



# HHS Public Access

Author manuscript

*J Clin Epidemiol.* Author manuscript; available in PMC 2018 January 01.

Published in final edited form as:

*J Clin Epidemiol.* 2017 January ; 81: 129–139. doi:10.1016/j.jclinepi.2016.09.012.

## A pseudo-random patient sampling method evaluated

Nicole L. De La Mata<sup>§,1</sup>, Mi-Young Ahn<sup>2</sup>, Nagalingeswaran Kumarasamy<sup>3</sup>, Penh Sun Ly<sup>4</sup>, Oon Tek Ng<sup>5</sup>, Kinh Van Nguyen<sup>6</sup>, Tuti Parwati Merati<sup>7</sup>, Thuy Thanh Pham<sup>8</sup>, Man Po Lee<sup>9</sup>, Nicolas Durier<sup>10</sup>, and Matthew G. Law<sup>1</sup>

<sup>1</sup>Kirby Institute, UNSW Australia, Sydney, NSW, Australia <sup>2</sup>Department of Internal Medicine and AIDS Research Institute, Yonsei University College of Medicine, Severance Hospital, Seoul, South Korea <sup>3</sup>Chennai Antiviral Research and Treatment Clinical Research Site (CART CRS), YRG CARE Medical Centre, VHS, Chennai, India <sup>4</sup>National Center for HIV/AIDS, Dermatology & STDs, Phnom Penh, Cambodia <sup>5</sup>Tan Tock Seng Hospital, Singapore <sup>6</sup>National Hospital for Tropical Diseases, Hanoi, Vietnam <sup>7</sup>Faculty of Medicine Udayana University & Sanglah Hospital, Bali, Indonesia <sup>8</sup>Bach Mai Hospital, Hanoi, Vietnam <sup>9</sup>Queen Elizabeth Hospital, Hong Kong, China <sup>10</sup>TREAT Asia, amfAR - The Foundation for AIDS Research, Bangkok, Thailand

### Abstract

**Objective**—To compare two HIV cohorts to determine whether a pseudo-random sample can represent the entire study population.

<sup>§</sup>Corresponding author: Nicole L. De La Mata The Kirby Institute, UNSW Australia, Wallace Wurth Building, Sydney, 2052, Australia. (02) 9385 9012, ndelamata@kirby.unsw.edu.au.  
MYA: honeybee63@yuhs.ac, NK: kumarasamy@yrgcare.org, PSL: penhsun@nchads.org, OTN: Oon\_Tek\_NG@ttsh.com.sg, KVN: kinhnv@nhtd.vn, TPM: tutiparwati@yahoo.com, TTP: pham\_t\_thuy@yahoo.com, MPL: leemp@ha.org.hk, ND: nicolas.durier@treatasia.org, ML: Mlaw@kirby.unsw.edu.au

TAHOD-LITE study members:

PS Ly and V Khol, National Center for HIV/AIDS, Dermatology & STDs, Phnom Penh, Cambodia;  
MP Lee, PCK Li, W Lam and YT Chan, Queen Elizabeth Hospital, Hong Kong, China;  
N Kumarasamy, S Saghayam and C Ezhilarasi, Chennai Antiviral Research and Treatment Clinical Research Site (CART CRS), YRG CARE Medical Centre, VHS, Chennai, India;  
S Pujari, K Joshi, S Gaikwad and A Chitalikar, Institute of Infectious Diseases, Pune, India;  
TP Merati, DN Wirawan and F Yuliana, Faculty of Medicine Udayana University & Sanglah Hospital, Bali, Indonesia;  
OT Ng, PL Lim, LS Lee and R Martinez-Vega, Tan Tock Seng Hospital, Singapore;  
JY Choi, Na S and JM Kim, Division of Infectious Diseases, Department of Internal Medicine, Yonsei University College of Medicine, Seoul, South Korea;  
TT Pham, DD Cuong and HL Ha, Bach Mai Hospital, Hanoi, Vietnam;  
KV Nguyen, HV Bui, DTH Nguyen and DT Nguyen, National Hospital for Tropical Diseases, Hanoi, Vietnam;  
AH Sohn, N Durier and B Petersen, TREAT Asia, amfAR - The Foundation for AIDS Research, Bangkok, Thailand;  
NL De La Mata, A Jiamsakul, DC Boettiger and MG Law, The Kirby Institute, UNSW Australia, Sydney, Australia.

**Publisher's Disclaimer:** This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

### Competing Interests

The authors do not have any competing interests to declare.

### Authors' contributions

NLD, MYA and ML contributed to the concept development. KN, PSL, OTN, KVN, TPM, TTP, MPL and MYA contributed data for the analysis. NLD performed the statistical analysis and wrote the first draft of the manuscript. All authors commented on the draft manuscript and approved of the final manuscript.

**Study Design and Setting**—HIV-positive patients receiving care at 8 sites in 7 Asian countries. TAHOD pseudo-randomly selected a patient sample, while TAHOD-LITE included all patients. We compared patient demographics, CD4 count and HIV viral load testing for each cohort. Risk factors associated with CD4 count response, HIV viral load suppression (<400 copies/mL) and survival were determined for each cohort.

**Results**—There were 2318 TAHOD patients and 14714 TAHOD-LITE patients. Patient demographics, CD4 count and HIV viral load testing rates were broadly similar between the cohorts. CD4 count response and all-cause mortality were consistent among the cohorts with similar risk factors. HIV viral load response appeared to be superior in TAHOD and many risk factors differed, possibly due to viral load being tested on a subset of patients.

**Conclusions**—Our study gives the first empirical evidence that analysis of risk factors for completely ascertained endpoints from our pseudo-randomly selected patient sample may be generalized to our larger, complete population of HIV-positive patients. However, results can significantly vary when analysing smaller or pseudo-random samples, particularly if some patient data are not completely missing at random, such as viral load results.

### Keywords

Asia; HIV; patient sampling; cohort; selection bias; observational data

---

### Introduction

Observational cohort studies are useful when evaluating the relationship between health related outcomes and exposures or when randomized controlled trials (RCTs) are not always feasible or ethical to be conducted [1, 2]. However, often there is little focus on the potential pitfalls of suboptimal patient sampling methods employed in observational cohorts. Also, selection bias is more likely to occur in cohorts than RCTs and may impact upon the validity and generalizability of the study findings [3-5].

Ideal patient sampling methods would produce a representative sample of the target population, with respect to patient demographics and disease-related variables. Although favorable, completely random sampling is not always feasible as recruitment is often costly and inefficient. As such, alternate sampling methods are used that are most appropriate to the given situation, including convenience sampling, quota sampling or homogenous sampling [6, 7]. For example, observational studies in emergency departments tend to use convenience sampling where patients presenting during “business hours” are selected as more staff are available to process recruitment data [8].

In HIV research, observational cohorts have been a key epidemiological resource with the ability to assess the natural history of HIV, antiretroviral treatment (ART) use and clinical outcomes within regions or target populations [9-12]. Early cohort studies of HIV-positive homosexual men were pivotal in identifying several HIV-related biomarkers that are still used for assessing disease progression [13]. However, HIV-positive patients require lifelong treatment and so, patients’ loss to follow up (LTFU) is a prominent concern [14]. Patients

LTFU is a major source of bias in cohort studies and if large, can significantly impact upon the validity of the results [3].

Most HIV observational studies either recruit all patients seen at a clinic, or a pseudo-random subset of patients are recruited. HIV observational studies in the Asia-Pacific region utilize pseudo-random sampling for patient recruitment. In 2003, the TREAT Asia HIV Observational database (TAHOD) began collecting data on HIV-positive patients presenting at clinical sites across the Asia-Pacific region. In order to minimize costs and LTFU rates but retain heterogeneity across a very diverse region, a limited number of patients with good retention in care were consecutively recruited from a number of sites [15]. Although convenient, this pseudo-random selection method can introduce another source of bias as patients LTFU are not completely at random [16]. In 2014, the TREAT Asia HIV Observational database-Low Intensity Transfer (TAHOD-LITE), a sub-study of TAHOD, was initiated where data was collected on all patients seen at certain clinical sites, from a nominated calendar year.

These two cohorts presented an opportunity to evaluate whether pseudo-random patient sampling methods produce a representative sample and reach similar study findings to sampling of entire programs. The study objective was to compare patient demographics, pre-ART HIV biomarkers and response to ART between TAHOD and TAHOD-LITE to determine whether the TAHOD sample suitably represents all of the patients seen in TAHOD-LITE.

## Methods

### Data collection and Participants

The TREAT Asia HIV Observational database (TAHOD) is a collection of 20 HIV treatment centres across the Asia-Pacific region including China (1 site), Hong Kong (1 site), Taiwan (1 site), South Korea (1 site), India (2 sites), Indonesia (2 sites), Malaysia (2 sites), Philippines (1 site), Singapore (1 site), Thailand (4 sites), Japan (1 site), Cambodia (1 site) and Vietnam (2 sites). Detailed data were collected for a subset of patients that attend care at the sites. Patients were not entirely randomly recruited instead each site consecutively selected patients that were likely to be retained in follow up, with those receiving and not receiving ART eligible to be selected. Patients were prospectively recruited from September 2003 however retrospective data on enrolled patients were also retrieved. To date, TAHOD has recruited over 8 000 patients, with over 5 000 in active follow up to March 2015. Further description of TAHOD protocols and methods has been described elsewhere [15].

The TREAT Asia HIV Observational database Low Intensity Transfer (TAHOD-LITE) is a sub-study of TAHOD and currently involves only 8 of the 20 TAHOD sites from Cambodia (1 site), Hong Kong (1 site), India (1 site), Indonesia (1 site), Singapore (1 site), South Korea (1 site) and Vietnam (2 sites). Conversely to TAHOD, TAHOD-LITE included data from all patients seen at a site from a certain nominated calendar time point. Hence, TAHOD-LITE is a collection of previously recruited TAHOD patients and all other patients that were not recruited to TAHOD. However, patient data were limited to a few variables. To date,

TAHOD-LITE included data on over 30 000 HIV-positive adult patients, with follow up to May 2014.

Ethical approvals were obtained for TAHOD and TAHOD-LITE from Institutional Review Boards (IRB) at each site, the University of New South Wales, and the coordinating center at TREAT Asia/amfAR. Written informed consent for data collection was only retrieved if required by the site-specific IRBs.

This analysis included all patients from TAHOD-LITE aged over 18 years, who had been receiving an ART regimen consisting of three or more drugs from 01 January 2003 to 31 December 2013 and who had at least one subsequent visit after the date of ART initiation. The TAHOD cohort was represented by the patients within TAHOD-LITE that were previously recruited into TAHOD. All other remaining eligible patients were used to represent the TAHOD-LITE cohort.

The primary endpoints included CD4 cell count, HIV viral load and overall mortality from ART initiation. The secondary endpoints included patient demographics, CD4 cell count and HIV viral load prior to ART initiation and, CD4 and HIV viral load testing rates.

### Statistical analyses

TAHOD and TAHOD-LITE were compared based on the findings of the patient response to treatment, measured as CD4 cell count change, HIV viral load suppression and survival from ART initiation. Analogous to an intention-to-treat approach, we only considered the first ART regimen initiated, with all modifications to treatment after ART initiation being ignored. Risk factors were selected *a priori* and included year of ART initiation, age at ART initiation, mode of HIV exposure, pre-ART HIV viral load, pre-ART CD4 cell count, first ART regimen, previous mono/dual therapy exposure, ever hepatitis B co-infection (HBV) and ever hepatitis C co-infection (HCV).

The CD4 cell count response to treatment was examined as the change in CD4 cell count, defined as the difference between the pre-ART CD4 cell count and the CD4 cell count closest to and within 90 days of the given time point from ART initiation. We examined the median change in CD4 cell count, with interquartile range (IQR), every 6 months up to 24 months from ART initiation, by cohort. We also compared the risk factors associated with the change in CD4 cell count at 12 months from ART initiation for each cohort using a linear regression model, adjusted by clinical site.

The virological response to treatment was examined as HIV viral load suppression, defined as achieving a HIV viral load <400 copies/mL closest to and within 90 days of the given time point from ART initiation. We examined the proportion achieving HIV viral load suppression, with 95% binomial confidence interval (95% CI), every 6 months up to 24 months from ART initiation, by cohort. The risk factors associated with HIV viral load suppression at 12 months from ART initiation was examined for each cohort using logistic regression model, stratified by clinical site.

Kaplan-Meier curves were used to compare survival estimates between the cohorts. A log rank test was used to determine whether the survival was significantly different between the

cohorts. Patient follow-up was from the start of ART and censored at the date of death or most recent clinic visit date, whichever was prior. As TAHOD prospectively recruited patients, TAHOD patients were left censored at the date of study recruitment. Cox proportional hazards model was used to evaluate the risk factors associated with mortality, within each cohort and stratified by clinical site.

To assess whether risk factors differed in TAHOD and TAHOD-LITE, interaction terms between the cohort and respective risk factor were also examined in all of the models, where a Wald test was used to determine if the interaction term was significant. Descriptive statistics were used to summarize the patient demographics, by cohort. The rate of CD4 and HIV viral load testing from ART initiation was calculated for each cohort. Rates were presented as per person-year of follow up (pys), with their 95% CI.

Data were analysed using Stata version 12 (Stata Corporation, College Station, Texas, USA).

## Results

A total of 18 441 patients from TAHOD-LITE had initiated an ART regimen between 01 January 2003 and 31 December 2013 and were aged over 18 years at ART initiation. Subsequently, 1 409 patients were excluded as they did not have a further clinic visit after ART initiation. The remaining 17 032 patients were included in the analysis, where 2 318 patients were previously recruited into TAHOD and the remaining 14 714 patients were in TAHOD-LITE only.

### Patient demographics comparison

Overall, the patient demographics were relatively similar between the cohorts (Table 1). Further country comparison also showed broad similarities in the patient demographics between the cohorts (Appendix 1). The distribution for mode of HIV exposure was somewhat different between the cohorts, where 82% in TAHOD-LITE and 68% in TAHOD indicated heterosexual exposure. This mainly arose from heterogeneity between the sites rather than between cohorts as the proportion reporting heterosexual contact varied from as low as 19% to as high as 99%. Only two sites had substantial differences in the proportion reporting heterosexual contact as mode of HIV exposure (Indonesia: 84% TAHOD-LITE vs 53% TAHOD; Singapore: 49% TAHOD-LITE vs 69% TAHOD).

The percentage of patients with HBV or HCV ever positive was higher in TAHOD (TAHOD vs TAHOD-LITE: HBV positive: 9% vs 5%; HCV positive: 14% vs 5%) however TAHOD-LITE had a greater percentage of patients who had never been tested (TAHOD vs TAHOD-LITE: HBV not tested: 16% vs 49%; HCV not tested: 22% vs 54%).

### CD4 cell count response from ART initiation

A further 243 TAHOD patients and 2 279 TAHOD-LITE patients were excluded from the CD4 cell count response analysis as they did not have a pre-ART CD4 cell count to determine the CD4 change over time. Hence, a total of 2075 TAHOD patients and 12435 TAHOD-LITE patients were used to describe the CD4 change from ART initiation. Across

all countries, the median CD4 cell count change from ART initiation was similar between the cohorts (Figure 1A).

A total of 9035 patients also had a CD4 cell count at 12 months from ART initiation and were included in the linear regression model (Table 2). Of the 9 035, 7 338 patients represented TAHOD-LITE and 1 697 patients represented TAHOD. Risk factors associated with CD4 cell count change that was similar between the cohorts included age at ART initiation and pre-ART HIV viral load. Other associated risk factors included year of ART initiation, sex, mode of HIV exposure, pre-ART CD4 cell count, first-line ART regimen and previous mono/dual therapy. TAHOD-LITE identified 4 more associated risk factors than TAHOD. The main effects of the cohort on the CD4 cell count change was not significant (Table 3), nor was the interaction effects with any of the risk factors (Table 2).

### **HIV viral load response from ART initiation**

Over 80% of TAHOD-LITE and 60% of TAHOD patients did not have regular HIV viral load testing and were excluded from the HIV viral load response analysis. Across all countries, the proportion of patients with a HIV viral load <400 copies/mL from ART initiation followed a similar trend between the cohorts (Figure 1B). However, TAHOD patients did have a slightly higher proportion up to 24 months from ART initiation.

A total of 3 448 patients were included in the logistic regression model, where 2 574 patients represented TAHOD-LITE and 874 patients represented TAHOD (Table 4). Associated risk factors that were similar between the cohorts included year of ART initiation, while other associated risk factors included age at ART initiation, sex, first-line ART regimen, previous mono/dual therapy and, HBV and HCV co-infection. The main effects of cohort was significant, where TAHOD patients had 3.15 times (95% CI: 2.38, 4.16) higher adjusted odds of achieving a HIV viral load <400 copies/mL (Table 3). Interaction effects were also significant for many of the risk factors (Table 4).

### **Mortality from ART initiation**

Overall, there were 903 deaths with 787 deaths occurring in TAHOD-LITE and 116 deaths occurring in TAHOD. Kaplan-Meier survival estimates from ART initiation indicated that the probability of survival not significantly different between the cohorts ( $p$ -value=0.797), up to 4 years follow-up (Figure 1C).

All eligible patients were included in the Cox Proportional Hazards regression model (Table 5). Risk factors that were associated with mortality that were identified in both cohorts included year of ART initiation, age at ART initiation and pre-ART CD4 cell count. Six other risk factors were associated with mortality in TAHOD-LITE only. The main effect of cohort on mortality was not significant (Table 3). Interaction effects were only significant for year of ART initiation and HBV co-infection (Table 5).

### **CD4 and HIV viral load testing rates**

Across all the countries, the CD4 testing rate was slightly higher in TAHOD-LITE than TAHOD, while the HIV viral load testing rate was similar between the cohorts. The CD4

testing rate was 1.77 tests per pys (95% CI: 1.76, 1.78) in TAHOD-LITE and 1.34 tests per pys (95% CI: 1.32, 1.35) in TAHOD. The HIV testing rate was 0.65 tests per pys (95% CI: 0.65, 0.66) in TAHOD-LITE and 0.65 tests per pys (95% CI: 0.64, 0.66) in TAHOD.

## Discussion

Overall, 2 318 patients and 14 714 patients represented TAHOD and TAHOD-LITE, respectively. Detailed comparison of the patient demographics, by country and across all countries, showed somewhat of a similarity between the cohorts. The CD4 cell count response and all-cause mortality from ART initiation was consistent among the cohorts. Additionally, both cohorts identified similar risk factors associated with CD4 cell count change and overall survival at 12 months from ART initiation. There was also little evidence to suggest that being in TAHOD or TAHOD-LITE interacted with any of the risk factors to produce differing CD4 cell count change or overall survival. However, the HIV viral load response from ART initiation was superior for TAHOD patients compared to TAHOD-LITE patients. The CD4 cell count and HIV viral load testing rates were also comparable between the cohorts.

Our findings did suggest that while CD4 and survival outcomes and risk factors were broadly similar between TAHOD and TAHOD-LITE, HIV viral load response differed between the cohorts, with interaction and main effects being significant. The virological suppression being superior in TAHOD patients could have arisen from differences in standard of care, where TAHOD patients may have been more engaged in care, perhaps attending the clinic more regularly or having better adherence [17, 18]. These data were not collected, so could not be thoroughly compared between the cohorts.

However, these findings may not reflect true differences in patient outcome, but rather may be a result of the lack of data. The HIV viral load response analysis was reliant on HIV viral load results that were missing for ~80% and ~60% of TAHOD-LITE and TAHOD patients, respectively. We believe the interpretation of our findings should be that analyses of endpoints that are not routinely collected in all patients are very prone to different results and conclusions. It is difficult to see how a statistical analysis technique could untangle the biases inherent in targeted endpoints, and we would recommend that analyses are only done for subsets for patients with routine testing performed.

There were limitations in our study. Although TAHOD-LITE collects data on a substantial number of patients, it is restricted to few variables relating to patient data. This limited our ability to explore other important variables relating to our outcomes, including duration and stage of HIV-infection. In addition, some variables that were not routinely collected had large proportions of missing data, such as HBV, HCV and HIV viral load, which could introduce another source of bias. In particular, viral load results have likely arisen from targeted testing where data are not missing completely at random (MCAR). Data not MCAR can affect the validity of the analysis results as the probability of missing data is partly reliant on other unobserved factors [19]. Hence, caution is advised not to overly interpret findings relating to these variables.

Patients LTFU is an expected reality in any study, but is especially a concern in observational studies. TAHOD-LITE did not have data from all the sites on whether patients had been transferred out to other clinics. Therefore, the overall patients LTFU could not be determined in TAHOD-LITE. However, for sites that did provide patient transfer data, the percentage of patients LTFU was quite low and comparable between the cohorts. TAHOD-LITE had a slightly higher percentage LTFU at 12% compared to TAHOD which had 10% LTFU. Our study also had limited sites per country, where eight clinical sites were used to represent seven countries. Hence, our findings cannot be generalized to the entire Asia-Pacific region or specific country, but rather are a reflective of the clinical sites themselves. In addition, there was heterogeneity between the clinical sites and the results were heavily weighted by the Indian site. Nonetheless, we have accounted for the heterogeneity and unequal weighting in the multivariate models with stratification by clinical site.

Random sampling is the optimal sampling method for producing a representative sample of the study population [20]. However, in practice, it is not always feasible to utilize random sampling methods and instead alternative methods are often used [2]. We believe that our analysis is the first to evaluate whether an observational patient sample is representative of the true study population. Our study suggests that pseudo-randomly patient sample cohorts generally produce risk factor estimates that were similar to those obtained from the entire study population.

In summary, we found that our pseudo-random patient sample, TAHOD, is representative of our larger, study population, TAHOD-LITE, and importantly produces comparable findings in an analysis of response to treatment for endpoints that are routinely ascertained. However, endpoints for data that are not routinely collected or not missing completely at random does introduce bias that can significantly impact upon subsequent analyses, particularly relating to viral load suppression. As such, our analyses should be limited to routinely collected data. Thus, our study gives the first empirical confirmation that analysis of risk factors for completely ascertained endpoints can be generalized from the pseudo-randomly selected patient sample to the larger, complete patient population in our HIV cohorts.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

TAHOD-LITE (TREAT Asia HIV Observational Database Low-Intensity TransfEr) is an initiative of TREAT Asia, a program of amfAR, The Foundation for AIDS Research, with support from the U.S. National Institutes of Health's National Institute of Allergy and Infectious Diseases, Eunice Kennedy Shriver National Institute of Child Health and Human Development, and National Cancer Institute, as part of the International Epidemiologic Databases to Evaluate AIDS (IeDEA; U01AI069907). TREAT Asia is also supported by ViiV Healthcare. The Kirby Institute is funded by the Australian Government Department of Health and Ageing, and is affiliated with the Faculty of Medicine, UNSW Australia (The University of New South Wales). The content of this publication is solely the responsibility of the authors and does not necessarily represent the official views of any of the governments or institutions mentioned above.



## References

1. Prentice RL. Design issues in cohort studies. *Stat Methods Med Res.* Dec; 1995 4(4):273–92. [PubMed: 8745127]
2. Soh SE, Saw SM. Cohort studies: design and pitfalls. *Am J Ophthalmol.* Jul; 2010 150(1):3–5. [PubMed: 20609703]
3. Howe CJ, Cole SR, Lau B, Napravnik S, Eron JJ Jr. Selection bias due to loss to follow up in cohort studies. *Epidemiology.* Oct 19.2015
4. Pizzi C, De Stavola BL, Pearce N, Lazzarato F, Ghiotti P, Merletti F, et al. Selection bias and patterns of confounding in cohort studies: the case of the NINFEA web-based birth cohort. *J Epidemiol Community Health.* Nov; 2012 66(11):976–81. [PubMed: 22147750]
5. Schooling CM, Cowling BJ, Jones HE. Selection bias in cohorts of cases. *Prev Med.* Sep; 2013 57(3):247–8. [PubMed: 23747980]
6. Bornstein MH, Jager J, Putnick DL. Sampling in Developmental Science: Situations, Shortcomings, Solutions, and Standards. *Dev Rev.* Dec; 2013 33(4):357–70. [PubMed: 25580049]
7. Suresh K, Thomas SV, Suresh G. Design, data analysis and sampling techniques for clinical research. *Ann Indian Acad Neurol.* Oct; 2011 14(4):287–90. [PubMed: 22346019]
8. Valley MA, Heard KJ, Ginde AA, Lezotte DC, Lowenstein SR. Observational studies of patients in the emergency department: a comparison of 4 sampling methods. *Ann Emerg Med.* Aug; 2012 60(2):139–45 e1. [PubMed: 22401950]
9. Hendel H, Cho YY, Gauthier N, Rappaport J, Schachter F, Zagury JF. Contribution of cohort studies in understanding HIV pathogenesis: introduction of the GRIV cohort and preliminary results. *Biomed Pharmacother.* 1996; 50(10):480–7. [PubMed: 9091061]
10. Pantazis N, Morrison C, Amornkul PN, Lewden C, Salata RA, Minga A, et al. Differences in HIV natural history among African and non-African seroconverters in Europe and seroconverters in sub-Saharan Africa. *PLoS One.* 2012; 7(3):e32369. [PubMed: 22412867]
11. Petoumenos K, Australian HIVOD. The role of observational data in monitoring trends in antiretroviral treatment and HIV disease stage: results from the Australian HIV observational database. *J Clin Virol.* Feb; 2003 26(2):209–22. [PubMed: 12600652]
12. Samet JM, Munoz A. Evolution of the cohort study. *Epidemiol Rev.* 1998; 20(1):1–14. [PubMed: 9762505]
13. Polk BF, Fox R, Brookmeyer R, Kanchanaraksa S, Kaslow R, Visscher B, et al. Predictors of the acquired immunodeficiency syndrome developing in a cohort of seropositive homosexual men. *N Engl J Med.* Jan 8; 1987 316(2):61–6. [PubMed: 3024007]
14. McManus H, Petoumenos K, Brown K, Baker D, Russell D, Read T, et al. Loss to follow-up in the Australian HIV Observational Database. *Antivir Ther.* Nov 7.2014
15. Zhou J, Kumarasamy N, Ditangco R, Kamarulzaman A, Lee CK, Li PC, et al. The TREAT Asia HIV Observational Database: baseline and retrospective data. *J Acquir Immune Defic Syndr.* Feb 1; 2005 38(2):174–9. [PubMed: 15671802]
16. Berger VW. Conservative handling of missing information. *J Clin Epidemiol.* Nov; 2012 65(11):1237–8. [PubMed: 23017643]
17. El-Khatib Z, Ekstrom AM, Coovadia A, Abrams EJ, Petzold M, Katzenstein D, et al. Adherence and virologic suppression during the first 24 weeks on antiretroviral therapy among women in Johannesburg, South Africa - a prospective cohort study. *BMC Public Health.* 2011; 11:88. [PubMed: 21303548]
18. Shah B, Walshe L, Saple DG, Mehta SH, Ramnani JP, Kharkar RD, et al. Adherence to antiretroviral therapy and virologic suppression among HIV-infected persons receiving care in private clinics in Mumbai, India. *Clin Infect Dis.* May 1; 2007 44(9):1235–44. [PubMed: 17407045]
19. Norris CM, Ghali WA, Knudtson ML, Naylor CD, Saunders LD. Dealing with missing data in observational health care outcome analyses. *J Clin Epidemiol.* Apr; 2000 53(4):377–83. [PubMed: 10785568]
20. Kalsbeek W, Heiss G. Building bridges between populations and samples in epidemiological studies. *Annu Rev Public Health.* 2000; 21:147–69. [PubMed: 10884950]

### What is new?

#### Key findings

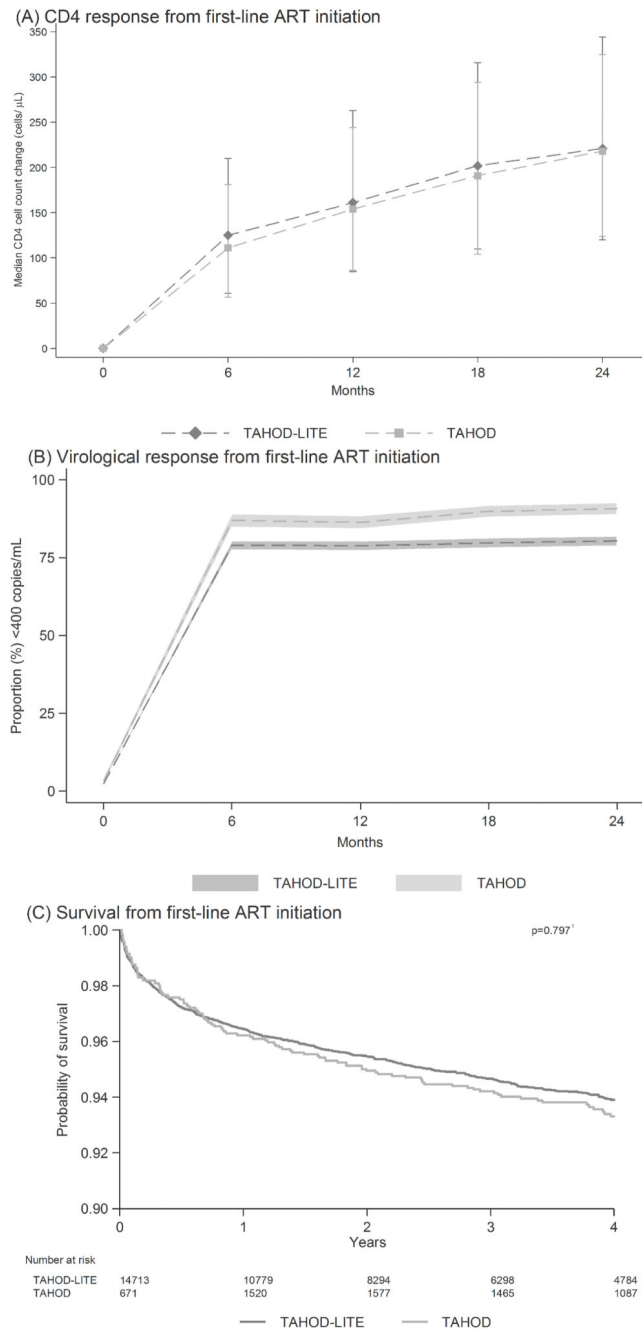
- Our pseudo-random patient sample, TAHOD, is representative of our larger study population, TAHOD-LITE, and produced comparable findings for routinely ascertained endpoints relating to the response to antiretroviral treatment.
- Endpoints that were not routinely collected or not missing completely at random may introduce bias that can significantly impact upon subsequent analyses, such as viral load suppression

#### What this study adds to what was known?

- Pseudo-random sampling of patients in observational cohorts is common, but can introduce bias. Our findings provide the first empirical evidence that a pseudo-random sample can produce comparable results seen in the larger, entire study population.

#### What is the implication and what should change now?

- Pseudo-random sampling methods should not be dismissed where random sampling is impractical, as analyses relating to routinely collected data may still produce comparable results.



**Figure 1. The response to first-line ART, by cohort**

(A) Median CD4 cell count change (cells/ $\mu$ L), with interquartile range, from ART initiation across all countries, by cohort. (B) Proportion of patients with HIV viral load <400 copies/mL since ART initiation across all countries, by cohort, where the shaded region represents the 95% binomial proportion confidence interval. (C) Probability of survival, as from Kaplan-Meier estimates, from ART initiation across all countries, by cohort ( $\tau$  = log rank test).

**Table 1**  
**Demographics of the patients by cohort**

	TAHOD-LITE n=14714			TAHOD n=2318		
	n	(%)	(%) <sup>1</sup>	n	(%)	(%) <sup>1</sup>
<b>Year of ART initiation</b>						
2003-05	2367	(16)		577	(25)	
2006-09	5500	(37)		748	(32)	
2010-13	6847	(47)		993	(43)	
<b>Age</b>						
30	3692	(25)		646	(28)	
31-40	6639	(45)		1033	(45)	
41-50	2922	(20)		439	(19)	
51 +	1461	(10)		200	(9)	
Median [IQR]	35	[30, 42]		35	[30, 41]	
<b>Sex</b>						
Male	10081	(69)		1610	(69)	
Female	4619	(31)		706	(30)	
Transgender	14	(<0.1)		2	(<0.1)	
<b>Mode of HIV exposure</b>						
Heterosexual	12026	(82)		1576	(68)	
Homosexual	1031	(7)		275	(12)	
Injecting drug user	527	(4)		255	(11)	
Other/Unknown	1130	(8)		212	(9)	
<b>Hepatitis C Co-infection<sup>2</sup></b>						
Positive	717	(5)	(11)	313	(14)	(17)
Negative	6029	(41)	(89)	1486	(64)	(83)
Not tested	7968	(54)	-	519	(22)	-
<b>Hepatitis B Co-infection<sup>3</sup></b>						
Positive	694	(5)	(9)	198	(9)	(10)
Negative	6824	(46)	(91)	1744	(75)	(90)
Not tested	7196	(49)	-	376	(16)	-
<b>Pre-ART CD4 (cells/<math>\mu</math>L)</b>						
50	2988	(20)	(24)	659	(28)	(32)
51-100	1997	(14)	(16)	298	(13)	(14)
101-200	3367	(23)	(27)	494	(21)	(24)
>200	4084	(28)	(33)	624	(27)	(30)
Not tested	2278	(15)	-	243	(10)	-
Median [IQR]	138	[53, 232]		116	[35, 216]	
<b>Pre-ART viral load (copies/mL)</b>						
10 <sup>5</sup>	1314	(9)	(47)	414	(18)	(51)

	TAHOD-LITE n=14714			TAHOD n=2318		
	n	(%)	(%) <sup>1</sup>	n	(%)	(%) <sup>1</sup>
>10 <sup>5</sup>	1491	(10)	(53)	402	(17)	(49)
Not tested	11909	(81)	-	1502	(65)	-
Median [IQR]	115000	[31791, 384000]		94675	[25900, 322000]	
<b>First ART regimen</b>						
NRTI+NNRTI <sup>4</sup>	13988	(95)		2105	(91)	
NRTI+PI <sup>5</sup>	632	(4)		188	(8)	
Other <sup>6</sup>	94	(1)		25	(1)	
<b>Previous mono/dual therapy</b>						
No	14217	(97)		2200	(95)	
Yes	497	(3)		118	(5)	

<sup>1</sup>Column percentages excluding frequencies for not tested.

<sup>2</sup>Hepatitis C antibody result where positive indicates ever positive result.

<sup>3</sup>Hepatitis B surface antigen result where positive indicates ever positive result

<sup>4</sup>Regimen combination consisting of nucleoside reverse transcriptase inhibitors (NRTIs) and nonnucleoside reverse transcriptase inhibitor (NNRTI).

<sup>5</sup>Regimen combination consisting of nucleoside reverse transcriptase inhibitors (NRTIs) and protease inhibitor (PI).

<sup>6</sup>Any regimen combination excluding NRTIs+NNRTI or NNRTIs+PI.

**Table 2**  
**Risk factors associated with CD4 change at 12 months from ART initiation, stratified by cohort**

	TAHOD-LITE n = 7 338				TAHOD n = 1 697				Interaction p value <sup>1</sup>	
	N	(%) <sup>2</sup>	Mean Diff. <sup>3</sup>	95% CI	P value <sup>4</sup>	N	(%) <sup>2</sup>	Mean Diff. <sup>3</sup>		95% CI
<b>Year of ART Initiation</b>					0.475					0.534
2003-2005	948	(13)	0	(-25, 2)	0.086	415	(24)	0	(-22, 16)	0.765
2006-2009	2938	(40)	-12	(-15, 13)	0.898	619	(36)	-3	(-0, 47)	<b>0.050</b>
2010-2013	3452	(47)	-1			663	(39)	<b>24</b>		
<b>Age at ART initiation (years)</b>					<0.001					0.585
30	1707	(23)	0	(-20, 0)	0.058	436	(26)	0	(-25, 10)	0.408
31-40	3213	(44)	-10	(-35, -10)	<0.001	748	(44)	-7	(-51, -8)	<b>0.007</b>
41-50	1548	(21)	-23	(-36, -6)	0.006	350	(21)	-14	(-41, 14)	0.333
51+	870	(12)	-21			163	(10)			
<b>Sex</b>										0.195
Male	4959	(68)	0	(10, 29)	<0.001	1179	(69)	0	(-11, 23)	0.514
Female	2379	(32)	20			518	(31)	6		
<b>Mode of HIV Exposure</b>					0.023					0.850
Heterosexual contact	5787	(79)	0	(3, 37)	0.020	1184	(70)	0	(-16, 37)	0.440
Homosexual contact	698	(10)	20	(-51, 4)	0.090	234	(14)	-6	(-42, 30)	0.743
Injecting drug use	266	(4)	-24	(-17, 13)	0.768	123	(7)	11	(-17, 39)	0.447
Other/unknown	587	(8)	-2			156	(9)			
<b>Pre-ART viral load (copies/mL)</b>					<0.001					0.962
100000	850	(12)	0	(19, 51)	<0.001	356	(21)	0	(33, 79)	<0.001
>100000	1011	(14)	35	(3, 34)	0.017	348	(21)	56	(3, 59)	<b>0.030</b>
Not tested	5477	(75)	-18			993	(59)	-31		
<b>Pre-ART CD4 (cells/<math>\mu</math>L)</b>					<0.001					0.537
										0.254

	TAHOD-LITE n =7 338					TAHOD n =1 697					Interaction p value <sup>1</sup>
	N (%) <sup>2</sup>	(%) <sup>3</sup>	Mean Diff.	95% CI	P value <sup>4</sup>	N (%) <sup>2</sup>	(%) <sup>3</sup>	Mean Diff.	95% CI	P value <sup>4</sup>	
50	1790 (24)	0	0		0.078	524 (31)	0			0.710	0.468
51-100	1076 (15)	10	10	(-3, 23)	0.140	231 (14)	3	(-20, 25)		0.829	
101-200	1943 (26)	3	3	(-9, 14)	0.622	410 (24)	6	(-14, 26)		0.564	
201+	2529 (34)	-17	-17	(-29, -6)	<b>0.003</b>	532 (31)	-9	(-29, 12)		0.403	
<b>First ART regimen</b>											
NRTI+NNRTI <sup>5</sup>	6934 (94)	0	0		0.078	1510 (89)	0			0.710	0.468
NRTI+PI <sup>6</sup>	351 (5)	1	1	(-19, 20)	0.960	167 (10)	12	(-18, 42)		0.440	
Other <sup>7</sup>	53 (1)	<b>55</b>	<b>55</b>	<b>(7, 104)</b>	<b>0.025</b>	20 (1)	-7	(-74, 61)		0.844	
<b>Previous mono/duo exposure</b>											
No	7099 (97)	0	0		0.363	1618 (95)	0			0.229	0.654
Yes	239 (3)	<b>-44</b>	<b>-44</b>	<b>(-66, -21)</b>	<b>&lt;0.001</b>	79 (5)	-29	(-64, 7)		0.118	
<b>Hepatitis B co-infection<sup>8</sup></b>											
Negative	4328 (59)	(91)	0		0.363	1336 (79)	0			0.229	0.796
Positive	443 (6)	(9)	-9	(-26, 8)	0.282	158 (9)	-3	(-27, 21)		0.821	
Not tested	2567 (35)	-	-8	(-24, 8)	0.312	203 (12)	-30	(-64, 4)		0.087	
<b>Hepatitis C co-infection<sup>9</sup></b>											
Negative	3924 (53)	(90)	0		0.256	1192 (70)	0			0.108	0.754
Positive	431 (6)	(10)	-15	(-37, 6)	0.162	194 (11)	-24	(-53, 5)		0.103	
Not tested	2983 (41)	-	7	(-11, 25)	0.436	311 (18)	19	(-12, 49)		0.228	

<sup>1</sup>Wald test for the interaction term between cohort (ie. TAHOD or TAHOD-LITE) and the respective risk factor.

<sup>2</sup>Column percentages for the respective risk factor, by cohort.

<sup>3</sup>Column percentages excluding frequencies for not tested in the respective risk factor, by cohort.

<sup>4</sup>Wald test for each level and global of the respective risk factor. Global p values for year of ART initiation, age and pre-ART CD4 count are test for trend while all other global p values are test for heterogeneity.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

- 5. Regimen combination consisting of nucleoside reverse transcriptase inhibitors (NRTIs) and nonnucleoside reverse transcriptase inhibitor (NNRTI).
- 6. Regimen combination consisting of nucleoside reverse transcriptase inhibitors (NRTIs) and protease inhibitor (PI).
- 7. Any regimen combination excluding NRTIs+NNRTI or NRTIs+PI.
- 8. Hepatitis B surface antigen result where positive indicates ever positive result.
- 9. Hepatitis C antibody result where positive indicates ever positive result.



**Table 3**  
**Summary of the main effect of cohort on the treatment response**

(A) CD4 cell count change at 12 months from ART initiation. (B) HIV viral load <400 copies/mL at 12 months from ART initiation. (C) All-cause mortality from ART initiation.

	Unadjusted			Adjusted <sup>1</sup>		
	Mean Diff.	95% CI	P <sup>2</sup> value	Mean Diff.	95% CI	P <sup>2</sup> value
<b>(A) CD4 Response</b>						
TAHOD-LITE	0			0		
TAHOD	-6	(-15, 4)	0.235	-4	(-14, 5)	0.359
	<b>OR</b>	<b>95% CI</b>	<b>P<sup>2</sup> value</b>	<b>OR</b>	<b>95% CI</b>	<b>P<sup>2</sup> value</b>
<b>(B) Viral suppression</b>						
TAHOD-LITE	1.00			1.00		
TAHOD	2.36	(1.83, 3.03)	<0.001	3.15	(2.38, 4.16)	<0.001
	<b>HR</b>	<b>95% CI</b>	<b>P<sup>2</sup> value</b>	<b>HR</b>	<b>95% CI</b>	<b>P<sup>2</sup> value</b>
<b>(C) Mortality</b>						
TAHOD-LITE	1.00			1.00		
TAHOD	0.90	(0.72, 1.13)	0.366	0.93	(0.74, 1.17)	0.556

<sup>1</sup> Adjusted for year of ART initiation, age at ART initiation, sex, mode of HIV exposure, pre-ART HIV viral load, pre-ART CD4 cell count, first-line ART regimen, previous mono/duo therapy exposure, hepatitis B co-infection (ever) and hepatitis C co-infection (ever).

<sup>2</sup>Wald test.

**Table 4**  
**Risk factors associated with HIV viral load < 400 copies/mL at 12 months from ART initiation, stratified by cohort**

	TAHOD-LITE n = 2 574				TAHOD n = 874				Interaction p value <sup>1</sup>	
	N	% with VL<400 copies/mL <sup>2</sup>	OR	95% CI	p value <sup>3</sup>	N	% with VL<400 copies/mL <sup>2</sup>	OR		95% CI
<b>Year of ART Initiation</b>					<0.001					<0.001
2003-2005	187	(4)	1.00			143	(14)	1.00		
2006-2009	628	(18)	<b>3.05</b>	<b>(1.82, 5.09)</b>	<0.001	349	(43)	<b>4.07</b>	<b>(2.13, 7.79)</b>	<0.001
2010-2013	1759	(78)	<b>20.14</b>	<b>(11.96, 33.92)</b>	<0.001	382	(44)	<b>6.37</b>	<b>(3.05, 13.30)</b>	<0.001
<b>Age at ART initiation (years)</b>					0.920					<b>0.031</b>
30	571	(22)	1.00			193	(20)	1.00		
31-40	1023	(38)	1.16	(0.83, 1.63)	0.379	367	(43)	<b>2.51</b>	<b>(1.40, 4.48)</b>	<b>0.002</b>
41-50	569	(23)	1.20	(0.80, 1.79)	0.380	212	(25)	1.85	(0.96, 3.59)	0.067
51+	411	(17)	0.98	(0.62, 1.56)	0.935	102	(12)	<b>2.83</b>	<b>(1.13, 7.11)</b>	<b>0.027</b>
<b>Sex</b>										0.686
Male	1995	(79)	1.00			694	(80)	1.00		
Female	579	(21)	<b>1.42</b>	<b>(1.03, 1.95)</b>	<b>0.031</b>	180	(20)	1.05	(0.56, 1.94)	0.883
<b>Mode of HIV Exposure</b>					0.303					<0.001
Heterosexual contact	1550	(54)	1.00			470	(54)	1.00		
Homosexual contact	581	(27)	1.53	(0.96, 2.45)	0.072	223	(27)	1.04	(0.51, 2.11)	0.921
Injecting drug use	135	(6)	1.15	(0.53, 2.52)	0.719	54	(4)	0.65	(0.21, 2.04)	0.458
Other/unknown	308	(13)	1.25	(0.82, 1.91)	0.306	127	(15)	1.06	(0.49, 2.33)	0.879
<b>Pre-ART HIV viral load (copies/mL)</b>					0.596					<0.001
100000	712	(28)	1.00			335	(41)	1.00		
>100000	782	(30)	1.25	(0.82, 1.91)	0.306	312	(37)	0.79	(0.42, 1.47)	0.458
Not Tested	1080	(43)	0.85	(0.59, 1.22)	0.372	227	(21)	0.70	(0.31, 1.58)	0.391
<b>Pre-ART CD4 cell count</b>					0.882					<0.001

	TAHOD-LITE n = 2 574				TAHOD n = 874				Interaction p value <sup>1</sup>	
	N	% with VL<400 copies/mL <sup>2</sup>	OR	95% CI	P value <sup>3</sup>	N	% with VL<400 copies/mL <sup>2</sup>	OR		95% CI
(cells/ $\mu$ L)										
50	583	(25)	1.00			209	(22)	1.00		
51-100	282	(11)	1.03	(0.64, 1.66)	0.892	101	(11)	1.26	(0.60, 2.67)	0.544
101-200	437	(16)	1.10	(0.72, 1.69)	0.664	187	(22)	1.67	(0.84, 3.33)	0.141
201+	1102	(41)	1.11	(0.75, 1.65)	0.591	339	(41)	1.34	(0.69, 2.60)	0.383
Not Tested	169	(6)	0.85	(0.49, 1.48)	0.576	38	(4)	1.42	(0.45, 4.44)	0.548
<b>First-line regimen</b>					<b>0.021</b>					<b>0.667</b>
NRTIs+NNRTI <sup>4</sup>	2252	(88)	1.00			692	(78)	1.00		
NRTIs+PI <sup>5</sup>	276	(10)	<b>0.52</b>	<b>(0.32, 0.83)</b>	<b>0.006</b>	167	(21)	1.38	(0.59, 3.23)	0.462
Other <sup>6</sup>	46	(2)	0.58	(0.16, 2.04)	0.392	15	(2)	1.77	(0.27, 11.65)	0.551
<b>Previous mono/duo exposure</b>										<b>&lt;0.001</b>
No	2513	(98)	1.00			834	(97)	1.00		
Yes	61	(2)	1.82	(0.79, 4.22)	0.162	40	(3)	<b>0.22</b>	<b>(0.09, 0.55)</b>	<b>0.001</b>
<b>Hepatitis B co-infection<sup>7</sup></b>					<b>&lt;0.001</b>					0.922
Negative	1996	(81)	1.00			706	(82)	1.00		
Positive	186	(8)	1.10	(0.64, 1.88)	0.724	63	(7)	0.88	(0.39, 2.00)	0.761
Not Tested	392	(11)	<b>3.35</b>	<b>(2.17, 5.19)</b>	<b>&lt;0.001</b>	105	(11)	0.87	(0.32, 2.37)	0.783
<b>Hepatitis C co-infection<sup>8</sup></b>					<b>&lt;0.001</b>					<b>0.978</b>
Negative	1755	(78)	1.00			590	(70)	1.00		
Positive	209	(9)	0.73	(0.39, 1.38)	0.335	73	(6)	1.09	(0.36, 3.33)	0.876
Not Tested	610	(13)	<b>0.22</b>	<b>(0.15, 0.34)</b>	<b>&lt;0.001</b>	211	(24)	0.95	(0.38, 2.38)	0.907

<sup>1</sup>Wald test for the interaction term between cohort (ie. TAHOD or TAHOD-LITE) and the respective risk factor.

<sup>2</sup>Column percentages of the patients with HIV viral load <400 copies/mL for the respective risk factor, by cohort.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

<sup>3</sup>Wald test for each level and global of the respective risk factor. Global p values for year of ART initiation, age and pre-ART CD4 count are test for trend while all other global p values are test for heterogeneity.

<sup>4</sup>Regimen combination consisting of nucleoside reverse transcriptase inhibitors (NRTIs) and nonnucleoside reverse transcriptase inhibitor (NNRTI).

<sup>5</sup>Regimen combination consisting of nucleoside reverse transcriptase inhibitors (NRTIs) and protease inhibitor (PI).

<sup>6</sup>Any regimen combination excluding NRTIs+NNRTI or NRTIs+PI.

<sup>7</sup>Hepatitis B surface antigen result where positive indicates ever positive result.

<sup>8</sup>Hepatitis C antibody result where positive indicates ever positive result.

**Table 5**  
**Risk factors associated with mortality from ART initiation, stratified by cohort**

	TAHOD-LITE n = 14 713				TAHOD n = 2 284				Interaction p value <sup>1</sup>		
	Total Patients	% of deaths <sup>2</sup>	HR	95% CI	P value <sup>3</sup>	Total Patients	% of deaths <sup>2</sup>	HR		95% CI	P value <sup>3</sup>
<b>Year of ART Initiation</b>					<b>0.044</b>					<b>0.048</b>	<b>0.015</b>
2003-2005	2367	(27)	1.00			562	(46)	1.00			
2006-2009	5500	(41)	0.83	(0.69, 1.01)	0.067	743	(27)	0.72	(0.41, 1.26)	0.246	
2010-2013	6846	(32)	<b>0.79</b>	<b>(0.63, 0.99)</b>	<b>0.044</b>	979	(27)	0.45	(0.20, 1.01)	0.054	
<b>Age at ART initiation (years)</b>					< <b>0.001</b>					<b>0.003</b>	<b>0.877</b>
30	3692	(18)	1.00			638	(26)	1.00			
31-40	6638	(41)	1.08	(0.88, 1.32)	0.458	1016	(38)	1.11	(0.68, 1.79)	0.680	
41-50	2922	(21)	1.23	(0.98, 1.56)	0.077	430	(19)	1.17	(0.65, 2.11)	0.604	
51+	1461	(20)	<b>2.23</b>	<b>(1.74, 2.86)</b>	< <b>0.001</b>	200	(17)	<b>3.35</b>	<b>(1.76, 6.37)</b>	< <b>0.001</b>	
<b>Sex</b>											<b>0.461</b>
Male	10094	(80)	1.00			1584	(80)	1.00			
Female	4619	(20)	<b>0.72</b>	<b>(0.60, 0.87)</b>	<b>0.001</b>	700	(20)	0.82	(0.49, 1.35)	0.431	
<b>Mode of HIV Exposure</b>					< <b>0.001</b>					0.517	0.463
Heterosexual contact	12025	(80)	1.00			1552	(59)	1.00			
Homosexual contact	1031	(4)	<b>0.64</b>	<b>(0.44, 0.94)</b>	<b>0.025</b>	271	(11)	1.40	(0.69, 2.84)	0.356	
Injecting drug use	527	(8)	<b>1.91</b>	<b>(1.29, 2.85)</b>	<b>0.001</b>	250	(21)	1.50	(0.75, 3.02)	0.253	
Other/unknown	1130	(8)	1.04	(0.80, 1.37)	0.760	211	(9)	1.42	(0.72, 2.82)	0.310	
<b>Pre-ART HIV viral load (copies/mL)</b>					< <b>0.001</b>					0.614	0.910
100000	1314	(6)	1.00			410	(12)	1.00			
>100000	1491	(12)	1.18	(0.82, 1.70)	0.366	398	(18)	0.84	(0.41, 1.72)	0.628	
Not tested	11908	(82)	<b>1.78</b>	<b>(1.26, 2.52)</b>	< <b>0.001</b>	1476	(70)	1.16	(0.52, 2.59)	0.717	
<b>Pre-ART CD4 cell count (cells/<math>\mu</math>L)</b>					< <b>0.001</b>					< <b>0.001</b>	0.394

	TAHOD-LITE n = 14 713					TAHOD n = 2 284					Interaction p value <sup>1</sup>
	Total Patients	% of deaths <sup>2</sup>	HR	95% CI	P value <sup>3</sup>	Total Patients	% of deaths <sup>2</sup>	HR	95% CI	P value <sup>3</sup>	
50	2987	(35)	1.00			653	(49)	1.00			
51-100	1997	(22)	0.90	(0.74, 1.10)	0.309	291	(15)	0.61	(0.35, 1.06)	0.079	
101-200	3367	(19)	<b>0.48</b>	<b>(0.39, 0.59)</b>	<b>&lt;0.001</b>	490	(21)	<b>0.58</b>	<b>(0.35, 0.94)</b>	<b>0.028</b>	
201+	4083	(10)	<b>0.26</b>	<b>(0.20, 0.33)</b>	<b>&lt;0.001</b>	618	(7)	<b>0.17</b>	<b>(0.08, 0.38)</b>	<b>&lt;0.001</b>	
Not tested	2278	(14)	<b>0.53</b>	<b>(0.42, 0.67)</b>	<b>&lt;0.001</b>	232	(8)	<b>0.40</b>	<b>(0.19, 0.85)</b>	<b>0.017</b>	
<b>First-line regimen</b>					<b>0.048</b>					0.996	0.560
NRTIs+NNRTI <sup>4</sup>	13987	(94)	1.00			2072	(95)	1.00			
NRTIs+PI <sup>5</sup>	632	(6)	<b>1.52</b>	<b>(1.08, 2.13)</b>	<b>0.015</b>	187	(5)	1.04	(0.39, 2.77)	0.932	
Other <sup>6</sup>	94	(<1)	1.34	(0.50, 3.64)	0.563	25	(-)	0.00		1.000	
<b>Previous mono/duo exposure</b>											0.202
No	14216	(95)	1.00			2167	(86)	1.00			
Yes	497	(5)	1.16	(0.84, 1.61)	0.357	117	(14)	1.72	(0.90, 3.30)	0.102	
<b>Hepatitis B co-infection</b> <sup>7</sup>					0.079					0.339	<b>0.018</b>
Negative	6823	(47)	1.00			1730	(74)	1.00			
Positive	694	(7)	<b>1.37</b>	<b>(1.03, 1.82)</b>	<b>0.029</b>	195	(7)	0.60	(0.29, 1.24)	0.168	
Not tested	7196	(46)	0.96	(0.72, 1.28)	0.774	359	(19)	0.78	(0.36, 1.69)	0.531	
<b>Hepatitis C co-infection</b> <sup>8</sup>					0.652					0.888	0.068
Negative	6028	(40)	1.00			1475	(59)	1.00			
Positive	717	(9)	1.17	(0.80, 1.70)	0.414	310	(18)	0.92	(0.47, 1.80)	0.807	
Not tested	7968	(51)	1.09	(0.78, 1.51)	0.620	499	(23)	1.15	(0.52, 2.56)	0.729	

<sup>1</sup>Wald test for the interaction term between cohort (ie. TAHOD or TAHOD-LITE) and the respective risk factor.<sup>2</sup>Column percentages of the patients who have died for the respective risk factor, by cohort.<sup>3</sup>Wald test for each level and global of the respective risk factor. Global p values for year of ART initiation, age and pre-ART CD4 count are test for trend while all other global p values are test for heterogeneity.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

- <sup>4</sup> Regimen combination consisting of nucleoside reverse transcriptase inhibitors (NRTIs) and nonnucleoside reverse transcriptase inhibitor (NNRTI).
- <sup>5</sup> Regimen combination consisting of nucleoside reverse transcriptase inhibitors (NRTIs) and protease inhibitor (PI).
- <sup>6</sup> Any regimen combination excluding NRTIs+NNRTI or NRTIs+PI.
- <sup>7</sup> Hepatitis B surface antigen result where positive indicates ever positive result.
- <sup>8</sup> Hepatitis C antibody result where positive indicates ever positive result.