Journal of Clinical Epidemiology 81 (2017) 129–139

Journal of Clinical Epidemiology

A pseudo-random patient sampling method evaluated

Nicole L. De La Mata[a,*], Mi-Young Ahn[b], Nagalingeswaran Kumarasamy[c], Penh Sun Ly[d], Oon Tek Ng[e], Kinh Van Nguyen[f], Tuti Parwati Merati[g], Thuy Thanh Pham[h], Man Po Lee[i], Nicolas Durier[j], Matthew G. Law[a]

[a]*The Kirby Institute, UNSW Australia, Wallace Wurth Building, Sydney, NSW 2052, Australia*
[b]*Department of Internal Medicine and AIDS Research Institute, Yonsei University College of Medicine, Severance Hospital, Seoul, South Korea*
[c]*Chennai Antiviral Research and Treatment Clinical Research Site (CART CRS), YRGCARE Medical Centre, VHS, Chennai, India*
[d]*National Center for HIV/AIDS, Dermatology & STDs, Phnom Penh, Cambodia*
[e]*Department of Infectious Diseases, Tan Tock Seng Hospital, Tan Tock Seng, Singapore*
[f]*National Hospital for Tropical Diseases, Hanoi, Vietnam*
[g]*Department of Internal Medicine, Udayana University, Sanglah Hospital, Bali, Indonesia*
[h]*Bach Mai Hospital, Hanoi, Vietnam*
[i]*Department of Medicine, Queen Elizabeth Hospital, Hong Kong, China*
[j]*TREAT Asia, amfAR—The Foundation for AIDS Research, Bangkok, Thailand*

Accepted 23 September 2016; Published online 19 October 2016

Abstract

Objectives: To compare two human immunodeficiency virus (HIV) cohorts to determine whether a pseudo-random sample can represent the entire study population.

Study Design and Setting: HIV-positive patients receiving care at eight sites in seven Asian countries. The TREAT Asia HIV Observational database (TAHOD) pseudo-randomly selected a patient sample, while TREAT Asia HIV Observational database-Low Intensity Transfer (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 2,318 TAHOD patients and 14,714 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.

Conclusion: Our study gives the first empirical evidence that analysis of risk factors for completely ascertained end points 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 analyzing 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

1. 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

Conflict of interest: The authors did not have any conflict of interest to declare.

* Corresponding author. Tel.: (02) 9385 9012; fax: (02) 9385 0940.
E-mail address: ndelamata@kirby.unsw.edu.au (N.L. De La Mata).

http://dx.doi.org/10.1016/j.jclinepi.2016.09.012
0895-4356/

What is new?

Key findings

- Our pseudo-random patient sample, the TREAT Asia HIV Observational database, is representative of our larger study population, TREAT Asia HIV Observational database-Low Intensity Transfer, and produced comparable findings for routinely ascertained end points relating to the response to antiretroviral treatment.
- End points 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.

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 human immunodeficiency virus (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 use 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. 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 substudy of TAHOD, was initiated where data were 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.

2. Methods

2.1. Data collection and participants

TAHOD is a collection of 20 HIV treatment centers across the Asia-Pacific region including China (one site), Hong Kong (one site), Taiwan (one site), South Korea (one site), India (two sites), Indonesia (two sites), Malaysia (two sites), Philippines (one site), Singapore (one site), Thailand (four sites), Japan (one site), Cambodia (one site), and Vietnam (two sites). Detailed data were collected for a subset of patients who attend care at the sites. Patients were not entirely randomly recruited instead each site consecutively selected patients who 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].

TAHOD-LITE is a substudy of TAHOD and currently involves only 8 of the 20 TAHOD sites from Cambodia (one site), Hong Kong (one site), India (one site), Indonesia (one site), Singapore (one site), South Korea (one site), and Vietnam (two 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 who 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.