



Brief Communication



OPEN ACCESS

Received: Oct 15, 2019

Accepted: Dec 9, 2019

Corresponding Author:

Jun Yong Choi, MD, PhD

Division of Infectious Diseases, Department of Internal Medicine, Yonsei University College of Medicine, 50-1, Yonsei-ro, Seodaemun-gu, Seoul 03722, Korea.

Tel: +82-2-2228-1974

Fax: +82-2-393-6884

E-mail: SERAN@yuhs.ac

Copyright © 2019 by The Korean Society of Infectious Diseases, Korean Society for Antimicrobial Therapy, and The Korean Society for AIDS

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ORCID iDs

Jun Hyung Kim

<https://orcid.org/0000-0002-0595-2110>

Jun Yong Choi

<https://orcid.org/0000-0002-2775-3315>

Funding

This research was supported by a fund of the major clinical and epidemiological indicators of HIV-infected people (Korea HIV/AIDS Cohort Study, 2019-ER5101-00) and for the Chronic Infectious Disease Cohort Study (Korea HIV/AIDS Cohort Study, 2013-E51006-00) from the Korea Centers for Disease Control and Prevention.

Comparison of Characteristics and Survival between Prospective and Retrospective Korea Human Immunodeficiency Virus/Acquired Immune Deficiency Syndrome Cohort Studies

Jun Hyung Kim ¹, Yunsu Choi², Joon Young Song³, Shin-Woo Kim⁴, Sang Il Kim⁵, Mee-Kyung Kee⁶, Bo Youl Choi², and Jun Yong Choi ¹

¹Department of Internal Medicine and AIDS Research Institute, Yonsei University College of Medicine, Seoul, Korea

²Department of Preventive Medicine, Hanyang University College of Medicine, Seoul, Korea

³Department of Internal Medicine, Korea University College of Medicine, Seoul, Korea

⁴Department of Internal Medicine, School of Medicine, Kyungpook National University, Daegu, Korea

⁵Division of Infectious Disease, Department of Internal Medicine, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, Seoul, Korea

⁶Division of Viral Disease Research, Center for Infectious Diseases Research, Korea, National Institute of Health, Cheongju, Korea

ABSTRACT

By comparing the data of prospectively and retrospectively enrolled cohorts, we evaluated whether the prospective cohort represented all patients in the retrospective cohort. The prospectively enrolled subjects were older and had lower CD4+ T cell counts, higher viral load. In addition, the initial antiretroviral treatment regimen of the prospective cohort consisted of less integrase strand transfer inhibitor-containing regimens. The 20-year survival rate was 51.8% in the prospective cohort and 84.6% in the retrospective cohort, respectively ($P = 0.844$). This study suggests the prospective cohort study may not represent all patients.

Keywords: HIV/AIDS; Cohort study; Survival rates

Prospective observational cohort studies provide useful information on the relationship between exposures and outcomes [1]. In addition, those studies are useful when randomized controlled trials are not feasible. However, selection bias is likely to occur in these cohort studies, and the bias may influence the results and limit the generalizability [2, 3]. In cohort studies, a group of individuals is sampled from a source population and followed-up over time to ascertain the occurrence of an outcome of interest [4]. Ideal patient enrollment methods would produce a cohort that represents the target populations with respect to subject demographics and core variables. However, completely representative recruitment is impossible as sampling is not always feasible due to costliness and inefficiency. Selection bias due to follow up loss also limits the internal validity of estimates derived from cohort studies



Conflict of Interest

No conflicts of interest

Author Contributions

Conceptualization: JYC. Formal analysis: BYC, YSC, JYC. Investigation: JYS, MKK, SWK, SIK, JYC. Methodology: BYC, YSC, JYC. Writing - original draft: JHK. Writing - review & editing: YSC, JYC.

[5]. Conversely, enrollment in cohort studies may also contribute towards subject outcomes because of unintended effects of the studies [6].

In 2006, the Korea Human immunodeficiency virus/Acquired immune deficiency syndrome (HIV/AIDS) cohort study (interval cohort study design) began collecting data on HIV-infected patients presenting at hospitals across Korea [7]. This prospective cohort has been an ongoing study since December 6, 2006. Among patients who visited these clinics, a limited number of patients were prospectively recruited. Patients with good retention in care or other specific characteristics were selectively enrolled for the cohort study.

To determine whether the data of prospectively enrolled subjects represent those of all HIV infected individuals in Korea, we retrospectively collected data on all HIV patients (clinical cohort study design) who visited any of the participating sites but not registered to The Korea HIV/AIDS cohort. This was done from November 2015 to August 2016.

The Korea HIV/AIDS cohort study is a multicenter prospective cohort study with ongoing enrollment of HIV-infected patients from 21 hospitals in Korea [8]. The investigators included HIV infected Korean patients older than 18 years, who were confirmed by HIV western blot and voluntarily consented to participate in this cohort study after obtaining informed consent [8]. At the time of registration, the enrolled participants were interviewed by a trained clinical researcher, and baseline information on clinical and epidemiological data was collected using a standardized protocol [7]. The prospective cohort study included questionnaires, information on medical history, physical findings, and laboratory findings, including immunological and virological status. The survey was conducted every 6 months. Enrolled subjects were among patients who visited any clinic at the participating sites. Patients were prospectively recruited from December 2006, but retrospective data on enrolled patients were also retrieved. Prospective cohort data collected the data from participants with national AIDS registry number who voluntarily consented. For patients who could not be surveyed because of withdrawal, death and hospital transfer, a follow-up investigation using national AIDS registry number was used once every year, to check whether or not the patients were deceased [8]. The Korea HIV/AIDS cohort study recruited 1,431 subjects, with over 900 in active follow-up up to December 2017.

From November 2015 to August 2016, researchers of the Korea HIV/AIDS cohort participating sites collected data of all HIV infected individuals of their institutions who have never been registered to the HIV/AIDS cohort study by retrospective medical record review. To compare with the Korea HIV/AIDS cohort study, data were retrospectively collected from all other patients seen at the participating sites. However, the retrospective data were limited to a few variables, which were collected from medical records, such as sex, age, transmission route, treatment regimen, CD4 T cell counts, HIV viral load test, mortality, and date of the last follow up. To August 2016, the retrospective data included data on 2,648 patients.

We compared the demographic and clinical characteristics of the prospectively enrolled patients with the retrospective data until August of 2016. A total of 4,079 participants are included in both of the data, 1,431 participants from prospective cohort data, and 2,648 participants from retrospective cohort data. In addition, survival rates from HIV diagnosis and factors associated with mortality were compared. Ethical approval was obtained from the Institutional Review Boards of all participating hospitals.

Data on prospectively enrolled patients and retrospective data were compared in terms of sex, age, CD4⁺ cell counts and viral loads at the time of diagnosis, initial treatment regimen, and mortality. Continuous variables are presented as means [standard deviation (SD)] or medians [interquartile range (IQR)], and categorical variables are presented as numbers and percentages. For continuous variables, a Student's *t* test or Mann-Whitney *U* test was used depending on the validity of the normality assumption. The χ -squared test or Fisher's exact test was used to assess categorical variables. Kaplan-Meier curves were used to compare survival proportion estimates considered with survival time between cohorts. A log-rank test was used to determine whether survival was significantly different between cohorts. Multiple logistic regression analysis was used to control for the effects of confounding variables and identify independent risk factors for mortality. For this analysis, variables with a *P* value <0.05 on univariate analysis and clinically important variables were entered. All *P* values were two-tailed, and a *P*-value <0.05 was considered statistically significant. All statistical analyses were performed using SAS Enterprise Guide 7.1 (SAS Institute Inc Cary, NC, USA).

A total of 1,431 and 2,648 patients were enrolled for the prospective and retrospective data, respectively. Male accounted for 93.1% of the total population, and the mean age was 39.8 years. Overall, sex distribution was similar between cohorts, but age at diagnosis was higher in the prospective cohort (41.0 ± 12.6 years *vs.* 39.0 ± 12.2 years, $P < 0.0001$; **Table 1**). Mean CD4⁺ T cell count at the initial test was lower in the prospective cohort (234 cells/mm³ *vs.* 270 cells/mm³, $P = 0.0008$). Viral load at the initial test was different between cohorts (66,243 copies/mL *vs.* 45,400 copies/mL, $P = 0.0002$). The initial antiretroviral treatment regimen consisting of integrase strand transfer inhibitor-containing regimens was less frequent in the prospective cohort (7.6% *vs.* 16.2%, $P < 0.0001$).

Overall, there were 245 deaths (6.0%), with 84 (5.9%) and 161 deaths (6.1%) occurring in the prospective and retrospective cohorts, respectively. There was no statistically significant difference in the Kaplan-Meier curve, but the 20-year survival rate was 51.8% in the prospective cohort and 84.6% in the retrospective cohort, respectively ($P = 0.844$; **Fig 1**). When the mortality rate per 1,000 patient-year was calculated by the follow-up period of infected persons, there was no difference between 10.8/1,000 patient-year in prospective cohort participants and 11.0/1,000 patient-year in retrospective cohort participants. Multiple logistic regression analyses indicated risk factors that were associated with mortality did not differ between cohorts. Older age and CD4⁺ T cell counts less than 100 cell/mm³ at the initial test were significantly associated with mortality in total study cohort (data not shown).

This study showed that the characteristics of the prospectively enrolled patients differed from all patients who visited any of the clinical sites. This may have been caused by sampling bias of both cohorts. For the prospective cohort study, patients were not entirely randomly recruited; each hospital site may have selected patients who were likely to be retained in follow-up and provided written informed consent for regular data collection. When compared with data from all patients who were not enrolled for the prospective cohort study, the prospectively enrolled subjects were older and had lower CD4⁺ T cell counts, higher viral load at the initial test. In addition, the initial antiretroviral treatment regimen of the prospective cohort comprised less integrase strand transfer inhibitor-containing regimens. The 20-year survival rate was 51.8% in the prospective cohort and 84.6% in the retrospective cohort, respectively. We should consider selection bias when interpreting results from the Korea HIV/AIDS cohort study.

Table 1. Comparison of patient characteristics between prospective and retrospective cohort studies

Characteristics	N (%)			P value
	Total (N = 4,079)	Prospective (N = 1,431)	Retrospective (N = 2,648)	
Male ^a	3,798 (93.1)	1,335 (93.3)	2,463 (93.0)	0.738
Age at HIV Diagnosis (yrs) ^{b,c}	39.8 ± 12.4	41.0 ± 12.6	39.0 ± 12.2	<0.0001
<20	70 (1.7)	14 (1.0)	56 (2.1)	<0.0001
20–29	876 (21.5)	265 (18.5)	611 (23.1)	
30–39	1,167 (28.6)	385 (26.9)	782 (29.5)	
40–49	1,089 (26.7)	387 (27.0)	702 (26.5)	
50–59	585 (14.3)	249 (17.4)	336 (12.7)	
60–69	228 (5.6)	107 (7.5)	121 (4.6)	
≥70	64 (1.6)	24 (1.7)	40 (1.5)	
Naïve ART treatment regimen ^a				
Missing	7 (0.2)	0 (0.0)	7 (0.2)	<0.0001
Untreated	983 (24.1)	417 (29.1)	566 (21.4)	
2NRTI + PI	1,606 (39.4)	552 (38.6)	1,054 (39.8)	
2NRTI + NNRTI	816 (20.0)	323 (22.6)	493 (18.6)	
2NRTI + INI	537 (13.1)	109 (7.6)	428 (16.2)	
3NRTI	5 (0.1)	1 (0.1)	4 (0.1)	
Combinations without NRTIs	12 (0.3)	5 (0.3)	7 (0.3)	
Others	113 (2.8)	24 (1.7)	89 (3.4)	
Initial CD4+ count (cell/mm ³) ^{a,d}	288 (117–541)	234.5 (89–363)	316.5 (134–741)	<0.0001
Missing	1,621 (39.7)	746 (52.0)	875 (33.0)	
≤100	612 (15.0)	185 (12.9)	427 (16.1)	<0.0001
101–200	346 (8.4)	112 (7.8)	234 (8.8)	
201–350	664 (16.3)	206 (14.3)	458 (17.3)	
351–500	364 (8.9)	108 (7.5)	256 (9.7)	
≥500	477 (11.7)	79 (5.5)	398 (15.0)	
Initial HIV viral load (copies/mL) ^{a,c}	52,427 (11,023.5–195,000)	66,243 (15,100–230,151)	45,400 (9,430–182,000)	0.0002
Missing	1,951 (47.8)	796 (55.6)	1,155 (43.6)	
≤100,000	1,377 (33.8)	388 (27.1)	989 (37.4)	<0.0001
>100,000	751 (18.4)	247 (17.3)	504 (19.0)	
Mortality				
All-Cause Mortality	245 (6.0)	84 (5.9)	161 (6.1)	0.767

^aχ²-squared test.

^bStudent's t-test (mean ± SD).

^cMean ± SE (Standard Error).

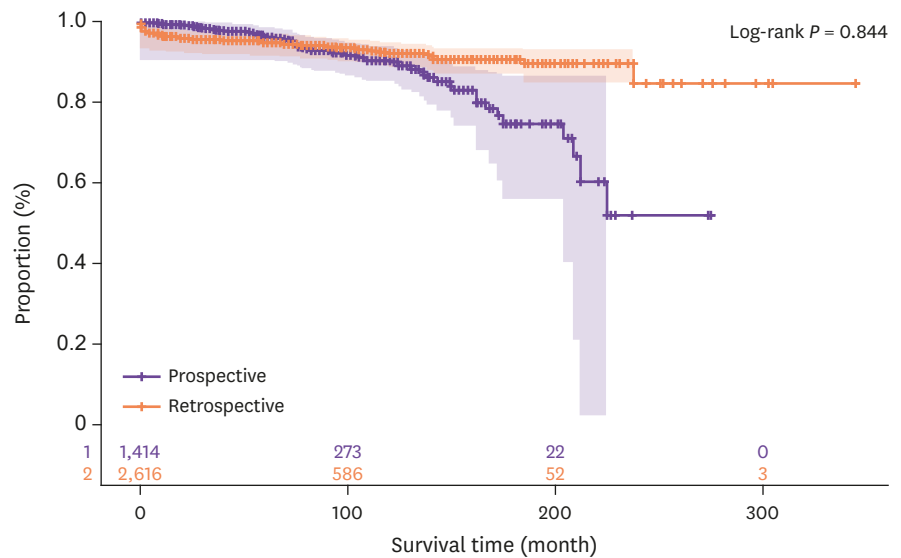
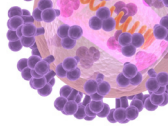
^dMann–Whitney U-test, median (interquartile range).

HIV, human immunodeficiency virus; ART, antiretroviral treatment; NRTI, nucleoside reverse transcriptase inhibitor; PI, protease inhibitor; NNRTI, non-nucleoside transcriptase inhibitor; INI, Integrase inhibitor.

A previous study showed missed clinical appointments were more frequent, and cumulative loss of time to follow-up was somewhat greater in cohort-independent patients [6]. It is possible that cohort enrollment itself promotes adherence to treatment.

There were several limitations in our study. First, the collected data for the retrospective cohort were restricted to a few variables so we could not compare other important variables associated with outcomes, including risk exposures, opportunistic diseases, and co-infections. Second, there were large proportions of missing data for some important variables. This might induce another bias. In addition, we could not perform sensitivity analyses considering the heterogeneity of the hospitals.

In spite of several limitations, our study suggests the prospective cohort study may not represent all patients who were not enrolled in the cohort study. Although random sampling is the best method for producing representative data, it is not always possible to randomly recruit subjects for a prospective cohort study [9]. An optimal method for sampling representative subjects and producing representative data should be developed. Otherwise, we must control for the selection biases that inhibit representative results.



Types of cohort studies	Total No. of subjects	No. of deaths	No. of censored cases	Censored cases (%)
Prospective	1,414	81	1,333	94.3
Retrospective	2,616	135	2,481	94.8
Total	4,030	216	3,814	94.6

Figure 1. Kaplan-Meier curve comparing survival rates of the prospective and retrospective cohorts.

ACKNOWLEDGEMENT

This study was supported by research grants for deriving the major clinical and epidemiological indicators of HIV-infected people (Korea HIV/AIDS Cohort Study, 2019-ER5101-00) and for the Chronic Infectious Disease Cohort Study (Korea HIV/AIDS Cohort Study, 2013-E51006-00) from the Korea Centers for Disease Control and Prevention.

REFERENCES

- Prentice RL. Design issues in cohort studies. *Stat Methods Med Res* 1995;4:273-92.
[PUBMED](#) | [CROSSREF](#)
- Howe CJ, Cole SR, Lau B, Napravnik S, Eron JJ Jr. Selection bias due to loss to follow up in cohort studies. *Epidemiology* 2016;27:91-7.
[PUBMED](#) | [CROSSREF](#)
- Schooling CM, Cowling BJ, Jones HE. Selection bias in cohorts of cases. *Prev Med* 2013;57:247-8.
[PUBMED](#) | [CROSSREF](#)
- Grimes DA, Schulz KF. Cohort studies: marching towards outcomes. *Lancet* 2002;359:341-5.
[PUBMED](#) | [CROSSREF](#)
- Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. *Epidemiology* 2004;15:615-25.
[PUBMED](#) | [CROSSREF](#)
- Song JY, Lee JS, Jeong HW, Choi HJ, Lee JS, Lee J, Eom JS, Cheong HJ, Chung MH, Kim WJ. Comparison of HIV-infected patients' characteristics and clinical outcomes between cohort-dependent and independent patients: Korean HIV cohort study. *J Acquir Immune Defic Syndr* 2011;57:e61-3.
[PUBMED](#) | [CROSSREF](#)

7. Kim YJ, Woo JH, Kim MJ, Park DW, Song JY, Kim SW, Choi JY, Kim JM, Han SH, Lee JS, Choi BY, Lee JS, Kim SS, Kee MK, Kang MW, Kim SI. Opportunistic diseases among HIV-infected patients: a multicenter-nationwide Korean HIV/AIDS cohort study, 2006 to 2013. *Korean J Intern Med* 2016;31:953-60.
[PUBMED](#) | [CROSSREF](#)
8. Choi BY, Choi JY, Han SH, Kim SI, Kee MK, Kim MJ, Kim SW, Kim SS, Kim YM, Ku NS, Lee JS, Lee JS, Choi Y, Park KS, Song JY, Woo JH, Kang MW, Kim J. Korea HIV/AIDS Cohort Study: study design and baseline characteristics. *Epidemiol Health* 2018;40:e2018023.
[PUBMED](#) | [CROSSREF](#)
9. Kalsbeek W, Heiss G. Building bridges between populations and samples in epidemiological studies. *Annu Rev Public Health* 2000;21:147-69.
[PUBMED](#) | [CROSSREF](#)