◆◆ Scientific
◆◆ Research

Acute-to-Chronic Ratio as a Metric for HIV Surveillance

Zonghui Hu, Dean A. Follmann

Biostatistics Research Branch, National Institute of Allergy and Infectious Diseases,
National Institutes of Health, Bethesda, USA
Email: huzo@niaid.nih.gov, dfollmann@niaid.nih.gov

ABSTRACT

Early diagnosis is important for HIV control on both the individual and the population level. Patients who are diagnosed shortly after infection have more treatment options which can result in delayed progression to AIDS. Early diagnosis can also help to reduce the spread of HIV: with viral loads at the highest level, the newly infected have the highest risk of transmitting HIV to partners. Accordingly it is important to evaluate at a population level whether individuals are diagnosed shortly after infection. For this purpose, we introduce the acute-to-chronic ratio, which is the odds of the infected individuals being diagnosed in the acute infection stage versus the chronic infection stage among those who seek HIV testing. We demonstrate this new metric through HIV testing results from sexually transmitted disease clinics of four cities, and compare the four populations with respect to how quickly the infected get HIV testing.

KEYWORDS

1. Introduction

We introduce a new metric in this article for evaluating the rapidity with which the infected seek HIV testing. It is a potentially useful measure to compare the effectiveness of HIV surveillance across different populations as well as monitoring a population over the time.

People go through different stages upon HIV infection. Immediately after infection, no antibodies are produced by the human immune system and the HIV viruses replicate very quickly. Characteristic of this initial stage is a high HIV viral load and no detectable antibody level. We refer to this early stage as the acute or acute infection [1]. After acute infection, antibodies are produced which, along with other processes, cause the viral load to drop to a lower and more stable level. Characteristic of this later stage is detectable HIV-specific antibodies. We refer to this late stage as the chronic infection. In HIV screening, standard tests such as the rapid HIV test (OraQuick) look for antibodies and can detect only the chronic infection. For acute infection, HIV RNA based tests such as PCR (Polymerase Chain Reaction) and NAAT (nucleic acid amplification test) are used which can detect the actual HIV viruses.

Detection of early HIV infection has been named as a priority [2]. Diagnosis of acute infections—in the antibody negative "window period"—is particularly important. Individuals diagnosed at the acute infection stage have more treatment options for reducing viral loads and enhancing HIV-specific immune response, and consequently have a better chance of delaying progression to AIDS. For public health, since the acutely infected have the maximal transmission potential due to the high viral loads [3], effective early diagnosis can help reduce the spread of HIV by allowing critical prevention services at the acute infection stage [4]. Consequently, early diagnosis is an important aspect for HIV control.

As the antibody-negative window is relatively short, about 3 weeks [5], diagnosis of acute infection requires

HIV testing shortly after infection. For a population at a given time, those going for HIV testing fall into three categories: the acutely infected, the chronically affected, and the uninfected. As diagnosis of acute infection requires HIV testing within the antibody negative window, a low proportion of acute infection can be due to a low proportion of infection or delayed testing beyond the window. We propose a new metric as the ratio of the proportion of acute infection versus that of chronic infection. It measures the odds of the infected being diagnosed during acute versus chronic stage. The new metric reflects the rapidity of the infected subjects seeking and obtaining HIV testing. All things being equal, that is, the population of those who seek testing is stable over the time or locations, a high acute infection to chronic ratio is desirable.

We illustrate the new metric with data from sexually transmitted (STD) clinics in Malawi, North Carolina [6], San Francisco [7], and Washington DC [8], which were collected around the years 2000 to 2003. **Table 1** lists the number of subjects going for HIV testing, the number of diagnosed acute infections, and the number of diagnosed chronic infections. Also the listed are the per capita incomes around the time of data collection. We are interested in whether the four regions share the same acute-to-chronic ratio; that is, whether the infected individuals in these regions have the same chance of being diagnosed during acute stage.

2. Acute-to-Chronic Ratio

Denote $\{N_i, X_i, Y_i\}$ as the subjects from the i-th population who go for HIV testing over a specific period of time, where N_i is the total number screened, X_i is the number with acute infection, Y_i is the number with chronic infection, and the rest $N_i - X_i - Y_i$ subjects are free of infection. Denote P_{ai} as the proportion of acute infection and P_{ci} as the proportion of chronic infection. Then (X_i, Y_i) follows a multinomial distribution (N_i, P_{ai}, P_{ci}). The log likelihood over the $i = 1, \cdots, n$ populations takes the form

$$l(P_{ai}, P_{ci}) = \sum_{i=1}^{n} x_i \log(P_{ai}) + y_i \log(P_{ci}) + (N_i - x_i - y_i) \log(1 - P_{ai} - P_{ci}). \tag{1}$$

Maximum likelihood estimation (MLE) gives the sample proportions

$$\tilde{P}_{ai} = x_i / N_i, \text{ and } \tilde{P}_{ci} = y_i / N_i, \text{ for } i = 1, \cdots, n.$$

Define π_i as the proportion of total infection in the i-th population, so that $\pi_i = P_{ai} + P_{ci}$. Among the infected who seek a test, let T be the time from infection to HIV testing. Let R be the period from infection to detectable antibodies, also known as the antibody negative window. If $T \leq R$, an acute infection is detected. If $T > R$, a prevalent infection is detected. Let F_i be the cumulative distribution function of $T_i - R_i$, then

$$P_{ai} = \pi_i \cdot \mathrm{pr}(T_i \leq R_i | \text{infected}) = \pi_i F_i(0),$$
$$P_{ci} = \pi_i \cdot \mathrm{pr}(T_i > R_i | \text{infected}) = \pi_i \{1 - F_i(0)\}.$$

We define the acute-to-chronic (AC) ratio as

$$\theta_i = P_{ai} / P_{ci} = F_i(0) / \{1 - F_i(0)\},$$

the odds of being diagnosed during acute stage versus chronic stage among the infected who seek an HIV test.

The distribution function F reflects how quickly the infected seek HIV testing. As the exact time of infection is hard to retrieve and the length of antibody negative window is variable over populations [9,10], both T and R are not easy to determine. It is thus hard to evaluate F_i and even harder to compare F_i across populations. The AC ratio can be considered as a parameter of the distribution which captures the F information

Table 1. **STD clinic data from four regions: the total number of subjects** (N_i)**, acute infections** (X_i)**, chronic infections** (Y_i)**, and the per capita income (** $S_i \times 1000$ **US dollar).**

Population	N_i	X_i	Y_i	S_i
Malawi	1450	21	588	0.6
North Carolina	109,250	23	606	20.3
San Francisco	3789	11	81	51.8
Washington DC	1553	6	64	28.6

around 0. It quantifies the chance of an infected being diagnosed within versus beyond the acute infection stage. A major attraction of the AC ratio is that it is conveniently constructed from regular testing results in STD clinics.

Note that this ratio is free of π_i, the overall infection rate in population i. While π_i is very important to monitor, θ_i gives important complementary information about what percentage of the infected is diagnosed during acute infection stage. All things being equal, larger values of θ are more desirable.

3. MLE Inference

3.1. Common AC Ratio

If the distribution functions F_i's are the same across the populations, then $\theta = P_{ai}/P_{ci} = F(0)/\{1 - F(0)\}$ is constant for $i = 1, \cdots, n$ regardless of the infection rate π_i within each population. The log likelihood under a common AC ratio is

$$l_1(\theta, P_{ci}) = \sum_{i=1}^{n} x_i \log(\theta P_{ci}) + y_i \log(P_{ci}) + (N_i - x_i - y_i) \log(1 - \theta P_{ci} - P_{ci}). \tag{2}$$

The number of parameters under model (2) is $n+1$. Maximum likelihood estimation of l_1 gives

$$\hat{\theta} = \sum_{i=1}^{n} y_i \left/ \sum_{i=1}^{n} x_i \right., \quad \hat{P}_{ci} = \frac{(x_i + y_i)/N_i}{1 + \hat{\theta}}, \quad \hat{P}_{ai} = \hat{\theta}\hat{P}_{ci}.$$

Model (2) can hold across populations even if the risks of infection are different. To test for a constant AC ratio we specify two hypotheses,

H_0^1: $\{(X_i, Y_i, N_i), i = 1, \cdots, n\}$ share a common AC ratio following model (2), versus
H_a: $\{(X_i, Y_i, N_i), i = 1, \cdots, n\}$ have their own AC ratios following model (1).

The log-likelihood ratio test statistics for testing between these two hypotheses is

$$lr_1 = -2\left\{l_1\left(\hat{\theta}, \hat{P}_{ci}\middle| x_i, y_i, N_i\right) - l\left(\tilde{P}_{ai}, \tilde{P}_{ci}\middle| x_i, y_i, N_i\right)\right\},$$

where $\hat{\theta}$, $\hat{P}_{ci}$ are the MLEs under (2) and $\tilde{P}_{ai}$, $\tilde{P}_{ci}$ the MLEs under (1). Under H_0^1, lr_1 is asymptotically χ^2 distributed with $n-1$ degree of freedom as $N_i \to \infty$ for all i, and H_0^1 can be accordingly tested. When the N_i's are not large, we can compute the p-value following a parametric bootstrap procedure.

Step 1. Generate $\left(X_i^{(l)}, Y_i^{(l)}\right)$ from multinomial distribution $\left(N_i, \hat{\theta}\hat{P}_{ci}, \hat{P}_{ci}\right)$ for $i = 1, \cdots, n$.
Step 2. Using $\left\{\left(N_i, X_i^{(l)}, Y_i^{(l)}\right) : i = 1, \cdots, n\right\}$, estimate $\hat{\theta}^{(l)}$ and $\hat{P}_{ci}^{(l)}$ from (2) and $\tilde{P}_{ai}^{(l)}$ and $\tilde{P}_{ci}^{(l)}$ from (1);
Step 3. Compute the log-likelihood ratio

$$lr_1^{(l)} = -2\left\{l_1\left(\hat{\theta}^{(l)}, \hat{P}_{ci}^{(l)}\middle| x_i^{(l)}, y_i^{(l)}, N_i\right) - l\left(\tilde{P}_{ai}^{(l)}, \tilde{P}_{ci}^{(l)}\middle| x_i^{(l)}, y_i^{(l)}, N_i\right)\right\}.$$

Step 4. Repeat Step 1 - 3 L times. The empirical p-value for rejecting H_0 is the percentage of $lr_1^{(l)}$ above the observed log-likelihood ratio over the L repetitions.

We can also evaluate whether two groups of populations, with a common AC ratio across the populations within each group, share a common AC ratio. This may happen, for example, with one group from a developed country versus the other from a developing country, or one group with an early HIV testing campaign versus the other without such a campaign. Let $\mathcal{S}$ represent one group of populations with a common AC ratio θ_s, and $\bar{\mathcal{S}}$ represent the other group of populations with a common AC ratio $\theta_{\bar{s}}$. The log likelihood under the null hypothesis of a common AC ratio over the two groups is

$$l_0'(\theta, P_{ci}) = \sum_{i=1}^{n} x_i \log(\theta P_{ci}) + y_i \log(P_{ci}) + (N_i - x_i - y_i) \log(1 - \theta P_{ci} - P_{ci}),$$

and the log likelihood under the alternative hypothesis of group specific AC ratio is

$$l_1'(\theta_s, \theta_{\bar{s}}, P_{ci}) = \sum_{i \in \mathcal{S}} x_i \log(\theta_s P_{ci}) + y_i \log(P_{ci}) + (N_i - x_i - y_i) \log(1 - \theta_s P_{ci} - P_{ci})$$

$$+ \sum_{i \in \bar{\mathcal{S}}} x_i \log(\theta_{\bar{s}} P_{ci}) + y_i \log(P_{ci}) + (N_i - x_i - y_i) \log(1 - \theta_{\bar{s}} P_{ci} - P_{ci}).$$

Whether the two group of populations share a common AC ratio can be tested via the log likelihood ratio test.

3.2. AC Ratio Depending on a Factor

Some population factors, such as the population education level and income level, may affect the AC ratio. For example, in developed countries, a majority of HIV infected individuals have access to HIV testing at the time of acute infection, but access is limited in developing countries [11]. Such ecological analyses are subject to substantial confounding, but may be useful to generate hypotheses. Here we use simple logistic regression to explore the impact of a factor on the AC ratio. Let S_i be a factor or a index measured at the population level such as the mean income for the catchment area of the HIV clinic(s) for population i. We assume $\theta_i = F_i(0)/\{1 - F_i(0)\} = \mathcal{L}(s_i)$ for some function $\mathcal{L}$. If there is knowledge about function $\mathcal{L}$, then P_{ai} and P_{ci} can be jointly estimated with $\mathcal{L}$. As $F_i(0)$ is the probability and θ_i is the odds, it is plausible to assume a logistic relationship; that is, $F_i(0)/\{1 - F_i(0)\} = \exp(\alpha_0 + \alpha_1 S_i)$. The log-likelihood function under an AC ratio that depends on S in this fashion is

$$l_2(\theta_i, P_{ci}) = \sum_{i=1}^{n} x_i \log(P_{ci}) + (N_i - x_i - y_i)\log(1 - P_{ci}\theta_i - p_{ci}) + y_i \log(P_{ci}\theta_i),$$

$$\theta_i = \exp(\alpha_0 + \alpha_1 s_i) \tag{3}$$

Maximization of (3) with respect to P_{ci} gives

$$p_{ci} = \frac{(x_i + y_i)/N_i}{1 + \theta_i}, \quad i = 1, \cdots, n,$$

and the profile log likelihood of $\boldsymbol{\alpha} = (\alpha_0, \alpha_1)^{\mathrm{T}}$ is

$$l(\boldsymbol{\alpha}) = c + \sum_{i=1}^{n} x_i \log\{1/(1 + \exp(\alpha_0 + \alpha_1 s_i))\} + y_i \log\{\exp(\alpha_0 + \alpha_1 s_i)/(1 + \exp(\alpha_0 + \alpha_1 s_i))\},$$

with

$$c = \sum_{i=1}^{n} (N_i - x_i - y_i)\log((N_i - x_i - y_i)/N_i) + (x_i - y_i)\log((x_i + y_i)/N_i).$$

The parameter $\boldsymbol{\alpha}$ can be estimated as the maximizer of $l(\boldsymbol{\alpha})$. In fact, $l(\boldsymbol{\alpha}) - c$ corresponds to a logistic regression of I_a versus S over the $\sum_{i=1}^{n} x_i + y_i$ infected subjects with I_a the indicator variable for acute infection. Thus, $\boldsymbol{\alpha}$ can be estimated by logistic regression where, within the i-th population, the $x_i + y_i$ subjects all have $S = s_i$, x_i subjects have the value of $I_a = 1$, and y_i subjects have the value of $I_a = 0$. Denote $\hat{\alpha}_0$ and $\hat{\alpha}_1$ as the estimates from $l(\boldsymbol{\alpha})$, then

$$\hat{\theta}_i = \exp(\hat{\alpha}_0 + \hat{\alpha}_1 s_i), \quad \hat{p}_{ci} = \frac{(x_i + y_i)/N_i}{1 + \hat{\theta}_i}, \quad \hat{p}_{ai} = \hat{p}_{ci}\hat{\theta}_i$$

A test of whether the AC ratio depends on S can be formed as a likelihood ratio test with hypotheses
$H_0^2: \{(X_i, Y_i, N_i), i = 1, \cdots, n\}$ have S-related AC ratio following model (3), versus
$H_a: \{(X_i, Y_i, N_i), i = 1, \cdots, n\}$ have their own AC ratios following model (1).
The corresponding log likelihood ratio is

$$lr_2 = -2\{l_2(\hat{\boldsymbol{\alpha}}, \hat{P}_{ci}) - l(\tilde{P}_{ai}, \tilde{P}_{ci})\},$$

which is asymptotically χ^2 distributed with $2n - (n + p)$ degree of freedom with p the number of parameters in $\mathcal{L}$ which is $p = 2$ in model (3). H_0^2 can be tested via log likelihood ratio or by bootstrap similar to Section 3.1 when N_i's are not large.

4. Application to STD Clinic Data

4.1. Testing for a Common AC Ratio

We first investigate whether there is a common AC ratio across the four populations in **Table 1**. The results are shown in **Figure 1**. On the left is the scatter plot of the proportion of acute infection versus that of chronic infection. Each circle represents the observed proportion of acute infection versus that of the chronic infection, which is also the MLE estimate under model (1). Around each circle are the 95% exact confidence bounds based on
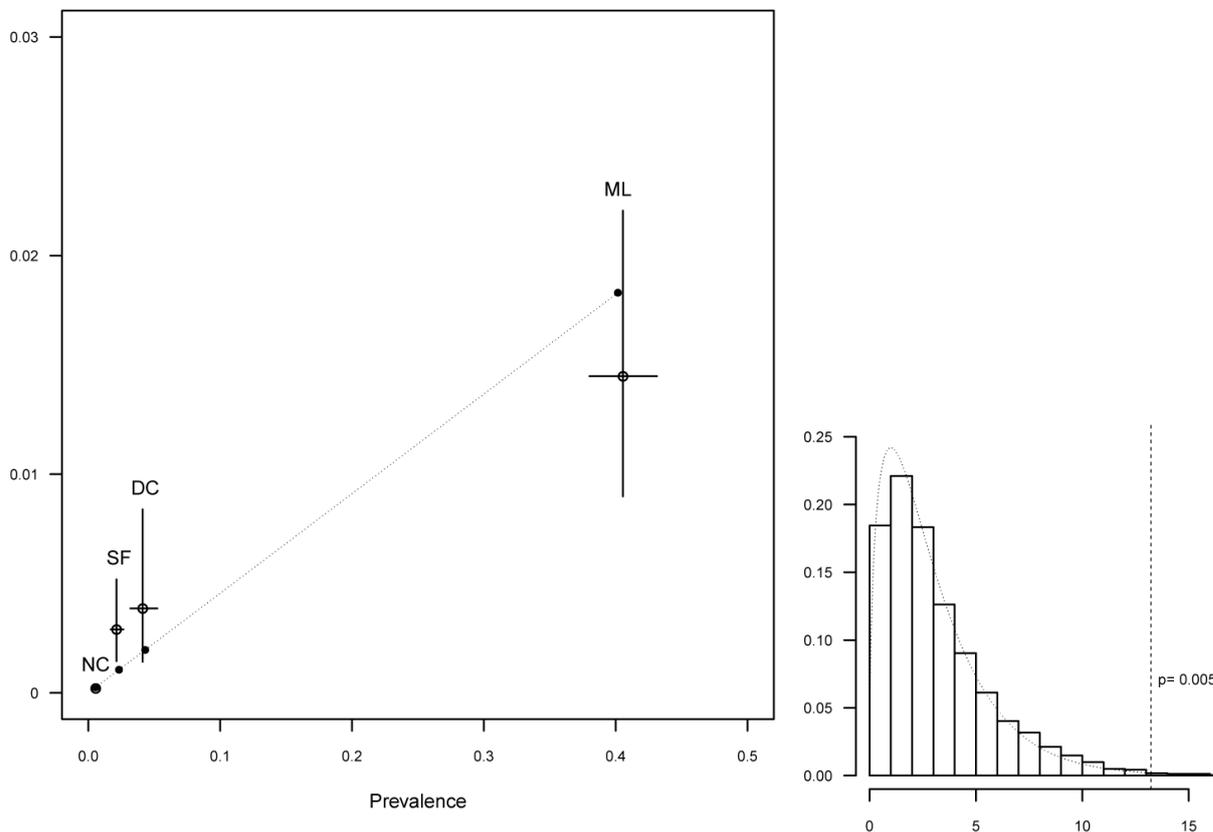
Figure 1. Test for common AC ratio over all 4 populations. On the left is the scatter plot of the proportion of acute infection P_{ai} versus the proportion of chronic infection P_{ci}: circles are the observed with 95% exact confidence bounds, solid dots are the MLE from model (2). On the right is the histogram of the log-likelihood ratio under H_0^1.

the binomial distributions (N_i, P_{ai}) and (N_i, P_{ci}) for X_i and Y_i, respectively. Each solid dot represents the MLE estimates under the common AC ratio model (2). Under (2), $\{(P_{ai}, P_{ci}), i = 1, \cdots, n\}$ are on the line through origin with slope θ. We observe the four solid dots falling on a line with slope $\hat{\theta} = 0.046$, indicating that there is about 1 acute infection for every 22 chronic infections should a common AC ratio be shared across the four populations. The observed AC ratio within each population is the slope of the line connecting the origin to the circle for that population (lines not shown in the Figure). For example, Malawi has an observed AC ratio of 0.036 indicating 1 acute infection for every 27 chronic infections; San Francisco has an observed AC ratio of 0.136 indicating 1 acute infection for every 7 chronic infections. We observe that $\hat{\theta}$ is lower than the observed AC at Washington DC and San Francisco but higher than the observed AC at North Carolina and Malawi. On the right of **Figure 1** is the histogram of the log-likelihood ratio under the null hypothesis H_0^1 from parametric bootstrap with $L = 9999$, where the over-imposed is the density of χ_3^2. The null hypothesis is rejected with empirical p value of $p = 0.005$ at the observed log-likelihood ratio of 13.23. Therefore, the four populations do not share a common AC ratio.

4.2. Testing for per Capita Income Related AC Ratio

For illustration, we investigate whether the AC ratio is related to the per capital income and follow (3). The estimation and testing results are presented in **Figure 2**. The scatter plot shows that the MLE estimates of P_{ai} and P_{ci} under model (3) are very close to the observed, all within the 95% confidence bounds. Under (3), $\{(P_{ai}, P_{ci}) : i = 1, \cdots, n\}$ and consequently the solid dots do not fall on a straight line. The empirical p value from bootstrap is $p = 0.183$, thus hypothesis H_0^2 is not rejected indicating that the AC ratio follows an increasing log-linear function of income. Since San Francisco has the highest and Malawi has the lowest per capita income, they have the highest and the lowest AC ratios, respectively. It could be that the per capital income is one cause for the differential AC ratios. It could also be that other social and demographic factors affect the AC ratio
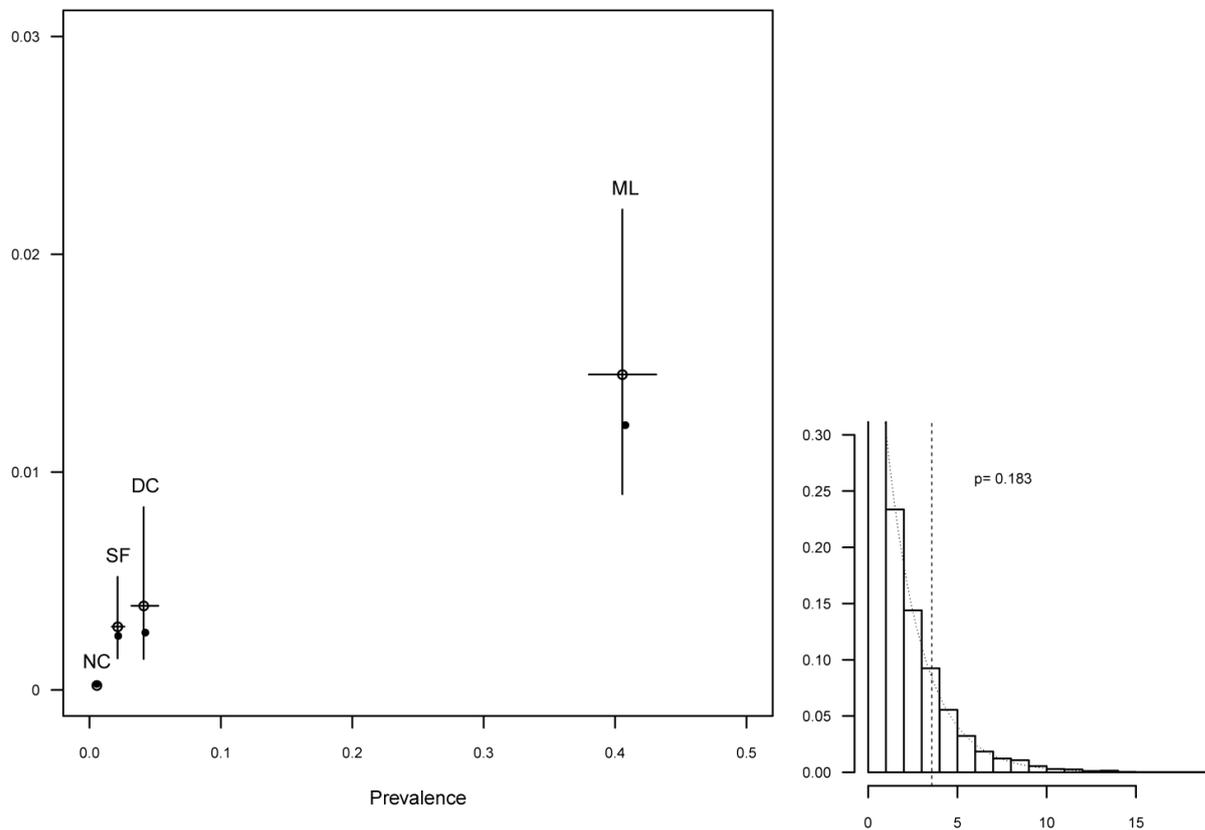
Figure 2. Test for capital income related AC ratio. On the left is the scatter plot of P_{ai} versus P_{ci} : circles are the observed with 95% confidence bounds, solid dots are the MLE from model (3). On the right is the histogram of the log likelihood ratio under H_0^2.

through the per capital income.

REFERENCES

[1] C. D. Pilcher, M. A. Price, I. F. Hoffman, S. Galvin, F. E. Martinson and P. N. Kazembe, *et al*., "Frequent Detection of Acute Primary HIV Infection in Men in Malawi," *AIDS*, Vol. 18, No. 3, 2004, pp. 517-524. http://dx.doi.org/10.1097/00002030-200402200-00019

[2] UNAIDS, "HIV Voluntary Counseling and Testing: A Gateway to Prevention and Care," *Joint United Nations Program on HIV/AIDS*, Geneva, 2002.

[3] C. D. Pilcher, J. J. Eron, P. L. Vemazza, M. Battegay, T. Harr and S. Yerly, "Sexual Transmission during the Incubation Period of Primary HIV Infection," *Journal of American Medical Association*, Vol. 286, No. 14, 2001, pp. 1713-1714. http://dx.doi.org/10.1001/jama.286.14.1713

[4] T. Flanigan and K. T. Tashima, "Diagnosis of Acute HIV Infection: It's Time to Get Moving," *Annals of Internal Medicine*, Vol. 134, No. 1, 2001, pp. 75-77. http://dx.doi.org/10.7326/0003-4819-134-1-200101020-00017

[5] R. Brookmeyer and T. C. Quinn, "Estimation of Current Human Immunodeficiency Virus Incidence Rates From a Cross-Sectional Survey Using Diagnosis Tests," *American Journal of Epidemiology*, Vol. 141, No. 2, 1995, pp. 166-172.

[6] C. D. Pilcher, S. A. Fiscus, T. Q. Nguyen, E. Foust, L. Wolf and D. Williams, "Detection of Acute Infections during HIV Testing in North Carolina," *New England Journal of Medicine*, Vol. 352, No. 18, 2005, pp. 1873-1883. http://dx.doi.org/10.1056/NEJMoa042291

[7] J. D. Klausner, R. M. Grant and C. K. Kent, "Detection of Acute HIV Infections," *New England Journal of Medicine*, Vol. 353, No. 6, 2005, pp. 631-633. http://dx.doi.org/10.1056/NEJM200508113530620

[8] A. M. Shahkolahi and P. Chiliade, "Early Diagnosis of HIV Infection at Whitman Walker Clinic STD Program Using NAATs on Pooled Blood Samples," *National STD Prevention Conference*, 9 May 2006.

[9]   S. Lindback, R. Thorstensson, A. C. Karlsson, M. Von Sydow, L. Falmholc and A. Blaxhult, "Diagnosis of Primary HIV-1 Infection and Duration of Follow-Up After HIV Exposure," *AIDS*, Vol. 14, 2000, pp. 2333-2339. http://dx.doi.org/10.1097/00002030-200010200-00014

[10] E. W. Fiebig, D. J. Wright, B. D. Rawal, P. E. Garret, R. T. Schumacher and L. Peddada, "Dynamics of HIV Viremia and Antibody Seroconversion in Plasma Donors: Implications for Diagnosis and Staging of Primary HIV Infection," *AIDS*, Vol. 17, No. 13, 2003, pp. 1871-1879. http://dx.doi.org/10.1097/00002030-200309050-00005

[11] T. A. Schacker, C. Collier, J. Hughes, T. Shea and L. Corey, "Clinical and Epidemiologic Features of Primary HIV Infection," *Annals of Internal Medicine*, Vol. 125, No. 4, 1996, pp. 257-264. http://dx.doi.org/10.7326/0003-4819-125-4-199608150-00001