Published in final edited form as:

J Clin Epidemiol. 2016 August ; 76: 183–192. doi:10.1016/j.jclinepi.2016.01.030.

The occurrence of Simpson's Paradox if site-level effect was ignored in the TREAT Asia HIV Observational Database (TAHOD)

Correspondence to: Awachana Jiamsakul, The Kirby Institute, UNSW Australia, Sydney NSW 2052, Australia, Ph: +61 2 9385 0900, Fax: +61 2 9385 0940, ajiamsakul@kirby.unsw.edu.au.

Stephen J Kerr: Stephen.k@hivnat.org

Ezhilarasi Chandrasekaran: biostats@yrgcare.org

Aizobelle Huelgas: bellehuelgas@yahoo.com

Sineenart Taecharoenkul: sineenart@rihes-cmu.org

Sirinya Teeraananchai: sirinya.t@hivnat.org

Gang Wan: wkingang@yahoo.com.cn

Penh Sun Ly: penhsun@nchads.org

Sasisopin Kiertiburanakul: sasisopin.kie@mahidol.ac.th

Matthew Law: mlaw@kirby.unsw.edu.au

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.

Author's contributions

AJ conceptualised analysis ideas, performed data collection, data analysis, drafting of manuscript and manuscript submission. ML initiated concept ideas, provided analysis inputs, reviewed and edited the manuscript. SK provided analysis inputs, reviewed and edited the manuscript. EC, AH, STa, STe and GW provided analysis inputs and reviewed the manuscript. PSL and SK reviewed the final manuscript. All authors have approved of the final version of the manuscript.

Conflicts of interest

There are no conflicts of interest.

The TREAT Asia HIV Observational Database

PS Ly* and V Khol, National Center for HIV/AIDS, Dermatology & STDs, Phnom Penh, Cambodia;

FJ Zhang*, HX Zhao and N Han, Beijing Ditan Hospital, Capital Medical University, Beijing, China;

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;

E Yuniastuti*, D Imran and A Widhani, Working Group on AIDS Faculty of Medicine, University of Indonesia/ Cipto

Mangunkusumo Hospital, Jakarta, Indonesia;

S Oka*, J Tanuma and T Nishijima, National Center for Global Health and Medicine, Tokyo, Japan;

JY Choi*, Na S and JM Kim, Division of Infectious Diseases, Department of Internal Medicine, Yonsei University College of Medicine, Seoul, South Korea;

BLH Sim*, YM Gani and R David, Hospital Sungai Buloh, Sungai Buloh, Malaysia;

A Kamarulzaman*, SF Syed Omar, S Ponnampalavanar and I Azwa, University Malaya Medical Centre, Kuala Lumpur, Malaysia;

M Mustafa and N Nordin, Hospital Raja Perempuan Zainab II, Kota Bharu, Malaysia;

R Ditangco*, E Uy and R Bantique, Research Institute for Tropical Medicine, Manila, Philippines;

WW Wong*, WW Ku and PC Wu, Taipei Veterans General Hospital, Taipei, Taiwan;

OT Ng*, PL Lim, LS Lee and R Martinez-Vega, Tan Tock Seng Hospital, Singapore;

P Phanuphak*, K Ruxrungtham, A Avihingsanon and P Chusut, HIV-NAT/Thai Red Cross AIDS Research Centre, Bangkok, Thailand;

S Kiertiburanakul*, S Sungkanuparph, L Chumla and N Sanmeema, Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand;

R Chaiwarith*, T Sirisanthana, W Kotarathitum and J Praparattanapan, Research Institute for Health Sciences, Chiang Mai, Thailand;

P Kantipong* and P Kambua, Chiangrai Prachanukroh Hospital, Chiang Rai, Thailand;

W Ratanasuwana* and R Sriodee, Faculty of Medicine, Siriraj Hospital, Mahidol University, Bangkok, Thailand;

KV Nguyen*, VH Bui, DTH Nguyen and DT Nguyen, National Hospital for Tropical Diseases, Hanoi, Vietnam;

TT Pham*, DD Cuong and HL Ha, Bach Mai Hospital, Hanoi, Vietnam;

AH Sohn*, N Durier* and B Petersen, TREAT Asia, amfAR - The Foundation for AIDS Research, Bangkok, Thailand;

DA Cooper, MG Law*, A Jiamsakul* and DC Boettiger, The Kirby Institute, UNSW Australia, Sydney, Australia.

* TAHOD Steering Committee member; † Steering Committee Chair; ‡ co-Chair

Awachana Jiamsakul¹, Stephen J Kerr^{2,3}, Ezhilarasi Chandrasekaran⁴, Aizobelle Huelgas⁵, Sineenart Taecharoenkul⁶, Sirinya Teeraananchai², Gang Wan⁷, Penh Sun Ly⁸, Sasisopin Kiertiburanakul⁹, Matthew Law¹, and on behalf of the TREAT Asia HIV Observational Database (TAHOD)

¹ The Kirby Institute, UNSW Australia, Sydney, Australia ² HIV-NAT, The Thai Red Cross AIDS Research Centre, Bangkok, Thailand ³ Department of Global Health, Academic Medical Center, University of Amsterdam, Amsterdam Institute for Global Health and Development, Amsterdam, The Netherlands ⁴ YRG CARE Medical Centre, Chennai, India ⁵ Research Institute for Tropical Medicine, Manila, Philippines ⁶ Research Institute for Health Sciences, Chiang Mai, Thailand ⁷ Beijing Ditan Hospital, Capital Medical University, Beijing, China ⁸ National Center for HIV/AIDS, Dermatology & STDs, Phnom Penh, Cambodia ⁹ Faculty of Medicine Ramathibodi Hospital, Mahidol University, Bangkok, Thailand

Abstract

Background—In multi-site HIV observational cohorts, clustering of observations often occur within sites. Ignoring clustering may lead to “Simpson's paradox” (SP) where the trend observed in the aggregated data is reversed when the groups are separated. This study aimed to investigate the SP in an Asian HIV cohort and the effects of site-level adjustment through various Cox-regression models.

Methods—Survival time from combination antiretroviral therapy (cART) initiation was analysed using four Cox models: (i) no site adjustment; (ii) site as a fixed effect; (iii) stratification through site; and (iv) shared frailty on site.

Results—A total of 6454 patients were included from 23 sites in Asia. SP was evident in the year of cART initiation variable. Model (i) shows the hazard ratio (HR) for years 2010-2014 was higher than the HR for 2006-2009, compared to 2003-2005 (HR = 0.68 vs 0.61). Models (ii)-(iv) consistently implied greater improvement in survival for those who initiated in 2010-2014 than 2006-2009 contrasting findings from Model (i). The effects of other significant covariates on survival were similar across four models.

Conclusions—Ignoring site can lead to SP causing reversal of treatment effects. Greater emphasis should be made to include site in survival models when possible.

Keywords

Simpson's paradox; clustering; HIV; cohort; Yule-Simpson; Cox

Introduction

In HIV research, time to event analyses are often analysed using the conventional Cox Proportional Hazards regression model [1, 2]. In multi-site cohort analyses, there is often heterogeneity across sites, resulting in potential clustering of observations within sites that must be accounted for when developing statistical models. For example, patients attending the same hospital are often correlated due to exposure to common treatment protocols. There may be differences in treatment guidelines across countries due to drug availability, and in

addition sites in resource-limited settings may not offer routine viral load testing which has been shown to be associated with treatment outcomes[3]. Patients from developed areas normally initiate combination antiretroviral therapy (cART) on a protease inhibitor (PI) based regimen while the standard first-line cART in resource-limited countries are nonnucleoside reverse transcriptase inhibitors (NNRTI) based[4]. Moreover, patient selection procedures for enrolment into a cohort study could be dependent upon the principal investigator of that site. Therefore when conducting multi-site analyses, site-level effect should be taken into account appropriately to ensure correct inferences are drawn.

It is common to find that the significant association with treatment outcomes diminishes once site has been accounted for in regression analyses. However, it cannot be assumed that the direction of the effect size, for example the hazard ratios (HR) in survival analysis, would remain the same in models with and without site adjustment. Adjusting for site clustering can lead to a situation where statistical significance with the outcome variable is retained, but with reversed association. This is called the “Simpson's paradox” (Yule-Simpson effect)[5], in which the trend observed in the aggregated data is reversed when the data is separated. This is generally due to large differences in baseline rates across sites[6], which is seen in our cohort, and leads to the underlying clustering of patients within and between sites.

The TREAT Asia HIV Observational Database (TAHOD)[7] is a multi-site adult HIV observational cohort with data from 13 countries across Asia. Patients enrolled in TAHOD receive clinical care according to treatment guidelines and protocols relevant to their treatment centre. As TAHOD sites are located across high, middle and low income countries, with multiple sites from the same country situated in different demographic and geographic regions, it can be assumed that there is potential clustering of patients within each site.

The aim of this study was to demonstrate the occurrence of Simpson's paradox in TAHOD when site-level clustering is ignored in survival analysis and the impact this has on interpretation of mortality estimates. To ensure that our inferences are not due to a particular behaviour of the different methods used to control for site, we repeated the procedure with three types of Cox regression models where site was adjusted as a fixed effect; through stratification; and shared frailty random effect methods.

Methods

Patients enrolled in TAHOD at the March 2014 data transfer, who had started cART from 2003 onwards without prior mono/dual therapy were included in this analysis. Risk for mortality began at cART initiation, with the exception of those who enrolled in TAHOD already receiving treatment, in which case they were left truncated at cohort entry. Survival time ended on the date of death, or on the date of last follow-up for subjects who were censored. Covariates adjusted in all models were year of cART initiation grouped into chronological periods of 2003-2005, 2006-2009 and 2010-2014, broadly corresponding to changes in HIV treatment guidelines[8, 9], age, sex, HIV mode of exposure, pre-cART viral load (VL) and CD4 count, initial cART regimen, hepatitis B/C co-infection, and previous

AIDS diagnoses. Covariates with missing values were categorised as “missing” but not included in the Wald's tests for heterogeneity. All covariates were adjusted as time fixed covariates and entered into the multivariate model concurrently. No variable selection was attempted. Covariates with $p < 0.05$ in the multivariate model were considered statistically significant.

We developed four types of Cox regression models to determine the effects of site-level correlation:

Model (i): No site adjustment

In this model, all covariates were fitted using Cox regression, with the exception of site. This model provides the effects of covariates on survival time assuming that site had no impact on survival and there was no within site correlation structure. All observations within and across sites were considered to be independent.

Model (ii): Site adjusted as a fixed effect

Site was included in the model as a fixed covariate, with differences between sites modelled as constant estimated parameters in the same way as other covariates. Adding site as a fixed covariate increases the degrees of freedom. However, the estimated HRs can be obtained for each site. Site effect was not reported due to confidentiality reasons.

Model (iii): Site adjusted through stratification

A stratified Cox regression analysis was used to account for clusters within sites. In this model, the baseline hazard for each site was allowed to vary, resulting in different Kaplan-Meier curves for each site. The stratified model, however, produces the same set of regression coefficients or hazard ratios for all strata. One advantage of this method is that it can be used to remedy violation of proportional hazards (PH) assumption and is easily interpreted. However, it is based on intracluster comparisons, and the effect of site as a stratification variable is considered a “nuisance” parameter and not estimated in the stratified Cox model.

Model (iv): Site adjusted in the shared frailty model

The Cox shared frailty model accounts for dependence of observations by assuming that subjects within the same cluster share the same frailty. For example, subjects within the same site have survival times dependent on the survival of other subjects, due to unobservable factors in that site. This is similar to a random effects model where the variance parameter is a measure of the degree of correlation within a cluster. In this analysis, gamma-distributed shared frailty Cox model was used causing a violation of the PH assumption at the population level. Therefore the HR can only be interpreted based on subjects sharing the same frailty[10].

Kaplan-Meier (K-M) curves were plotted for survival time with and without site adjustment. A maximum of five year survival time was displayed on these K-M curves.

Ethics approvals were obtained from UNSW Australia Ethics Committee, Western Institutional Review Board, and respective local ethics committee from each TAHOD participating site and coordinating centre (TREAT Asia/amfAR, Bangkok, Thailand). All data management and statistical analyses were performed using SAS software version 9.3 (SAS Institute Inc., Cary, NC, USA) and Stata software version 13.1 (StataCorp, College Station, TX, USA).

Results

A total of 6454 patients from 23 sites in 13 countries across Asia were included. Table 1 describes the characteristics of these patients. There were a total of 271 deaths (4.2%) with a mortality rate of 1.06 per 100 person-years (100PYS), and 95% confidence interval (95%CI) of (0.94-1.19) as shown in Table 2. The crude mortality rates for year of cART initiation shows that those starting treatment in 2003-2005 had a rate of 1.06 per 100PYS, which decreased to 0.89 per 100PYS in 2006-2009, and increased back to 1.50 per 100PYS for 2010-2014. These rates imply that without accounting for any confounders, patients initiating cART in 2010-2014 had poorer survival than those initiating in earlier years, while those who started treatment between 2006-2009 had the lowest mortality rate.

In Table 3, univariate and multivariate results from Models (i) to (iv) are presented. In Model (i), where site was assumed to not have an effect on survival, the overall multivariate results found that older age, being hepatitis C positive and having had prior AIDS illnesses were associated with poorer survival. Later years of cART initiation and higher CD4 count helped reduced the risk of mortality. Examining the adjusted HRs for the year of cART initiation variable reveals that compared to 2003-2005, mortality was lower if patients initiated cART in 2006-2009 (HR = 0.61, 95%CI (0.45-0.84)) and 2010-2014 (HR = 0.68, 95%CI (0.47-0.98)). However, when comparing the HRs for the two time periods: 2006-2009 (HR = 0.61) and 2010-2014 (HR = 0.68), the HR for 2010-2014 was unexpectedly higher than the HR for 2006-2009. This is contradictory to the improved survival time that one was expected to see for each incremental time period of cART initiation. Figure 1a illustrates this phenomenon through the K-M plot for survival time by year of cART initiation. From this plot, we see that those who initiated in 2006-2009 had better survival than patients from other time groups.

We then explored Model (ii) which included site as a fixed covariate. The effects of the significant covariates in the adjusted model remained similar to Model (i) with the exception of the year of cART initiation. When adjusting for site as a fixed effect, year of cART initiation maintained its statistical significance in the presence of other confounders, but the relationship of the HRs between different periods has now changed. The 2010-2014 time period now has a smaller HR (HR = 0.45, 95%CI (0.28-0.74)) relative to 2006-2009 (HR = 0.58, 95%CI (0.41-0.81)) in comparison with 2003-2005 as the reference time period. We attribute this reversed association with the introduction of site variable into the model as all other covariates have remained the same. Figure 1b shows the K-M plot adjusted for site. Here we see that survival time was better for patients in the 2010-2014 group.

Next, we performed a stratified Cox regression (Model (iii)) treating site as a stratification variable. In this model, the HRs and the 95%CI for each of the significant covariates in the multivariate model were very similar to Model (ii). Initiating cART between 2010-2014 had the lowest hazard for mortality (HR = 0.42, 95%CI (0.25-0.70)), while the HR for 2006-2009 was a little higher (HR = 0.57, 95%CI (0.40-0.63)), compared to the reference time period.

Lastly, the Cox shared frailty model (Model (iv)) was applied by treating patients from the same site as having the same frailty. The HRs for the year of cART covariate increased slightly compared to Models (ii)-(iii), however the direction of the effects and the significance of the variable remained comparable to the previous two models (2006-2009 HR = 0.60, 95%CI (0.43-0.83); and 2010-2014 HR = 0.55, 95%CI (0.36-0.84)).

In terms of other covariates adjusted in the models, the effect sizes of the significant covariates in all four models were similar to each other. The overall interpretation of these covariates are that irrespective of whether site clustering was taken into account, patients initiating cART at age older than 50 years had more than triple the hazard for mortality compared to those aged ≤ 30 years. Those with hepatitis C co-infection and prior AIDS diagnoses had HRs that were approximately doubled those of subjects without the co-infection or AIDS diagnoses. Higher pre-treatment CD4 counts were associated with better survival time, where CD4 >200 cells/ μ L reduced the hazard for mortality by roughly 40-50% and CD4 count between 101-200 cells/ μ L reduced the hazard by approximately 30%, compared to those with CD4 ≤ 50 cells/ μ L.

The significance of clustering within sites, however, is illustrated through the year of cART initiation variable. Once site-level correlation has been adjusted in Models (ii)-(iv), the HRs for the different time periods in this variable showed a reversed relationship compared to Model (i). This can be seen in both the univariate and multivariate results. Model (i) shows that improvement in survival time was better if initiating cART in 2006-2009, compared to 2010-2014. Models (ii)-(iv) consistently indicate that improvement in survival was higher in 2010-2014. The reversion of the HRs in Models (ii)-(iv) points to the underlying existence of Simpson's paradox suggesting that site variable in TAHOD plays a crucial role in determining the effects of predictive factors for survival outcomes.

Discussion

TAHOD, a multi-site HIV observational cohort, captures data arising from routine clinical care from sites across Asia. As TAHOD participating sites were selected from major HIV referral centres, with some being the only site representing their country, it was expected that there would be variability in treatment protocols and clinical care across these sites. For example, the World Health Organization (WHO) 2013 guidelines recommend VL testing as the preferred method to monitor ART failure due to poor correlation of other monitoring methods with VL failure[4, 11, 12]. However, in resource-limited settings, VL monitoring may not form part of routine clinical care due to its high cost, or lack of resources and infrastructure. In such cases, immunologic and/or clinical failure is often used to define treatment failure. Studies have shown that by using methods other than VL to assess ART

response, there may be a delay in switching to second-line ART which has been shown to be associated with mortality. Furthermore, patients with immunological failure were shown to have higher VL and higher rates of HIV drug resistance mutations compared to those with virological failure, which could impact second-line treatment options[13, 14]. As TAHOD sites are located in high and low-income countries[15], the frequency of VL testing in TAHOD sites varies substantially from an average of less than once per year to more than three times per year[3]. It was therefore not unreasonable to assume that clinical outcomes within each site would be highly correlated.

Many studies have used different strategies to account for site differences in time to event analyses, by way of adjusting or stratification by site or cohort [16, 17]. However, we continue to see examples of studies where the site-level effect was ignored[18, 19]. In this situation, there may be underlying reasons for the lack of adjustment for site/cohort such as small sample sizes[20]. In HIV research, it is known that there is usually an improvement in survival for patients initiating cART in later years[21, 22]. Our current TAHOD analyses illustrate that not adjusting for site differences can lead to a complete misinterpretation of the hazard for mortality, rather than simply producing biased estimates for the HRs causing changes in the 95% CI and p-values.

The Simpson's paradox effect, in which the aggregated result is the opposite in direction to the adjusted one, is reported infrequently in medical research, however the occurrence may be much more common than conventionally thought[6, 23, 24]. One example of Simpson's paradox described in previous literature is the babies' birth weight example where babies who were born with low birth weight (LBW) were known to have higher risk of mortality compared to those with higher birth weight. However, babies with LBW from high-risk population (e.g. from mothers who smoke) tend to have lower mortality than LBW babies from low-risk population (e.g. mothers who were non-smokers)[25]. Another example is the kidney stone study where the overall success rates of kidney stone removal between the 1970's to 1980's from percutaneous nephrolithotomy was reported to be 83% compared to 78% from open surgery. When the stone diameter was taken into account, open surgery was shown to have higher success rates over percutaneous nephrolithotomy for both diameter groups[26]. A more recent scientific study has found that lipids common to plasma membranes of three life domains (archaea, eubacteria, and eukaryotes) showed a similar number of carbon atoms in eubacteria compared to eukaryotes. However, when analysing mutually exclusive subsets of the same data, the number of atoms found in eukaryotes was higher than eubacteria[27]. A study looking at gender differences in health care utilisation found that not adjusting for nursing home residence led to different conclusions than if a nursing home indicator was adjusted for[28]. There have been several investigations examining the nature of the Simpson's paradox including the relationship with causal inferences[29, 30]. Evaluations of these causality interpretations have also been conducted by other investigators [31, 32]. Recommendations to reduce or avoid Simpson's paradox have been suggested, including a detailed checklist for each research process stage[33] and the use of cross-sum ratio rather than the cross-product ratio (odds ratio) based on a 2×2 table[34].

We found that the three methods used to adjust for site in our study (fixed effect, stratification and shared frailty) provided similar HRs and statistical significance for the adjusted covariates. Since the conclusions drawn from these three models in relations to the effects of other significant covariates were consistent across models, determining which type of Cox regression to fit was less vital than the acknowledgement that one must not ignore site-level correlation when performing these analyses. The findings are consistent with a previous study exploring seven Cox regression models to account for heterogeneity across HIV cohorts. The study concluded that although adjusting for site differences was crucial in these analyses involving multiple and disparate sites, the specific technique used to account for heterogeneity played a less important role[35].

One limitation of this study was that we assumed that all covariates satisfied the PH assumption. This study was intended to demonstrate the significance of including site in the Cox model to account for dependence of observations within the same site. If a covariate violates the PH assumption, one option would be to stratify or add a time-varying covariate into the model. If multiple covariates violate the assumption, the complexity of the model increases which could obscure the underlying objectives for this study, that heterogeneity across sites should not be ignored. However, we have tested for proportionality of hazards for our main covariate of interest (Year of cART initiation) which showed no evidence of assumption violation. For all other covariates, we assumed that the PH assumption was satisfied. Another limitation was the fact that we only included baseline or time-fixed covariates in the models. CD4 cell count and viral load can be fitted in a Cox regression as time-varying covariates to allow a person to contribute risk time in more than one category of a covariate. This would produce HRs that took into account all CD4 and viral load levels after cART initiation rather than pre-cART levels only. Since the focus of this study was to illustrate the occurrence of Simpson's paradox as a consequence of not adjusting for site, we simplified our analyses by only utilising time-fixed covariates.

Conclusions

Our study demonstrates the importance of adjusting for site differences in time to event analyses. Ignoring an effect of site can lead to Simpson's paradox which causes the reversion of the association between explanatory variables leading to misinterpretation of treatment effects, rather than simply changing the statistical significance. Methods used to adjust for site did not cause considerable changes to the multivariate results suggesting that it is important to consider including site in regression analyses but that the technique used to fit site in the regression did not matter greatly in this situation. Researchers should be aware that when presented with counterintuitive statistical results, they maybe interacting with the Simpson's paradox. It is crucial that researchers have a thorough understanding of their dataset and the ability to recognise reversed statistical findings which could have a major scientific impact on the wider community.

Acknowledgments

The TREAT Asia HIV Observational Database 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. Cornell M, Lessells R, Fox MP, Garone DB, Giddy J, Fenner L, et al. Mortality among adults transferred and lost to follow-up from antiretroviral therapy programmes in South Africa: a multicenter cohort study. *J Acquir Immune Defic Syndr*. 2014; 67:e67–75. [PubMed: 24977471]
2. Ebissa G, Deyessa N, Biadgilign S. Predictors of early mortality in a cohort of HIV-infected children receiving high active antiretroviral treatment in public hospitals in Ethiopia. *AIDS care*. 2015:1–8.
3. Oyomopito R, Lee MP, Phanuphak P, Lim PL, Ditangco R, Zhou J, et al. Measures of site resourcing predict virologic suppression, immunologic response and HIV disease progression following highly active antiretroviral therapy (HAART) in the TREAT Asia HIV Observational Database (TAHOD). *HIV Med*. 2010; 11:519–29. [PubMed: 20345881]
4. World Health Organization. Consolidated guidelines on the use of antiretroviral drugs for treatment and preventing HIV infection: Recommendations for a public health approach June. 2013
5. Simpson EH. The Interpretation of Interaction in Contingency Tables. *Journal of the Royal Statistical Society*. 1951; 13:238–41.
6. Kievit RA, Frankenhuis WE, Waldorp LJ, Borsboom D. Simpson's paradox in psychological science: a practical guide. *Frontiers in psychology*. 2013;4. [PubMed: 23378839]
7. 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*. 2005; 38:174–9. [PubMed: 15671802]
8. World Health Organization. Antiretroviral therapy for HIV infection in adults and adolescents: recommendations for a public health approach. 2006 revision.
9. World Health Organization. Antiretroviral therapy for HIV infection in adults and adolescents: recommendations for a public health approach. 2010 revision.
10. Kleinbaum, DG.; Klein, M. Survival Analysis. A Self-Learning Text. Second ed. Springer Science +Business Media, Inc.; 2005.
11. Rutherford GW, Anglemyer A, Easterbrook PJ, Horvath T, Vitoria M, Penazzato M, et al. Predicting treatment failure in adults and children on antiretroviral therapy: a systematic review of the performance characteristics of the 2010 WHO immunologic and clinical criteria for virologic failure. *AIDS*. 2014; 28(Suppl 2):S161–9. [PubMed: 24849476]
12. Ingole N, Mehta P, Pazare A, Paranjpe S, Sarkate P. Performance of immunological response in predicting virological failure. *AIDS Res Hum Retroviruses*. 2013; 29:541–6. [PubMed: 23137294]
13. Petersen ML, Tran L, Geng EH, Reynolds SJ, Kambugu A, Wood R, et al. Delayed switch of antiretroviral therapy after virologic failure associated with elevated mortality among HIV-infected adults in Africa. *AIDS*. 2014; 28:2097–107. [PubMed: 24977440]
14. Sungkanuparph S, Win MM, Kiertiburanakul S, Phonrat B, Maek-a-nantawat W. HIV-1 drug resistance at virological failure versus immunological failure among patients failing first-line antiretroviral therapy in a resource-limited setting. *International journal of STD & AIDS*. 2012; 23:316–8. [PubMed: 22648883]
15. The World Bank.
16. Engsig FN, Zangerle R, Katsarou O, Dabis F, Reiss P, Gill J, et al. Long-term mortality in HIV-positive individuals virally suppressed for >3 years with incomplete CD4 recovery. *Clin Infect Dis*. 2014; 58:1312–21. [PubMed: 24457342]
17. Phanuphak P, Sirivichayakul S, Jiamsakul A, Sungkanuparph S, Kumarasamy N, Lee MP, et al. Transmitted Drug Resistance and Antiretroviral Treatment Outcomes in Non-Subtype B HIV-1-Infected Patients in South East Asia. *J Acquir Immune Defic Syndr*. 2014; 66:74–9. [PubMed: 24413039]

18. Hasse B, Ledergerber B, Furrer H, Battegay M, Hirschel B, Cavassini M, et al. Morbidity and aging in HIV-infected persons: the Swiss HIV cohort study. *Clin Infect Dis*. 2011; 53:1130–9. [PubMed: 21998280]
19. Pacheco YM, Jarrin I, Rosado I, Campins AA, Berenguer J, Iribarren JA, et al. Increased risk of non-AIDS-related events in HIV subjects with persistent low CD4 counts despite cART in the CoRIS cohort. *Antiviral Res*. 2015
20. Jiamsakul A, Sungkanuparph S, Law M, Kantor R, Praparattanapan J, Li PC, et al. HIV multi-drug resistance at first-line antiretroviral failure and subsequent virological response in Asia. *J Int AIDS Soc*. 2014; 17:19053. [PubMed: 25141905]
21. Grimsrud A, Balkan S, Casas EC, Lujan J, Van Cutsem G, Poulet E, et al. Outcomes of antiretroviral therapy over a 10-year period of expansion: a multicohort analysis of African and Asian HIV programs. *J Acquir Immune Defic Syndr*. 2014; 67:e55–66. [PubMed: 24977472]
22. Boule A, Schomaker M, May MT, Hogg RS, Shepherd BE, Monge S, et al. Mortality in patients with HIV-1 infection starting antiretroviral therapy in South Africa, Europe, or North America: a collaborative analysis of prospective studies. *PLoS Med*. 2014; 11:e1001718. [PubMed: 25203931]
23. Terwilliger, J.; Schield, M. Frequency of Simpson's Paradox in NAEP Data. The American Education Research Association (AERA). San Diego: 2004.
24. Pavlides MG, Perlman MD. How Likely Is Simpson's Paradox? *Am Stat*. 2009; 63:226–33.
25. Wilcox AJ. On the importance--and the unimportance--of birthweight. *Int J Epidemiol*. 2001; 30:1233–41. [PubMed: 11821313]
26. Julious SA, Mullee MA. Confounding and Simpson's paradox. *BMJ*. 1994; 309:1480–1. [PubMed: 7804052]
27. Bansal S, Mittal A. A statistical anomaly indicates symbiotic origins of eukaryotic membranes. *Molecular biology of the cell*. 2015
28. Kronman AC, Freund KM, Hanchate A, Emanuel EJ, Ash AS. Nursing home residence confounds gender differences in Medicare utilization an example of Simpson's paradox. *Women's health issues : official publication of the Jacobs Institute of Women's Health*. 2010; 20:105–13.
29. Pearl, J. Causality. Second ed. Cambridge University Press; 2009.
30. Spirtes, P.; Glymour, C.; Scheines, R. Causation, Prediction, and Search. Second ed. MIT Press; 2000.
31. Bandyopadhyay PS, Nelson D, Greenwood M, Brittan G, Berwald J. The logic of Simpson's paradox. *Synthese*. 2010; 181:185–208.
32. Bandyopadhyay PS, Greenwood M, Dcruz DWF, Venkata RR. Simpson's Paradox and Causality. *Am Philos Quart*. 2015; 52:13–25.
33. Smith ML, Goltz HH. What is hidden in my data? Practical strategies to reveal Yule-Simpson's paradox and strengthen research quality in health education research. *Health promotion practice*. 2012; 13:637–41. [PubMed: 22923774]
34. Rudas T. Informative allocation and consistent treatment selection. *Statistical Methodology*. 2010; 7:323–37.
35. Giganti MJ, Luz PM, Caro-Vega Y, Cesar C, Padgett D, Koenig S, et al. A comparison of seven Cox regression-based models to account for heterogeneity across multiple HIV treatment cohorts in Latin America and the Caribbean. *AIDS Res Hum Retroviruses*. 2015

What's new?

Key findings

- Simpson's paradox can occur if no adjustment has been made for site-level clustering causing the interpretations of hazard ratios to be counterintuitive. We found that without accounting for site, those who initiated antiretroviral therapy in 2010-2014 had higher hazard for mortality compared to the hazard for 2006-2009, with 2003-2005 as the reference group.
- Once site is adjusted through fixed effect, stratification or shared frailty methods, the hazard ratios were now reversed showing that patients who started treatment in later years had better survival. This indicates that greater emphasis should be made to include site in survival models and that the methods used for site adjustment played a less vital role.

What this adds to what is known?

- Previous studies have shown the occurrence of Simpson's paradox in various research fields. However, to our knowledge, this study was the first in HIV research to specifically show the reversal of hazard ratios as a consequence of ignoring site-level clustering.

What is the implication, what should change now?

- Ignoring site effect in multi-site cohort analyses can cause the reversion of the association between explanatory variables leading to misinterpretation of treatment effects, rather than simply causing biased estimates of the hazard ratios and changing the statistical significance.
- Researchers should be aware that when presented with counterintuitive statistical results, there may be a possible violation of the basic statistical assumptions, such as the independence of observations. Greater awareness would allow investigators to take appropriate actions in order to avoid reaching paradoxical conclusions.

Figure 1.

(a) Kaplan-Meier plot without site adjustment; and (b) Kaplan-Meier plot adjusted for site

Table 1

Patient Characteristics

	Total (%) 6454 (100)	No of deaths (%) 271 (4.2)
Year of cART initiation		
2003-2005	1627 (25.3)	107 (39.5)
2006-2009	2926 (45.3)	101 (37.3)
2010-2014	1901 (29.5)	63 (23.2)
Age at cART initiation (years)	median = 35, IQR (29-41)	median = 37, IQR (31-47)
30	1969 (30.5)	59 (21.8)
31-40	2736 (42.4)	105 (38.7)
41-50	1231 (19.1)	55 (20.3)
>50	518 (8.0)	52 (19.2)
Sex		
Male	4481 (69.4)	211 (77.9)
Female	1973 (30.6)	60 (22.1)
Mode of HIV Exposure		
Heterosexual contact	4043 (62.6)	173 (63.8)
Homosexual contact	1331 (20.6)	36 (13.3)
Injecting drug use	607 (9.4)	40 (14.8)
Other/unknown	473 (7.3)	22 (8.1)
Pre-cART viral load (copies/mL)	median = 100000, IQR (31000-280000)	median = 150000, IQR (67568-500000)
<100000	1628 (25.2)	49 (18.1)
100000	1677 (26.0)	82 (30.3)
Missing	3149 (48.8)	140 (51.7)
Pre-cART CD4 (cells/μL)	median = 124, IQR (38.5-220)	median = 56, IQR (18-161)
50	1690 (26.2)	114 (42.1)
51-100	787 (12.2)	37 (13.7)
101-200	1406 (21.8)	47 (17.3)

	Total (%) 6454 (100)	No of deaths (%) 271 (4.2)
>200	1665 (25.8)	37 (13.7)
Missing	906 (14.0)	36 (13.3)
Initial cART Regimen		
NRTI+NNRTI	5642 (87.4)	244 (90.0)
NRTI+PI	738 (11.4)	23 (8.5)
Other	74 (1.1)	4 (1.5)
Hepatitis B co-infection		
Negative	4583 (71.0)	168 (62.0)
Positive	523 (8.1)	30 (11.1)
Not tested	1348 (20.9)	73 (26.9)
Hepatitis C co-infection		
Negative	3996 (61.9)	146 (53.9)
Positive	781 (12.1)	47 (17.3)
Not tested	1677 (26.0)	78 (28.8)
Previous AIDS		
No	4000 (62.0)	100 (36.9)
Yes	2454 (38.0)	171 (63.1)

Crude mortality rates

Table 2

	No patients	Follow up (years)	No of deaths	Mortality rate (/100pys)	95% CI
Total	6454	25621.84	271	1.06	(0.94, 1.19)
Year of cART initiation					
2003-2005	1627	10090.4	107	1.06	(0.88, 1.28)
2006-2009	2926	11329.53	101	0.89	(0.73, 1.08)
2010-2014	1901	4201.91	63	1.50	(1.17, 1.92)
Age at cART initiation (years)					
30	1969	7261.48	59	0.81	(0.63, 1.05)
31-40	2736	11230.93	105	0.93	(0.77, 1.13)
41-50	1231	5146.02	55	1.07	(0.82, 1.39)
>50	518	1983.41	52	2.62	(2.00, 3.44)
Sex					
Male	4481	17578.61	211	1.20	(1.05, 1.37)
Female	1973	8043.23	60	0.75	(0.58, 0.96)
Mode of HIV Exposure					
Heterosexual contact	4043	16545	173	1.05	(0.90, 1.21)
Homosexual contact	1331	5315.18	36	0.68	(0.49, 0.94)
Injecting drug use	607	1730.76	40	2.31	(1.70, 3.15)
Other/Unknown	473	2030.9	22	1.08	(0.71, 1.65)
Pre-cART viral load (copies/mL)					
<100000	1628	6296.28	49	0.78	(0.59, 1.03)
100000	1677	6633.63	82	1.24	(1.00, 1.53)
Missing	3149	12691.93	140	1.10	(0.93, 1.30)
Pre-cART CD4 (cells/μL)					
50	1690	6577.23	114	1.73	(1.44, 2.08)
51-100	787	3060.11	37	1.21	(0.88, 1.67)

	No patients	Follow up (years)	No of deaths	Mortality rate (/100pys)	95% CI
101-200	1406	5736.14	47	0.82	(0.62, 1.09)
>200	1665	6263.67	37	0.59	(0.43, 0.82)
Missing	906	3984.7	36	0.90	(0.65, 1.25)
Initial cART Regimen					
NRTI+NNRTI	5642	21839.65	244	1.12	(0.99, 1.27)
NRTI+PI	738	3508.87	23	0.66	(0.44, 0.99)
Other combination	74	273.32	4	1.46	(0.55, 3.90)
Hepatitis B co-infection					
Negative	4583	18382	168	0.91	(0.79, 1.06)
Positive	523	2061.01	30	1.46	(1.02, 2.08)
Not tested	1348	5178.83	73	1.41	(1.12, 1.77)
Hepatitis C co-infection					
Negative	3996	16560.66	146	0.88	(0.75, 1.04)
Positive	781	2466.39	47	1.91	(1.43, 2.54)
Not tested	1677	6594.8	78	1.18	(0.95, 1.48)
Previous AIDS					
No	4000	15764.43	100	0.63	(0.52, 0.77)
Yes	2454	9857.41	171	1.73	(1.49, 2.02)

Four Cox regression models for survival time

Table 3

(i) No site adjustment				(ii) Site adjusted as a fixed effect				(iii) Site adjusted through stratification				(iv) Site adjusted in the shared frailty model				
Univariate		Multivariate (df=22)		Univariate		Multivariate (df=44)		Univariate		Multivariate (df=22)		Univariate		Multivariate (df=22)		
HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	
Year of cART initiation																
2003-2005	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
2006-2009	0.54	(0.40, 0.72)	0.61	(0.45, 0.84)	0.43	(0.31, 0.60)	0.58	(0.41, 0.81)	0.42	(0.30, 0.59)	0.57	(0.40, 0.81)	0.46	(0.33, 0.63)	0.60	(0.43, 0.83)
2010-2014	0.64	(0.45, 0.89)	0.68	(0.47, 0.98)	0.30	(0.19, 0.48)	0.45	(0.28, 0.74)	0.28	(0.17, 0.45)	0.42	(0.25, 0.70)	0.39	(0.26, 0.59)	0.55	(0.36, 0.84)
Age at cART initiation (years)																
30	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
31-40	1.18	(0.86, 1.63)	1.13	(0.82, 1.56)	1.21	(0.87, 1.67)	1.08	(0.78, 1.50)	1.20	(0.87, 1.66)	1.08	(0.78, 1.49)	1.21	(0.88, 1.67)	1.10	(0.80, 1.52)
41-50	1.31	(0.91, 1.89)	1.38	(0.94, 2.00)	1.37	(0.94, 2.00)	1.29	(0.88, 1.89)	1.35	(0.93, 1.98)	1.27	(0.87, 1.87)	1.36	(0.93, 1.98)	1.31	(0.89, 1.91)
>50	3.09	(2.13, 4.49)	3.64	(2.47, 5.38)	3.52	(2.40, 5.18)	3.52	(2.35, 5.26)	3.51	(2.38, 5.17)	3.52	(2.35, 5.28)	3.42	(2.33, 5.02)	3.52	(2.36, 5.25)
Sex																
Male	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
Female	0.64	(0.48, 0.85)	0.77	(0.57, 1.05)	0.61	(0.45, 0.82)	0.74	(0.54, 1.02)	0.61	(0.45, 0.82)	0.74	(0.54, 1.01)	0.62	(0.46, 0.83)	0.77	(0.56, 1.05)
Mode of HIV Exposure																
Heterosexual contact	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
Homosexual contact	0.60	(0.42, 0.87)	0.78	(0.52, 1.16)	0.50	(0.32, 0.78)	0.74	(0.47, 1.16)	0.49	(0.32, 0.77)	0.73	(0.46, 1.14)	0.52	(0.34, 0.78)	0.70	(0.45, 1.09)
Injecting drug use	2.01	(1.42, 2.84)	1.44	(0.92, 2.26)	1.66	(1.10, 2.51)	1.09	(0.67, 1.78)	1.60	(1.05, 2.42)	1.04	(0.64, 1.70)	1.84	(1.24, 2.74)	1.25	(0.77, 2.01)
Other/Unknown	1.03	(0.66, 1.61)	1.05	(0.67, 1.66)	1.03	(0.63, 1.67)	1.16	(0.71, 1.90)	1.04	(0.64, 1.69)	1.14	(0.70, 1.87)	0.99	(0.61, 1.60)	1.06	(0.65, 1.72)
Pre-cART viral load (copies/mL)																
<100000	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	
100000	1.64	(1.15, 2.33)	1.14	(0.79, 1.63)	1.42	(0.99, 2.03)	1.08	(0.75, 1.55)	1.41	(0.98, 2.01)	1.06	(0.73, 1.53)	1.47	(1.03, 2.10)	1.09	(0.76, 1.57)
Missing																
Pre-cART CD4 (cells/ μ L)																

(i) No site adjustment				(ii) Site adjusted as a fixed effect				(iii) Site adjusted through stratification				(iv) Site adjusted in the shared frailty model			
Univariate		Multivariate (df=22)		Univariate		Multivariate (df=44)		Univariate		Multivariate (df=22)		Univariate		Multivariate (df=22)	
HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI	HR	95% CI
50	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
51-100	0.68 (0.47, 0.99)	0.77 (0.53, 1.12)	0.72 (0.49, 1.04)	0.82 (0.56, 1.20)	0.73 (0.50, 1.06)	0.83 (0.57, 1.21)	0.80 (0.55, 1.16)	0.83 (0.57, 1.21)	0.70 (0.48, 1.02)	0.80 (0.55, 1.16)	0.80 (0.55, 1.16)	0.80 (0.55, 1.16)	0.80 (0.55, 1.16)	0.80 (0.55, 1.16)	0.80 (0.55, 1.16)
101-200	0.47 (0.33, 0.66)	0.67 (0.47, 0.95)	0.50 (0.36, 0.71)	0.68 (0.47, 0.98)	0.51 (0.36, 0.72)	0.68 (0.48, 0.98)	0.67 (0.46, 0.95)	0.68 (0.48, 0.98)	0.49 (0.34, 0.69)	0.67 (0.46, 0.95)	0.67 (0.46, 0.95)	0.67 (0.46, 0.95)	0.67 (0.46, 0.95)	0.67 (0.46, 0.95)	0.67 (0.46, 0.95)
>200	0.30 (0.21, 0.44)	0.51 (0.34, 0.77)	0.35 (0.24, 0.51)	0.59 (0.39, 0.89)	0.34 (0.23, 0.51)	0.59 (0.39, 0.89)	0.55 (0.37, 0.84)	0.59 (0.39, 0.89)	0.33 (0.22, 0.48)	0.55 (0.37, 0.84)	0.55 (0.37, 0.84)	0.55 (0.37, 0.84)	0.55 (0.37, 0.84)	0.55 (0.37, 0.84)	0.55 (0.37, 0.84)
Missing															
Initial cART Regimen															
NRTI+NNRTI	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
NRTI+PI	0.68 (0.44, 1.04)	0.85 (0.53, 1.35)	1.10 (0.63, 1.94)	1.27 (0.72, 2.23)	1.22 (0.69, 2.15)	1.36 (0.77, 2.43)	1.05 (0.62, 1.78)	1.36 (0.77, 2.43)	0.90 (0.54, 1.50)	1.05 (0.62, 1.78)	1.05 (0.62, 1.78)	1.05 (0.62, 1.78)	1.05 (0.62, 1.78)	1.05 (0.62, 1.78)	1.05 (0.62, 1.78)
Other combination	1.20 (0.45, 3.22)	1.51 (0.54, 4.19)	1.94 (0.68, 5.52)	1.80 (0.60, 5.38)	2.04 (0.71, 5.86)	1.91 (0.63, 5.83)	1.64 (0.56, 4.78)	1.91 (0.63, 5.83)	1.65 (0.59, 4.62)	1.64 (0.56, 4.78)	1.64 (0.56, 4.78)	1.64 (0.56, 4.78)	1.64 (0.56, 4.78)	1.64 (0.56, 4.78)	1.64 (0.56, 4.78)
Hepatitis B co-infection															
Negative	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Positive	1.57 (1.06, 2.31)	1.44 (0.98, 2.14)	1.52 (1.03, 2.25)	1.46 (0.99, 2.17)	1.52 (1.03, 2.25)	1.46 (0.98, 2.16)	1.45 (0.98, 2.15)	1.46 (0.98, 2.16)	1.52 (1.03, 2.26)	1.45 (0.98, 2.15)	1.45 (0.98, 2.15)	1.45 (0.98, 2.15)	1.45 (0.98, 2.15)	1.45 (0.98, 2.15)	1.45 (0.98, 2.15)
Not tested															
Hepatitis C co-infection															
Negative	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Positive	2.12 (1.53, 2.95)	1.65 (1.08, 2.51)	2.05 (1.42, 2.94)	1.78 (1.17, 2.72)	2.01 (1.39, 2.89)	1.78 (1.16, 2.71)	1.76 (1.15, 2.67)	1.78 (1.16, 2.71)	2.10 (1.47, 3.00)	1.76 (1.15, 2.67)	1.76 (1.15, 2.67)	1.76 (1.15, 2.67)	1.76 (1.15, 2.67)	1.76 (1.15, 2.67)	1.76 (1.15, 2.67)
Not tested															
Previous AIDS															
No	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Yes	2.83 (2.21, 3.63)	2.06 (1.57, 2.71)	2.54 (1.96, 3.30)	1.92 (1.45, 2.54)	2.55 (1.97, 3.30)	1.93 (1.46, 2.55)	1.96 (1.48, 2.60)	1.93 (1.46, 2.55)	2.68 (2.07, 3.46)	1.96 (1.48, 2.60)	1.96 (1.48, 2.60)	1.96 (1.48, 2.60)	1.96 (1.48, 2.60)	1.96 (1.48, 2.60)	1.96 (1.48, 2.60)

Bold values represent significant covariates at p<0.05 in the final model.

Abbreviations: cART = combination antiretroviral therapy; NRTI = nucleoside reverse transcriptase inhibitors; NNRTI = non-nucleoside reverse transcriptase inhibitors; PI = protease inhibitors; uL = micro litre; mL = millilitre; df = degrees of freedom; HR = hazard ratio; and CI = confidence interval.