJM199602153340703

14. Peto, T. (1996). Surrogate markers in HIV disease. *Journal of Antimicrobial Chemotherapy, 37*, 161-170. DOI:10.1093/jac/37.suppl_B.161

15. Prentice, R. L. (1989). Surrogate Endpoints in Clinical Trials : Definition and Operational Criteria. *Statistics in Medicine, 8*, 431-440.DOI:10.1002/sim.4780080407

16. Weir, C. J., & Walley, R. J. (2005, October). Statistical evaluation of biomarkers as surrogate endpoints:a literature review. *Statistics in Medicine, 25*, 183–203.DOI:10.1002/sim.2319