



---

UW Biostatistics Working Paper Series

---

1-13-2014

# Change Point Testing in Logistic Regression Models with Interaction Term

Youyi Fong

*Fred Hutchinson Cancer Research Center, yfong@fhcrc.org*

Chongzhi Di

*Fred Hutchinson Cancer Research Center, cdi@fredhutch.org*

Sallie Permar

*Duke University Medical Center, sallie.permar@duke.edu*

---

## Suggested Citation

Fong, Youyi; Di, Chongzhi; and Permar, Sallie, "Change Point Testing in Logistic Regression Models with Interaction Term" (January 2014). *UW Biostatistics Working Paper Series*. Working Paper 400.  
<http://biostats.bepress.com/uwbiostat/paper400>

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.

Copyright © 2011 by the authors

# Change Point Testing in Logistic Regression Models with Interaction Term

Youyi Fong<sup>†</sup>, Chongzhi Di<sup>‡</sup>, Sallie Permar<sup>§</sup>

January 10, 2014

## Abstract

The threshold effect takes place in situations where the relationship between an outcome variable and a predictor variable changes as the predictor value crosses a certain threshold/change point. Threshold effects are often plausible in a complex biological system, especially in defining immune responses that are protective against infections such as HIV-1, which motivates the current work. We study two hypothesis testing problems in change point models. We first compare three different approaches to obtaining a p-value for the maximum of scores test in a logistic regression model with change point variable as a main effect. Next, we study the testing problem in a logistic regression model with the change point variable both as a main effect and as part of an interaction term. We propose a test based on the maximum of likelihood ratio statistics and show that the correct significance level can be obtained by transforming random samples from a multivariate normal distribution. In simulation studies, we show the optimality of the maximum of likelihood statistics test among change point model-based methods, and demonstrate the performance trade-off when compared to dichotomizing the predictor variable at median across a range of true thresholds. We illustrate the utility of the change point model-based testing methods with a real data example from a recent study of immune responses that are associated with the risk of mother to child transmission (MTCT) of HIV-1.

**Keywords.** change point testing, effect modifier, maximum of score statistics, maximum of likelihood ratio statistics, mother to child transmission of HIV-1

## 1 Background

In this paper we study a change point model, also known as a threshold model, in which a covariate has no effect before reaching an unknown threshold and has a

---

\*Vaccine and Infectious Disease Division and Public Health Sciences Division, Fred Hutchinson Cancer Research Center

<sup>†</sup>Correspondence to: yfong@fhcrc.org, 1100 Fairview Ave N, Seattle, WA 98006

<sup>‡</sup>Public Health Sciences Division, Fred Hutchinson Cancer Research Center

<sup>§</sup>Human Vaccine Institute, Duke University Medical Center

constant effect after that. This type of change point model is a popular approach for handling nonlinearity in the relationship between two variables without over-parameterization, and have been widely used in econometrics, quality control, human genetics, and many more fields of study. Our interests in change point models arise from the study of immune responses to HIV-1 in subjects vaccinated with a HIV-1 vaccine [1] and HIV-1-infected mothers [2] and their association with HIV-1 infection risk. Several factors motivate us to consider change point models. First, it is often unclear how to properly transform a continuous immune response variable to be used in the regression. Second, we often create score variables which are combinations of individual immune response measurements. The relationship between the outcome variable and a score is more likely to be nonlinear than individual components of the score. Third, our current understanding of how immune systems operate is consistent with the existence of threshold effect, i.e. only an immune response above a certain quality and quantity threshold can result in protection from HIV-1 infection or transmission. Scientific considerations also prompt us to study change point models with interaction term because each immune response variable measures one aspect of the immune responses and different aspects of the multifaceted immune responses may need to work together synergistically to offer protection. Alternatively, one aspect of the immune responses may prevent other aspects of the immune responses from working effectively against the virus [1].

A particular challenge of the change point model testing problem is that under the null hypothesis, the threshold parameter becomes unidentifiable. This type of problems are often referred to as the Davies problems [3, 4, 5], and have motivated many previous works in the biostatistical literature on change point testing [3, 4, 6, 7, 8, 9, 10, 11, 12, 13, 14]. However, some methods, e.g. [4], don't apply to the model we study in this paper, which is discontinuous in the change point variable. Furthermore, some of the proposed methods, e.g. [11], do not have simulation studies to demonstrate the actual Type I error rates in finite samples. To provide a practical guidance on how to choose a method for performing a maximum of scores test in a logistic regression model with a change point variable as a main effect, in Section 2, we compare the finite sample performance of three different procedures for obtaining the p-value. In Section 3, we study a logistic regression model in which the change point variable appears both as a main effect and in an interaction term. We propose a test based on the maximum of likelihood ratio statistics, and use the same principle as in Section 2 to obtain the p-value associated with the test statistic. Of particular relevance to Section 3 is the work by [7], which was concerned with identifying a change point in a predictor that modifies the effect of treatment in a randomized two-arm clinical trial. Also of interest is [10], which proposed a maximum of likelihood ratio test in a different context which do not involve interaction terms. In Section 4, we conduct two simulation studies to study the Type I error rates and powers of the proposed tests. In Section 5, we use a real data example from a study of humoral immune responses that are associated with the risk of HIV-1 MTCT to illustrate the use of the proposed methods. We end with a discussion in Section 6.

## 2 Change Point Tests for Models with Main Effect Only

We first review the asymptotic theory for the score statistics in a logistic model with a change point variable as a main effect. Consider the following model

$$\text{logit} \{ \Pr(Y = 1) \} = \alpha^T z + \beta I(x > e), \quad (1)$$

where  $Y$  is a binary variable,  $z$  is a vector of covariates,  $x$  is the change point variable.  $\alpha$  is the vector of coefficients associated with  $z$ ,  $\beta$  is the effect size associated with the change point variable, and  $e$  is the threshold parameter. We are interested in testing the null hypothesis  $\beta = 0$ . The score test, which is based on the behavior of the test statistic under the null, is a natural choice here because the score statistics are asymptotically normally distributed. Let  $L_i$  denote the log likelihood for the  $i^{\text{th}}$  observation. The score with respect to  $\beta$  evaluated at  $\beta = 0$  is

$$\left. \frac{\partial L_i}{\partial \beta} \right|_{\beta=0} = I(x_i > e) \left\{ y_i - \frac{1}{1 + \exp(-\alpha^T z_i)} \right\} = I(x_i > e) (y_i - \mu_i),$$

where  $\mu_i = \text{expit}(\alpha^T z_i)$ . For a given  $e$ , let  $k = \# \{x_i > e\}$ . Denote  $w_i(e) = 1/k$  if  $x_i > e$  and 0 otherwise. Let  $w(e) = [w_1(e), \dots, w_n(e)]^T$ . Plug in the maximum likelihood estimate for  $\alpha$  under the null model and we have the score statistics, as a function of  $e$ ,

$$S_1(e) = w(e)^T (Y - \hat{\mu}) = \frac{1}{k} \sum_{i: x_i > e} (y_i - \hat{\mu}_i). \quad (2)$$

The score statistics is simply the sum of residuals for all observations with  $x$  greater than the threshold. This makes intuitive sense because when the data are from a model with  $\beta > 0$ , the residuals are likely to be greater than 0 for observations with  $x$  greater than the true threshold and less than 0 for observations with  $x$  less than the true threshold.

The score statistics depends on the threshold parameter which can not be estimated under the null because it is not part of the null model. If there is knowledge about the plausible value for  $e$  in a specific application, we will wish to use that knowledge. More often we have no idea where the threshold may be, a common strategy is to take the maximum of the score statistics evaluated at a sequence of  $M$  thresholds  $e_1, \dots, e_M$  for a fixed  $M$ . The asymptotic joint distribution of  $[S_1(e_1) \ \cdots \ S_1(e_M)]^T$  under the null model is given by the following theorem. The proof is straightforward using standard generalized linear model theory.

**Theorem 1** Denote  $W_1 = [w(e_1) \ \cdots \ w(e_M)]$ . Under the null,

$$\sqrt{n} [S_1(e_1) \ \cdots \ S_1(e_M)]^T = \sqrt{n} W_1^T (Y - \hat{\mu}) \Rightarrow N(0, V_1)$$

where  $V_1 = W_1^T A D A^T W_1$ ,  $D = \text{diag} \{ \mu(1 - \mu) \}$  and  $A = I - D Z (Z^T D Z)^{-1} Z^T$ .

There are many ways to conduct the test based on the above theory. We will examine four and compare their performance with a Monte Carlo study. Let  $\hat{D} = \text{diag} \{ \hat{\mu} (1 - \hat{\mu}) \}$  and  $\hat{V}_1(e) = w(e)^T \hat{A} \hat{D} \hat{A} w(e)$ . The first testing method is based on individually standardized score statistics,  $T(e) = \sqrt{n} |S_1(e)| / \sqrt{\hat{V}_1(e)}$ . Each of the marginal distributions of  $[T(e_1) \cdots T(e_M)]^T$  can be approximated by a normal distribution of mean 0 and variance 1. We take the test statistics to be

$$T_{\max} = \max \{ T(e_1), \dots, T(e_M) \}.$$

The distribution of  $T_{\max}$  can be approximated by the maximum of a multivariate normal distribution with mean 0, variance 1 and a correlation matrix derived from  $\hat{V}_1$ , and the p-value can be obtained by comparing  $T_{\max}$  with random samples from this multivariate normal distribution.

The first method requires the simulation of a multivariate normal distribution with an estimated correlation matrix. In the second method, we explore the use of more asymptotic theory to bypass the need to perform random sampling during testing. One potential approach is to apply formula (2.2) from [4] to derive an upper bound for the tail probability  $\Pr(T_{\max} > c)$ . This approach would not work however because it requires taking derivative of  $T(e)$  with respect to  $e$  and the upper bound explodes. Instead we apply results from [11], which provides an analytical approximation of the tail probability of the maximum of a normalized score statistics evaluated at a grid of thresholds. Let  $T_{\max}^A = \max \left\{ |S_1(e)| / \sqrt{1^T \hat{D} 1 k(n-k)/n} \right\}$ . By Theorem 3.1 and Remark 3.2 of [11], we have

$$\Pr \left( T_{\max}^A < \sqrt{2 \log \log n} + \frac{\log \log \log n}{2\sqrt{2 \log \log n}} + \frac{t - \frac{1}{2} \log \pi}{\sqrt{2 \log \log n}} \right) \rightarrow \exp \{ -2 \exp(-t) \}$$

as  $n \rightarrow \infty$ . Let  $t = \sqrt{2 \log \log n} T_{\max}^A - 2 \log \log n - \frac{1}{2} \log \log \log n + \frac{1}{2} \log \pi$ . The p-value associated with an observed  $T_{\max}^A$  is then  $1 - \exp \{ -2 \exp(-t) \}$ .

The third method we explore also does not require Monte Carlo sampling. Let  $U = \hat{V}_1^{-1/2} \sqrt{n} [S_1(e_1) \cdots S_1(e_M)]^T$ . The distribution of  $U$  can be approximated by a multivariate standard normal distribution, and we obtain a p value by comparing  $U^T U$  with a chi-squared distribution of degree  $M$ .

### 3 Maximum of Likelihood Ratio Test for Change Point Models with Interaction Term

We now consider a logistic regression model with the change point variable appearing both as a main effect and as part of an interaction term

$$\text{logit} \{ \Pr(Y = 1) \} = \alpha^T z + \beta_1 I(x > e) + \beta_2 z_1 I(x > e), \quad (3)$$

where  $z_1$  is a component of the covariate vector  $z$ ,  $\beta_1$  is the effect size associated with the main effect of the change point variable, and  $\beta_2$  is the effect size

associated with the interaction term involving the change point variable. We are interested in testing  $\beta_1 = \beta_2 = 0$ . Although the null model is the same as the null model in testing  $\beta = 0$  in model (1), the test here can potentially be a more powerful test against the null because sometimes we may fail to detect a weak main effect, but succeed in detecting a weak main effect and a weak interaction effect together.

The score vector with respect to  $\beta_1$  and  $\beta_2$  evaluated at  $\beta_1 = \beta_2 = 0$  are

$$\left[ \begin{array}{c} \partial L_i / \partial \beta_1 \\ \partial L_i / \partial \beta_2 \end{array} \right] \Big|_{\beta_1 = \beta_2 = 0} = \left[ \begin{array}{c} 1 \\ z_{i,1} \end{array} \right] I(x_i > e) (y_i - \mu_i).$$

Plug in the estimate for  $\mu_i$  in the null model and we have a vector of score statistics  $[ S_1(e) \ S_2(e) ]$ , where  $S_1(e)$  is as defined in (2) and  $S_2(e)$  is defined as follows:

$$S_2(e) = (w * z_1)^T (Y - \hat{\mu}) = \frac{1}{k} \sum_{i: x_i > e} z_{1,i} (y_i - \hat{\mu}_i), \quad (4)$$

where  $w * z_1$  is elementwise multiplication of two vectors  $w$  and  $z_1$ . For a sequence of  $M$  potential thresholds, we can form a score statistics vector of length  $2M$ . Its asymptotic joint distribution under the composite null  $\beta_1 = \beta_2 = 0$  is given by the following theorem. The proof is similar to the proof of theorem 1 and follows from generalized linear model theory.

**Theorem 2** Denote  $W_3 = [ w(e_1) \ w(e_1) * z_1 \ \dots \ w(e_M) \ w(e_M) * z_1 ]$ . Under the null  $\beta_1 = \beta_2 = 0$ ,

$$\sqrt{n} [ S_1(e_1) \ S_2(e_1) \ \dots \ S_1(e_M) \ S_2(e_M) ]^T = \sqrt{n} W_3^T (Y - \hat{\mu}) \Rightarrow N(0, V_3),$$

where  $V_3 = W_3^T A D A^T W_3$ ,  $D = \text{diag} \{ \mu(1 - \mu) \}$  and  $A = I - DZ (Z^T DZ)^{-1} Z^T$ .

One approach to testing, based on the lessons learned in testing the main effect, is to let the test statistics be the maximum element of the standardized multivariate score statistics, namely

$$T_{\max}^{MI} = \max \left\{ [ S_1(e_1) \ S_2(e_1) \ \dots \ S_1(e_M) \ S_2(e_M) ] / \sqrt{\text{diag}(\hat{V}_3)} \right\},$$

where  $\text{diag}(\hat{V}_3)$  is the diagonal of the estimated covariance matrix  $\hat{V}_3$  and the division is elementwise division. The p-value can be obtained by comparing  $T_{\max}^{MI}$  with random samples of maximum of a multivariate normal distribution with mean 0, variance 1 and a correlation matrix derived from  $\hat{V}_3$ .

There are two related issues with tests based on  $T_{\max}^{MI}$ . The first issue is that the test p value for the same dataset changes with affine transformation of the covariate  $z_1$ . The second issue is that tests based on  $T_{\max}^{MI}$  are not as powerful as it can be. These two issues are connected by the fact that if we let  $z_1^*$  denote the

standardized  $z_1$  and  $\bar{S}_2(e)$  denote the corresponding score statistics.  $S_2(e)$  can be expressed as a linear combination of  $S_1(e)$  and  $\bar{S}_2(e)$  depending on the affine transformation of  $z_1$ . Different linear combinations have different powers. The best combination is a function of  $\beta_1$  and  $\beta_2$  when the data is generated from the model (3). Without knowing the true  $\beta_1$  and  $\beta_2$ , we could try maximizing over a grid of thresholds and linear combinations. This is computationally intensive, however. We can achieve a similar level of performance by realizing that under a fixed  $e^*$ , model (3) becomes

$$\text{logit}\{\Pr(Y = 1)\} = \alpha^T z + \beta_1 I(x > e^*) + \beta_2 z_1 I(x > e^*), \quad (5)$$

and we are interested in testing  $\beta_1 = \beta_2 = 0$ . It is well known that for regular models, the likelihood ratio test, Wald test and score test are asymptotically equivalent under the null; yet likelihood ratio test often has better power. Hence we propose a maximum of likelihood ratio test. For each  $e^*$ , we fit model (5) and obtain a likelihood ratio statistics  $Q(e^*)$  against the null model  $\text{logit}\{\Pr(Y = 1)\} = \alpha^T z$ . Our test statistics is

$$Q_{\max} = \max\{Q(e_1), \dots, Q(e_M)\}.$$

To obtain the reference distribution for  $Q_{\max}$ , we use the fact that under the null, each  $Q(e)$  can be asymptotically expressed as

$$Q(e) = [S_1(e) \ S_2(e)] \hat{I}_{\beta\beta,\alpha}^{-1}(e) \begin{bmatrix} S_1(e) \\ S_2(e) \end{bmatrix} + o_p(1),$$

where  $\hat{I}_{\beta\beta,\alpha}(e)$  is a plug-in estimate of  $I_{\beta\beta,\alpha}(e)$ ,  $I_{\beta\beta,\alpha}(e) = I_{\beta\beta}(e) - I_{\beta\alpha}(e) I_{\alpha\alpha}^{-1}(e) I_{\alpha\beta}(e)$  and  $\begin{bmatrix} I_{\alpha\alpha}(e) & I_{\beta\alpha}(e) \\ I_{\alpha\beta}(e) & I_{\beta\beta}(e) \end{bmatrix}^T$  is a partition of the Fisher information matrix for model (5) with  $\beta = [\beta_1 \ \beta_2]$  here. Let  $J$  denote a  $2M \times 2M$  block diagonal matrix with  $\hat{I}_{\beta\beta,\alpha}^{-1/2}(e)$  on the diagonal. Denote by  $V_4$  the variance-covariance matrix of  $\begin{bmatrix} [S_1(e_1) \ S_2(e_1)] \hat{I}_{\beta\beta,\alpha}^{-1/2}(e_1) & \dots & [S_1(e_M) \ S_2(e_M)] \hat{I}_{\beta\beta,\alpha}^{-1/2}(e_M) \end{bmatrix}$ .  $V_4$  can be approximated by  $JV_3J$  asymptotically. Because the diagonal of  $V_4$  are 1's, in practice, better performance can be obtained by approximating  $V_4$  with the correlation matrix derived from  $JV_3J$ . Hence, to obtain random samples that approximate the reference distribution of  $Q_{\max}$ , we simulate from a multivariate normal distribution with mean 0, variance 1, and correlation matrix corresponding to  $JV_3J$ . Each random sample is a vector of  $2M$  elements and can be viewed as a sequence of  $M$  pairs of random variables. Taking the sum of squares for each pair, the distribution of the maximum of the  $M$  sums of squares approximates the distribution of  $Q_{\max}$ .

## 4 Simulation Studies

### 4.1 Main effect only

We simulate data from

$$\text{logit}\{\Pr(Y = 1)\} = \alpha + \log(1.4) Z$$

where  $[Z, X]^T$  comes from a bivariate normal distribution with mean 0, variance 1 and correlation  $\rho$ . We test two values for  $\rho$ : 0 and 0.3.  $\alpha$  is chosen so that the proportion of cases in the dataset is 1/3. The sample size is 250. We test whether there is a threshold effect in  $x$  with a nominal Type 1 error rate of 5%. The Type 1 error rate estimates from 10,000 replicates are shown in Table 1. The standard deviation of the estimate at the nominal rate is 0.22%.

Results in Table 1 show that approximating the distribution of  $T_{\max}$  using the estimated covariance matrix gives Type I error rates that are close to the nominal level. Approximating the distribution of  $T_{\max}^A$  with the asymptotic expression from [11] leads to a conservative test. Testing with  $U^T U$  is also conservative, presumably due to the difficulty in estimating the full variance-covariance matrix. Additional simulation results (not shown) show that the higher  $M$ , the number of thresholds examined, is, the more conservative the test becomes. From now on, our method of choice for change point model-based testing will be  $T_{\max}$ .

We compare the maximum of score test based on the change point model with two other tests not based on change point models. In one, we encode the covariate  $x$  as a binary variable dichotomized at median, and use a Wald test to test the coefficient associated with the variable. In the other, we encode the covariate  $x$  as a trichotomous variable with cut points chosen as the 33% and 67% percentiles, and use a generalized Wald test to test the overall hypothesis that there is no association between the outcome and  $x$ . The estimated Type I error rates for the two tests from 10,000 replicates are 5.11 and 4.89, respectively.

To examine the powers of these tests, we simulate data from

$$\text{logit}\{\Pr(Y = 1)\} = \alpha + \log(1.4)Z + \beta I(X > e)$$

where  $Z \sim N(0, 1)$ ,  $X \sim N(4.7, sd = 1.6)$  and the correlation  $\rho$  between  $Z$  and  $X$  is either 0 or 0.3. Let  $e$  take values from 3.4 to 6.  $\beta$  is chosen from three levels of odds ratios:  $\log(0.8)$ ,  $\log(0.6)$  and  $\log(0.4)$ .  $\alpha$  is chosen so that the proportion of cases in each dataset is 1/3 on average. The estimated powers from 2,000 replicates for all methods are shown in Table 2 and Figure 1. The results suggest that the performance of the tests based on binary encoding depends greatly on how close the true threshold is to the median of the covariate distribution. For example, when the true odds ratio is 0.4 and  $\rho = 0$ , the powers of the tests based on binary encoding range from 85% to 24%. On the other hand, under the same setting, the powers of the tests based on change point models vary much less, ranging from 48% to 80%. In other words, at a price of slightly less power when the true threshold is close to the median, the change point method guarantees a relatively high level of power when the true threshold is away from the median of the covariate. Thus, the maximum of score test based on the change point model is an omnibus procedure that is powerful against alternatives under a wide range of threshold values. Trichotomizing the covariate  $x$  is more powerful than dichotomization when the true threshold is away from the median, but



almost always inferior to the change point method. Results from other  $\beta$ 's and  $\rho = 0.3$  are similar.

## 4.2 Main effect plus interaction

We now study the performance of the testing procedures for interaction model. We simulate data from

$$\text{logit}\{\Pr(Y = 1)\} = \alpha + \log(1.4)Z + \beta_1 I(X > e) + \beta_2 \times Z \times I(X > e),$$

where the covariates are simulated as previously described.  $\alpha$  is chosen so that the proportion of cases in the dataset is  $1/3$ . The sample size is 250. We test the null hypothesis that  $\beta_1 = \beta_2 = 0$  at 5% alpha level. Based on 10,000 replicates, the estimated Type 1 error rates for  $T_{\max}^{MI}$ -based test are 5.1% and 5.0% for  $\rho = 0$  and 0.3, respectively, the estimated Type 1 error rates for  $T_{\max}$ -based test are 5.0% and 4.9% for  $\rho = 0$  and 0.3, respectively, and the estimated Type 1 error rates for  $Q_{\max}$ -based test are slightly elevated at 5.9% for both  $\rho = 0$  and 0.3, respectively. We compare the change point-based testing methods with setting threshold at median and use a likelihood ratio test to test the null hypothesis. The estimated Type I error rate are 5.0% and 4.6% for  $\rho = 0$  and 0.3, respectively, based on 10,000 replicates.

To examine the powers of the test, we simulate data from two  $\beta_1$ :  $\{\log(1.5), -\log(1.5)\}$  and three  $\beta_2$ :  $\{\log(0.8), \log(0.6), \log(0.4)\}$ . Results from 2,000 replicates are shown in Table 3 and Figure 1. First, let us focus on  $\rho = 0$  and  $\beta_1 = \log(0.67)$ . When the odds ratio for the interaction term,  $\exp(\beta_2)$ , is 0.8, the change point testing method based on  $T_{\max}$  is slightly more powerful than the change point testing method based on  $T_{\max}^{MI}$  and  $Q_{\max}$ . As the odds ratio for the interaction term drops to 0.6 and 0.4, the powers of tests based on  $T_{\max}$  remain relatively constant, while the powers of tests based on  $T_{\max}^{MI}$  and  $Q_{\max}$  become significantly more powerful. This suggests that jointly testing the main effect and interaction of terms can be more powerful than testing the main effect marginally, in the presence of a moderate to strong interaction effect. Tests based on  $Q_{\max}$  are consistently more powerful than tests based on  $T_{\max}^{MI}$ , making  $Q_{\max}$  our most preferred choice among these tests. The difference in performance between the tests based on binary encoding and  $T_{\max}^{MI}$  is similar to the difference in the previous subsection, namely that the change point tests trade some power loss at the median threshold for greater power gain at the more extreme thresholds. Results from  $\beta_1 = -\log(0.67)$  and  $\rho = 0.3$  are similar.

## 5 Data Examples

In this section, we illustrate the use of change point model-based testing methods using the Mother To Child HIV-1 Transmission (MTCT) humoral immune

correlates study [2]. Every year more than 300,000 infants acquire HIV-1 despite the fact that there now exist effective antiretroviral prophylaxis regimens. Development of maternal or infant HIV-1 vaccine will undoubtedly hasten the elimination of pediatric HIV-1. To further our understanding of effective human immune responses that will prevent HIV-1 infection, the MTCT immune correlates study was carried out to identify maternal HIV-1 specific immunologic biomarkers that are associated with the risk of HIV-1 MTCT. The study used samples from the historical Women and Infants Transmission Study cohort of U.S. HIV-infected mother-infant pairs enrolled prior to the availability of antiretroviral drugs in an observational study of vertical HIV-1 transmission and pathogenesis. Eighty-three HIV-transmitting mothers and 165 nontransmitting mothers with available plasma samples were selected for the study. None of the mothers breast-fed or received any antiretroviral prophylaxis.

One immune response variable of particular interest is V3\_score. This variable is a linear combination of several variables measuring the strength of IgG antibody binding to the variable loop (V3) region [15] of several variants of HIV-1 Envelope proteins. We examined the association between HIV-1 transmission and V3\_score via three models. All models include several clinical factors known to be associated with the risk of vertical transmission including viral load, gestational age etc., and they differ in how V3\_score is encoded. In the first model, V3\_score is treated as a continuous variable; in the second, it is treated as a binary variable dichotomized at median; and in the third, it is modeled as a change point variable. The p-values from Wald tests in the first two models are 0.04 and 0.35, respectively. The discrepancy of the two results can be explained by the result from the third model, where the maximum of score test p value is 0.04 and the threshold that yields the maximum score statistics is at 10% of the V3\_score distribution, far from the median.

Different immune biomarkers measure distinct aspects of the immune response to HIV-1. To have an effective defense against HIV-1, these different aspects of the immune response may have to work together synergistically. This suggests that it is important to study the interaction between immune biomarkers. For illustration, we focus on the interactions between a variable NAb\_score, which measures the amount and breadth of neutralizing antibodies [16], and eight other continuous immune response variables selected for their scientific importance [2]. NAb\_score by itself does not have a significant association with transmission risk whether it is studied as a continuous variable, median-dichotomized variable or change point variable. When NAb\_score is studied as a continuous variable or as a median-dichotomized variable, only one immune response variable, IgG\_Mngp41, is found to interact with it, which measures the IgG antibody binding to the gp41 protein of the HIV-1 Envelope. When NAb\_score is studied as a change point variable using the maximum of likelihood ratio statistics, three more variables in addition to IgG\_Mngp41 are found to interact with it. These variables measure antibody avidity, IgG antibody binding to the gp120 protein on the HIV-1 Envelope and IgA antibody binding to the gp41 protein on HIV-1 Envelope, respectively, and they do not show significant association with transmission risk on their own. The

broad pattern of interaction uncovered by the change point method accentuates the importance of eliciting multiple immune responses for any successful HIV-1 vaccine.

## 6 Discussion

This paper is motivated by the need to detect threshold effect in the study of synergistic human immune responses to HIV-1 virus. First, we compare three methods for computing p values for maximum of score tests in a logistic regression model with a main effect modulated by a change point. Second, we study testing methods for models containing a change point variable both as a main effect and as part of an interaction term. We propose a test based on the maximum of likelihood ratio statistics and a method to approximate its asymptotic distribution. Through simulation study, we show that the proposed methods have proper Type I error rates. The powers of the proposed maximum of score tests and maximum of likelihood ratio tests are much less sensitive to the value of the true threshold than simply dichotomizing a variable at its median. We apply the proposed methods to a study of the impact of immune responses in HIV-1 carrying mothers on the risks of transmitting HIV-1 to infants. Tests based on maximum of likelihood ratio statistics reveal multiple significant interactions, suggesting the importance of eliciting multiple immune responses when designing a HIV-1 vaccine. The proposed testing methods have been implemented in a R package *chnqpt*, and can be downloaded from the Comprehensive R Archive Network (CRAN).

We often apply a change point model in a two stage process. First, we perform hypothesis testing to see if there is a threshold effect; second, we fit a change point model to obtain estimates of the threshold and the regression coefficients. In such a context, the parameter estimates from the second stage may be biased towards greater absolute values and do not follow the established asymptotic theory for the sampling distribution of the parameter estimates under change point models [17]. Establishing valid confidence intervals for parameter estimates in a change point model is currently under investigation.

## Acknowledgements

The authors thank Peter Gilbert for helpful comments on the manuscript. This work was supported by the U.S. Military HIV Research Program grant W81XWH-07-2-0067, the National Institute of Allergy and Infectious Diseases (NIAID) grant UM1-AI-068618 to the HIV Vaccine Trials Network, the NIAID grant UM1 AI100645-02 to the Duke Center for HIV/AIDS Vaccine Immunology and Immunogen Discovery, the NIAID grant AI104370-01 to Y.F., and the NIEHS grant ES-022332 to C.D.

## References

- [1] Haynes BF, Gilbert PB, McElrath MJ, Zolla-Pazner S, Tomaras GD, Alam SM, Evans DT, Montefiori DC, Karnasuta C, Sutthent R, *et al.*. Immune-correlates analysis of an hiv-1 vaccine efficacy trial. *New England Journal of Medicine* 2012; **366**(14):1275–1286. NIHMSID:NIHMS371846.
- [2] Permar SR, Fong Y, Fouda GG, Vandergrift N, Gilbert P, Datta S, Parks R, Lloyd K, Liebl B, Overman G, *et al.*. Maternal hiv-1 envelope variable region 3-specific igg responses are a correlate of risk of perinatal transmission. *New England Journal of Medicine* 2013; **submitted**.
- [3] Davies R. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 1977; **64**(2):247–254.
- [4] Davies R. Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* 1987; **74**(1):33–43.
- [5] Davies R. Hypothesis testing when a nuisance parameter is present only under the alternative: Linear model case. *Biometrika* 2002; **89**(2):484–489.
- [6] Ulm K. A statistical method for assessing a threshold in epidemiological studies. *Statistics in medicine* 1991; **10**(3):341–349.
- [7] Koziol JA, Wu SCH. Change-point statistics for assessing a treatment-covariate interaction. *Biometrics* 1996; :1147–1152.
- [8] Xu R, Adak S. Survival analysis with time-varying regression effects using a tree-based approach. *Biometrics* 2002; **58**(2):305–315.
- [9] Mazumdar M, Smith A, Bacik J. Methods for categorizing a prognostic variable in a multivariable setting. *Statistics in medicine* 2003; **22**(4):559–571.
- [10] Pastor-Barriuso R, Guallar E, Coresh J. Transition models for change-point estimation in logistic regression. *Statistics in medicine* 2003; **22**(7):1141–1162.
- [11] Antoch J, Gregoire G, Jarušková D. Detection of structural changes in generalized linear models. *Statistics & probability letters* 2004; **69**(3):315–332.
- [12] Zheng G, Chen Z. Comparison of maximum statistics for hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrics* 2005; **61**(1):254–258.
- [13] Vexler A, Gurevich G. Average most powerful tests for a segmented regression. *Communications in Statistics - Theory and Methods* 2009; **38**(13):2214–2231.

- [14] Lee S, Seo M, Shin Y. Testing for threshold effects in regression models. *Journal of the American Statistical Association* 2011; **106**(493):220–231.
- [15] AYYAVOO V, UGEN KE, FERNANDES LS, GOEDERT JJ, RUBINSTEIN A, WILLIAMS WV, WEINER DB. Analysis of genetic heterogeneity, antigenicity, and biological characteristics of hiv-1 in a maternal transmitter and nontransmitter patient pair. *DNA and cell biology* 1996; **15**(7):571–580.
- [16] Li M, Gao F, Mascola JR, Stamatatos L, Polonis VR, Koutsoukos M, Voss G, Goepfert P, Gilbert P, Greene KM, *et al.*. Human immunodeficiency virus type 1 env clones from acute and early subtype b infections for standardized assessments of vaccine-elicited neutralizing antibodies. *Journal of virology* 2005; **79**(16):10 108–10 125.
- [17] Hansen BE. Sample splitting and threshold estimation. *Econometrica* 2000; **68**(3):575–603.



$\rho$	$T_{\max}$	Antoch	$U^T U$
0	5.2	0.7	1.6
0.3	5.1	2.0	1.3

Table 1: Type I error rates (percent) of hypothesis testing procedures for logistic regression models with change point variable as a main effect.  $\rho$ : correlation of predictor  $Z$  and  $X$ .



$e$ (quantile)	$\rho = 0$					$\rho = 0.3$				
	0.21	0.36	0.46	0.58	0.79	0.21	0.43	0.53	0.65	0.82
$\beta = \log(0.8)$										
trichotomized	7.0	10.0	9.3	9.4	7.5	7.1	9.2	8.2	7.4	6.3
dichotomized	6.3	9.2	11.6	10.6	6.7	6.1	8.5	11.6	10.1	6.2
$T_{\max}$ -based	8.1	10.0	10.7	10.0	7.4	8.2	9.0	8.8	8.8	7.0
$\beta = \log(0.6)$										
trichotomized	15.8	31.4	27.1	25.1	14.4	14.8	28.9	24.3	23.2	13.6
dichotomized	11.9	27.2	40.9	34.3	11.8	11.9	24.4	37.8	31.6	10.8
$T_{\max}$ -based	23.4	31.1	32.2	30.6	18.6	21.9	28.2	29.8	27.6	17.5
$\beta = \log(0.4)$										
trichotomized	41.2	77.6	69.7	68.0	33.1	37.4	75.3	66.5	65.6	32.2
dichotomized	28.4	65.8	85.0	77.9	24.3	24.6	61.1	82.8	76.5	22.9
$T_{\max}$ -based	62.2	76.3	80.2	76.8	48.2	60.0	74.7	77.5	73.3	47.6

Table 2: Powers of hypothesis testing procedures for logistic regression models with change point variable as a main effect.  $\rho$ : correlation of predictor  $Z$  and  $X$ .

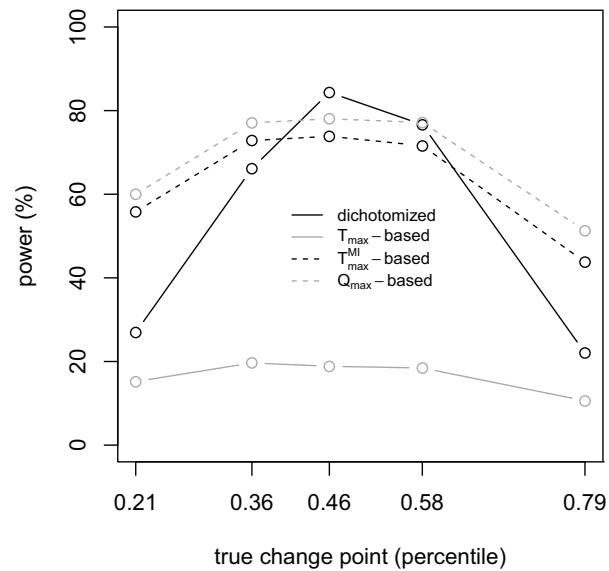
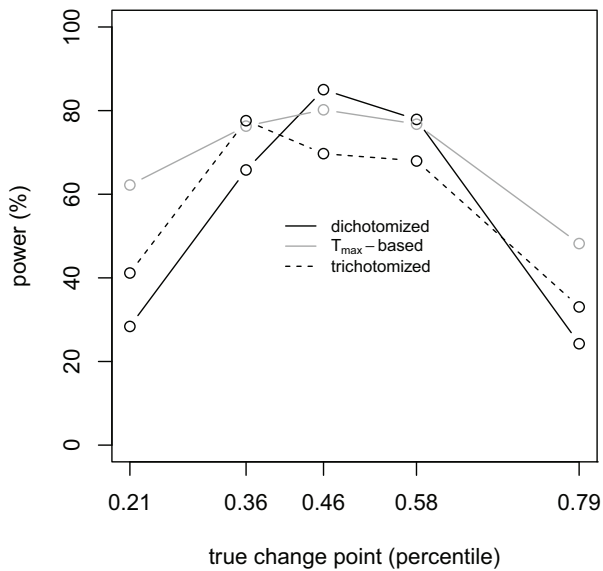


Figure 1: Powers of hypothesis testing procedures for logistic regression models with change point variable as a main effect only ( $\beta = \log(0.4)$ , left) and both as a main effect and as part of an interaction term ( $\beta_1 = \log(0.67), \beta_2 = \log(0.4)$ , right). Correlation of predictor  $Z$  and  $X$ :  $\rho = 0$ .



$e$ (quantile)	$\rho = 0$					$\rho = 0.3$				
	0.21	0.36	0.46	0.58	0.79	0.21	0.43	0.53	0.65	0.82
$\beta_1 = \log(0.67), \beta_2 = \log(0.8)$										
dichotomized	8.9	18.4	26.6	23.9	9.4	7.6	16.4	23.6	22.9	9.4
$T_{\max}$ -based	16.3	21.8	22.1	21.3	12.8	12.8	18.1	20.9	21.2	16.7
$T_{\max}^{MI}$ -based	14.3	18.6	19.1	19.3	11.0	11.2	15.8	18.1	18.1	14.6
$Q_{\max}$ -based	14.9	19.1	21.0	20.9	13.6	12.6	17.2	18.7	19.5	16.3
$\beta_1 = \log(0.67), \beta_2 = \log(0.6)$										
dichotomized	13.3	31.6	47.6	41.2	12.8	10.5	27.2	42.9	43.4	16.2
$T_{\max}$ -based	15.8	21.1	22.0	21.8	13.6	9.2	15.4	19.7	24.5	22.2
$T_{\max}^{MI}$ -based	23.2	32.6	32.4	31.5	18.9	14.0	23.6	28.1	32.2	27.5
$Q_{\max}$ -based	26.9	38.1	39.1	37.9	22.8	19.8	30.7	35.6	39.4	30.2
$\beta_1 = \log(0.67), \beta_2 = \log(0.4)$										
dichotomized	27.0	66.1	84.4	76.6	22.1	21.8	59.7	80.2	77.8	25.7
$T_{\max}$ -based	15.2	19.4	18.8	18.5	10.9	7.8	13.5	17.1	24.4	27.0
$T_{\max}^{MI}$ -based	56.5	73.2	73.9	71.7	44.5	43.1	62.4	66.6	71.2	55.3
$Q_{\max}$ -based	60.0	77.1	78.0	77.2	51.3	49.6	69.2	73.9	77.7	60.1
$\beta_1 = -\log(0.67), \beta_2 = \log(0.8)$										
dichotomized	8.9	16.8	26.3	23.8	9.2	10.1	17.6	26.2	22.0	7.8
$T_{\max}$ -based	13.8	19.2	20.2	20.3	14.7	16.0	20.5	21.0	19.1	11.9
$T_{\max}^{MI}$ -based	11.4	15.6	18.0	18.4	13.5	14.8	17.7	18.1	16.8	10.0
$Q_{\max}$ -based	14.4	19.5	21.3	20.5	13.8	17.8	20.8	21.1	19.7	12.3
$\beta_1 = -\log(0.67), \beta_2 = \log(0.6)$										
dichotomized	13.7	33.7	52.4	46.2	14.5	17.7	36.5	50.0	42.2	11.8
$T_{\max}$ -based	13.8	20.4	21.2	20.9	15.5	24.8	25.0	22.8	17.9	8.3
$T_{\max}^{MI}$ -based	21.4	32.5	35.7	33.6	23.7	33.4	38.5	36.2	29.5	13.8
$Q_{\max}$ -based	27.2	41.3	44.4	41.6	26.9	35.8	43.0	43.2	38.6	20.7
$\beta_1 = -\log(0.67), \beta_2 = \log(0.4)$										
dichotomized	28.6	67.5	87.4	82.3	27.3	36.1	70.5	86.2	77.5	22.9
$T_{\max}$ -based	14.5	18.7	19.8	20.5	16.4	33.1	29.3	23.9	16.4	6.3
$T_{\max}^{MI}$ -based	56.2	74.5	76.7	76.0	55.2	71.5	77.3	77.4	71.0	37.6
$Q_{\max}$ -based	60.1	79.0	81.3	81.5	60.2	69.0	80.0	81.0	77.5	47.6

Table 3: Powers of hypothesis testing procedures for logistic regression models with change point variable both as a main effect and as part of an interaction term.  $\rho$ : correlation of predictor  $Z$  and  $X$ .