At-Risk Alcohol Use Among HIV-Positive Patients and the Completion of Patient-Reported Outcomes

Jacqueline E. Rudolph[1] · Stephen R. Cole[1] · Jessie K. Edwards[1] ·
Richard Moore[2] · Conall O'Cleirigh[3] · Wm. Christopher Mathews[4] ·
Katerina Christopoulos[5] · For the Center for AIDS Research Network of Integrated
Clinical Systems

Abstract Heavy drinking is prevalent among people living with HIV. Studies use tools like patient-reported outcomes (PROs) to quantify alcohol use in a detailed, timely manner. However, if alcohol misuse influences PRO completion, selection bias may result. Our study included 14,145 adult HIV patients (133,036 visits) from CNICS who were eligible to complete PROs at an HIV primary care visit. We compared PRO completion proportions between patients with and without a clinical diagnosis of at-risk alcohol use in the prior year. We accounted for confounding by baseline and visit-specific covariates. PROs were completed at 20.8% of assessed visits. The adjusted difference in PRO completion proportions was −3.2% (95% CI −5.6 to −0.8%). The small association between receipt of an at-risk alcohol use diagnosis and decreased PRO completion suggests there could be modest selection bias in studies using the PRO alcohol measure.

Keywords Patient-reported outcomes · PROs · HIV · Alcohol consumption · Selection bias

✉ Jacqueline E. Rudolph
  jerudolp@live.unc.edu

[1] Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, 2101 McGavran Greenberg Hall, CB# 7435, Chapel Hill, NC 27599, USA

[2] Department of Medicine, Johns Hopkins University, Baltimore, MD, USA

[3] Harvard Medical School, Boston, MA, USA

[4] Department of Medicine, University of California at San Diego, San Diego, CA, USA

[5] Department of Medicine, University of California at San Francisco, San Francisco, CA, USA

Introduction

In the United States, people living with human immunodeficiency virus (HIV) exhibit a higher prevalence of heavy drinking and alcohol use disorders than the general population [1–3]. Long-term heavy drinking has been linked to worsened health in HIV-positive populations, through such mechanisms as immune dysregulation and decreased adherence to antiretroviral therapy (ART) [1, 3–5]. To inform alcohol interventions in these populations, it is critical that future work examine the relationship between alcohol consumption and HIV outcomes like mortality or viremia. Before these relationships can be examined, though, we must assess the measurement properties of the metrics used to estimate alcohol intake. Here, we examine whether using patient-reported outcome (PRO) questionnaires that measure alcohol consumption would likely provide valid estimates of the effects of alcohol in the Center for AIDS Research (CFAR) Network of Integrated Clinical Systems (CNICS), a large, geographically diverse HIV cohort.

PROs are an increasingly powerful tool for collecting rich, timely data [6–8]. Patient responses recorded using individual tablet computers (a common means of collecting PROs) are expected to be less affected by social desirability bias than those given during in-person interviews [8, 9]. This is especially the case for private information like substance use and sexual behavior. Thus, PROs are expected to identify more risky behaviors than physician-reported diagnoses.

However, in many large cohorts like CNICS, only a subset of participants may complete PROs. Nevertheless, it is a goal of CNICS to have active patients complete the PROs, and the responses from the subset completing the PROs are sometimes used as data to draw inferences [10].

It is unknown whether any specific factors affect a patient's willingness to participate in completion of a PRO, and the subgroup completing PROs could be a select sample of the full cohort [11].

We sought to learn about the subsample of CNICS patients that complete the PROs and to determine whether their self-reported data is susceptible to selection bias. Selection bias can occur when a non-random sample is selected and factors associated with selection are directly affected by both exposure and outcome or have a common cause with the exposure and with the outcome [12, 13]. We investigated the potential for bias by assessing whether having received an at-risk alcohol use diagnosis, as a marker for heavy drinking, was associated with completion of the alcohol consumption questions on the CNICS PRO. We hypothesized that patients with an at-risk alcohol use diagnosis would be less likely to complete the PRO alcohol questions.

Methods

Cohort Description

We used data from CNICS, a clinic-based cohort consisting of over 32,000 HIV-infected adults aged 18 or older who sought care at one of eight clinical sites (https://www.uab.edu/cnics/). The study sites are located in geographically diverse, urban areas of the United States: Birmingham, AL; Baltimore, MD; Seattle, WA; San Francisco, CA; San Diego, CA; Chapel Hill, NC; Boston, MA; and Cleveland, OH.

The data center for CNICS draws together select information from clinic, administrative, and medical records as well as PROs. Patient visits occurred approximately every 3–6 months (although time between visits varied by patient), with a goal that PROs would be filled out every 4–6 months [10]. The questionnaires (referred to by CNICS as PROs) collect information on patient outcomes like body morphology and HIV symptoms as well as patient behaviors like alcohol consumption. Currently, PROs are available at seven of the eight CNICS sites. Institutional review boards at each site approved study procedures. Participants provided written informed consent to be included in the CNICS cohort or contributed administrative or clinical data with a waiver of written informed consent where approved by local institutional review board(s). This study was reviewed by the Office of Human Research Ethics at the University of North Carolina at Chapel Hill and was determined to not constitute human subjects research as defined under federal regulations.

Sample Selection

CNICS patients were eligible for inclusion if they attended at least one HIV primary care visit between implementation of the PROs at one of the seven CNICS clinics collecting PRO data and November 2014. Patients were allowed to contribute multiple visits. Prior to the application of any exclusion criteria, there were 178,877 recorded visits contributed by 16,028 patients.

We included only visits on which a patient was likely to be eligible to complete a PRO. Visits that occurred within 108 days of the last visit on which a patient completed a PRO were excluded (29,947 visits), based on the CNICS protocol that PROs can only be offered to patients at visits more than 108 days apart. Visits were excluded if the patient's recorded CNICS start date occurred after the medical record visit date (1099 visits).

Patients were excluded if they had no baseline data, e.g. race/ethnicity, HIV risk factor, and age at baseline (6214 visits). Patients were also excluded if they were not White non-Hispanic, Black non-Hispanic, or Hispanic because the other non-Hispanic group was small and likely heterogeneous (8581 visits). After accounting for these restrictions, 14,145 patients contributing 133,036 visits were included in this study sample.

Variable Definitions

Our main exposure was defined as receipt of at least one "at-risk alcohol use" clinical diagnosis (i.e. a diagnosis recorded by the physician in the electronic medical record but no ICD codes were provided) within the year prior to an eligible CNICS visit, as our best estimate of recent heavy drinking for the full study sample given the available data. A patient was not considered exposed if the diagnosis was received on the day of the visit, due to concerns that completing a PRO might prompt a patient to inform their doctor of their alcohol misuse. Our main outcome was defined as completion of the PRO alcohol consumption questions on the same day as an eligible CNICS visit. The required PRO questions were as follows: (1) how often do you have a drink containing alcohol; (2) how many drinks containing alcohol do you have on a typical day when you are drinking; and (3) how often do you have five or more drinks on one occasion? [14, 15] These are the first three questions of the Alcohol Use Disorders Identification Test (AUDIT) questionnaire or all of the questions on the shortened AUDIT consumption (AUDIT-C) questionnaire. Here, a patient was considered to have the outcome if he completed all three questions or if he answered "Never" to the first question (the latter causes the computer-based questionnaire to automatically skip the remaining alcohol questions) [7].

We considered several covariates in our analyses. Variables that were collected at baseline for all patients included enrollment date, sex, race/ethnicity, study site, and HIV risk category (i.e. injection drug user or men who have sex with men). Variables that were updated at each eligible HIV primary care visit were prior clinical diagnosis of substance use, including amphetamines, cocaine, and opiates; whether ART had been initiated; most recent and nadir CD4 counts; most recent and peak log_{10} HIV1 RNA level (viral load); age; time since PRO introduction at study site; past completion of a PRO; and time since entry into CNICS. To ensure we were not controlling for variables affected by exposure, all potential confounders had to occur prior to the date on which exposure status was recorded.

Statistical Analyses

We first assessed whether the subset completing PROs differed from those who did not complete PROs. We compared the distribution of patient characteristics listed above both for first eligible visits as well as for all eligible visits. Categorical variables were compared using a Chi Square test, and continuous variables were compared using a two-sample T test (both using an α level of 0.05).

In our main analysis, we used a linear binomial model to estimate a difference in the proportion of patients completing the required PRO alcohol questions comparing patients who had received an "at-risk alcohol use" diagnosis within one year of the eligible CNICS visit to those who had not. A log binomial model was used to estimate a ratio of proportions. We accounted for repeated visits using generalized estimating equations and excluded patients missing any covariate data; 1% of visits (1564 visits) were excluded due to missing data.

We conducted a crude and covariate-adjusted analysis. In our adjusted analysis, we controlled for the measured baseline and visit-specific covariates using stabilized inverse probability of treatment weights (IPTW) [16]. Restricted quadratic splines with four knots were used to flexibly model continuous variables [17].

We also undertook several sensitivity analyses. First, we assessed whether the results changed if different definitions of the exposure and outcome were used. The alternate exposure definition considered was receipt of an "at-risk alcohol use" diagnosis at any point prior to the eligible visit. Alternate outcome definitions examined were whether the patient initiated a PRO on the eligible visit date and, among those who initiated a PRO, whether they completed the alcohol questions. The linear binomial model used in the main analysis was run for all definition changes.

Next, we assessed the effect of exposure misclassification. As a method of determining whether a patient has a recent history of heavy drinking or an alcohol use disorder, the physician-reported diagnoses are expected to be specific but not sensitive. The most severe cases are likely recorded, but mild or moderate cases might be missed. We assumed the specificity of the diagnoses to be one and recalculated the difference in proportions by "correcting" our weighted counts of exposed and unexposed events and totals for a range of sensitivities between 0.25 and 1. The lower bound was chosen based on the smallest estimate found for the percent of alcoholics whose disorder was detected by their physician, and the upper bound reflects the assumed sensitivity in the main analysis [18, 19]. All statistical analyses were carried out in SAS software version 9.4 (SAS Institute Inc., Cary, NC).

Results

At their first eligible visit, 18.4% of eligible patients completed the required PRO alcohol questions (Table 1). Both PRO completers and non-completers were primarily male. Completers were more likely to be both injection drug users and MSM than non-completers. PRO-completers were more likely to be Black non-Hispanic and less likely to be Hispanic than non-completers (35.9 and 32.0% Black, respectively; 11.4 and 17.5% Hispanic). Mean ages at baseline and visit were comparable. Patients completing the PROs had slightly higher most recent CD4 cell counts, lower nadir CD4 cell counts, lower recent viral loads, but similar peak viral loads as non-completers. Greater time had elapsed since study entry and since PRO implementation at the study site for completers than non-completers. Compared to non-completers, patients completing the PROs were more likely to be on ART (79.4% of completers and 70.8% of non-completers). Distributions of past substance use diagnoses were similar.

Patients completed the PRO alcohol questions at 20.8% of all eligible HIV primary care visits (Table 2). Similar distributions for baseline patient characteristics were observed as for first visits. Visits where the alcohol questions were completed were less likely to be contributed by patients with a history of substance use for all drugs examined and more likely to be contributed by patients on ART. Visits where PROs were completed were also more likely to be contributed by a patient who had completed a PRO in the past (68.1% of completion visits and 19.0% of non-completion visits). The completion and non-completion subsamples had nearly identical CD4 cell counts and viral loads. The age distributions were comparable. Visits where PROs were completed occurred, on average, more

Table 1 Baseline characteristics by PRO alcohol question completion, at first eligible visit and all eligible visits

Characteristics	First visit			All visits		
	PRO completed (n = 2601)	PRO not completed (n = 11,544)	P-value	PRO completed (n = 27,668)	PRO not completed (n = 105,368)	P-value
Male, n (%)	2185 (84.0)	9799 (84.9)	0.3	23,385 (84.5)	90,589 (86.0)	<0.0001
Race/ethnicity, n (%)						
White non-Hispanic	1371 (52.7)	5824 (50.5)	<0.0001	14,267 (51.6)	57,339 (54.4)	<0.0001
Black non-Hispanic	933 (35.9)	3698 (32.0)		9276 (33.5)	26,273 (24.9)	
Hispanic	297 (11.4)	2022 (17.5)		4125 (14.9)	21,756 (20.7)	
HIV risk category, n (%)						
MSM	1625 (63.2)	7372 (64.8)	0.0007	17,804 (65.0)	69,578 (66.9)	<0.0001
IDU	170 (6.6)	745 (6.6)		1420 (5.2)	8158 (7.8)	
Both	136 (5.3)	405 (3.6)		1287 (4.7)	3969 (3.8)	
Other	641 (24.9)	2854 (25.1)		6901 (25.2)	22,355 (21.5)	
Baseline age, mean (SD)	39.7 (10.4)	39.7 (10.5)	0.8	39.8 (10.1)	40.2 (10.0)	<0.0001

CNICS Center for AIDS Research Network of Integrated Clinical Systems, *HIV* human immunodeficiency virus, *MSM* men who have sex with men, *IDU* injection drug user, *SD* standard deviation

years after the PROs were implemented than non-completion visits (3.1 years and 2.4 years, respectively).

In the main analysis, 21.0% of those unexposed to and 15.8% of those exposed to an at-risk alcohol use diagnosis within the year prior to a CNICS HIV primary care visit completed the PRO alcohol questions on the day of said visit (Table 3). Comparing exposed to unexposed patients, the crude difference in PRO-completion proportions was −5.2% (95% CI −6.6 to −3.8%). After adjusting for confounding, the difference in the percent of patients completing the alcohol questions was −3.2% (95% CI −5.6 to −0.8%) comparing exposed to unexposed patients.

In the alternate definition analyses (Table 4), when any prior receipt of an at-risk alcohol use diagnosis was the exposure, estimates were closer to the null (crude −1.5%, 95% CI −2.7 to −0.4%; weighted −1.3%, 95% CI −2.6 to −0.0%). Using PRO initiation as the outcome did not substantially change the difference in proportion estimates. With the main analysis exposure definition, the crude difference in PRO completion proportions was −5.1% (95% CI −6.6 to −3.7%), and the weighted difference was −3.1% (95% CI −5.6 to −0.6%).

Within the group of visits where a PRO was initiated (28,624 visits), the difference in proportions of patients completing the alcohol questions was attenuated compared to the main analysis results. For the main analysis exposure definition, the crude difference was −1.8% (95% CI −3.6 to 0.1%), and the weighted difference was −1.6% (95% CI −4.2 to 1.1%). Patients were highly likely to complete the alcohol questions if they had already initiated the PRO;

94.1% of visits contributed by a patient with an at-risk alcohol diagnosis in the year prior met this outcome criterion.

Lastly, in the analysis assessing the effects of non-differential exposure misclassification, adjustment of the weighted visits resulted in estimates that were further from the null (Table 5). The distance from the observed estimate increased as the sensitivity decreased; however, changes were small. At the lowest sensitivity of 0.25, the adjusted difference in PRO-completion proportions was −3.6%, only 0.4 percentage points lower than the estimate when assuming perfect sensitivity.

Discussion

In this study, we estimated the difference in the proportion of patients completing the PRO alcohol questions, comparing patients with an at-risk alcohol use diagnosis in the year prior to a CNICS HIV primary care visit to those without. We observed that PROs were completed at very few of the assessed visits and that patients with a diagnosis were less likely to complete the PRO alcohol questions. However, neither the crude nor adjusted difference in proportions indicated a particularly strong relationship between at-risk alcohol use and completion of the PROs. The strength of association decreased further if the exposure was any prior receipt of an at-risk alcohol use diagnosis rather than receipt in the prior year. After adjustment for exposure misclassification, estimates were further from

Table 2 Visit-specific characteristics by PRO alcohol question completion, at first eligible visit and all eligible visits

Characteristics	First visit			All visits		
	PRO completed ($n = 2601$)	PRO not completed ($n = 11,544$)	P-value	PRO completed ($n = 27,668$)	PRO not completed ($n = 105,368$)	P-value
Age at visit, mean (SD)	42.6 (10.7)	42.3 (10.8)	0.3	44.5 (10.5)	44.7 (10.5)	0.006
Past year alcohol diagnosis, n (%)	89 (3.4)	468 (4.1)	0.1	778 (2.8)	4141 (3.9)	<0.0001
Any prior alcohol diagnosis, n (%)	417 (16.0)	1614 (14.0)	0.007	4701 (17.0)	19,300 (18.3)	<0.0001
Any prior cocaine diagnosis, n (%)	314 (12.1)	1476 (12.8)	0.3	3206 (11.6)	15,107 (14.3)	<0.0001
Any prior amphetamine diagnosis, n (%)	287 (11.0)	1163 (10.1)	0.1	3228 (11.7)	15,718 (14.9)	<0.0001
Any prior opiates diagnosis, n (%)	99 (3.8)	472 (4.1)	0.5	1051 (3.8)	7032 (6.7)	<0.0001
On ART, n (%)	2066 (79.4)	8167 (70.7)	<0.0001	25,429 (91.9)	93,914 (89.1)	<0.0001
CD4 counts, mean (SD)						
Most recent	500.6 (298.5)	486.2 (297.9)	0.03	545.7 (301.7)	520.1 (300.1)	<0.0001
Nadir since baseline	294.1 (245.4)	310.2 (254.2)	0.004	255.0 (216.1)	261.0 (216.2)	<0.0001
Log_{10} viral load, mean (SD)						
Most recent	2.5 (1.4)	2.7 (1.4)	<0.0001	2.1 (1.0)	2.3 (1.1)	<0.0001
Peak since baseline	4.3 (1.4)	4.2 (1.4)	0.08	4.5 (1.3)	4.5 (1.3)	0.0002
Past PRO completion, n (%)	0 (0.0)	0 (0.0)		18,836 (68.1)	19,969 (19.0)	<0.0001
Years since PRO introduction, mean (SD)	1.7 (1.8)	1.5 (1.8)	<0.0001	3.1 (1.7)	2.4 (1.7)	<0.0001
Years since entry into CNICS, mean (SD)	3.3 (3.6)	3.0 (3.7)	0.0002	5.2 (4.0)	5.0 (4.1)	<0.0001

CNICS Center for AIDS Research Network of Integrated Clinical Systems, *SD* standard deviation, *ART* antiretroviral therapy, *PRO* patient reported outcomes

Table 3 Association between receipt of an at-risk alcohol use diagnosis in the year prior to a CNICS HIV primary care visit and completion of the PRO alcohol questions

Exposure	Total visits	No. PRO completions	PRO-completion prop.	Difference in prop.	95% CI	Ratio of prop.	95% CI
Crude analysis							
Alcohol diagnosis in past year	4889	774	0.158	−0.052	−0.066, −0.038	0.75	0.69, 0.82
No alcohol diagnosis in past year	126,583	26,638	0.210	0	Reference	1	Reference
Weighted analysis[a]							
Alcohol diagnosis in past year	4821	855	0.177	−0.032	−0.056, −0.008	0.85	0.74, 0.97
No alcohol diagnosis in past year	125,842	26,384	0.210	0	Reference	1	Reference

CNICS Center for AIDS Research Network of Integrated Clinical Systems, *HIV* human immunodeficiency virus, *PRO* patient reported outcomes, *No.* number, *Prop.* proportion, *CI* confidence interval

[a] Stabilized inverse probability of treatment weights were used to control for confounding by sex, race/ethnicity, study site, HIV risk category, prior substance use diagnoses, ART initiation, most recent and peak CD4 counts, most recent and nadir log_{10} viral load, age, time since PRO introduction at study site, past completion of a PRO, and time since entry into CNICS

Table 4 Results of the alternate definition analyses

Exposure	Total Visits	No. PRO completions	PRO-completion prop.	Crude difference in prop.	95% CI	Weighted[a] difference in prop.	95% CI
Main outcome definition: all alcohol PRO questions completed							
Alternate exposure definition 1							
Any prior alcohol diagnosis	23,876	4679	0.196	−0.015	−0.027, −0.004	−0.013	−0.026, −0.000
Never alcohol diagnosis	107,596	22,733	0.211	0	Reference	0	Reference
Alternate outcome definition 1: PRO initiated							
Main exposure definition							
Alcohol diagnosis in past year	4889	823	0.168	−0.051	−0.066, −0.037	−0.031	−0.056, −0.006
No alcohol diagnosis in past year	126,583	27,801	0.220	0	Reference	0	Reference
Alternate exposure definition 1							
Any prior alcohol diagnosis	23,876	4944	0.207	−0.013	−0.025, −0.001	−0.013	−0.026, 0.001
Never alcohol diagnosis	107,596	23,680	0.220	0	Reference	0	Reference
Alternate outcome definition 2: among those who initiated a PRO, completion of alcohol questions							
Main exposure definition							
Alcohol diagnosis in past year	823	774	0.941	−0.018	−0.036, 0.001	−0.016	−0.042, 0.011
No alcohol diagnosis in past year	27,801	26,638	0.958	0	Reference	0	Reference
Alternate exposure definition 1							
Any prior alcohol diagnosis	4944	4679	0.946	−0.014	−0.022, −0.006	−0.008	−0.016, −0.000
Never alcohol diagnosis	23,680	22,733	0.960	0	Reference	0	Reference

PRO patient reported outcomes, *No.* number, *Prop.* proportion, *RD* risk difference, *CI* confidence interval

[a] Stabilized inverse probability of treatment weights were used to control for confounding by sex, race/ethnicity, study site, HIV risk category, prior substance use diagnoses, ART initiation, most recent and peak CD4 counts, most recent and nadir $\log_{10}$ viral load, age, time since PRO introduction at study site, past completion of a PRO, and time since entry into CNICS

the null, though still small. When examining patients who initiated a PRO, we saw a high proportion of completion of the alcohol questions in both exposure groups.

These findings are informative for researchers working in CNICS and related settings. Alcohol consumption is often a variable of interest in studies of people living with HIV due to the high prevalence of heavy drinking and alcohol use disorders in this population [1, 2]. Some studies have reported that the level of heavy drinking in people living with HIV is twice the level in the general US population [2]. Moreover, alcohol use has been found to be adversely associated with HIV disease progression. Alcohol misuse can lead to tissue inflammation, immune dysregulation, increased viral replication, and higher susceptibility to opportunistic infections [1, 3, 4, 20]. Alcohol use is also associated with decreased ART

adherence [3, 20]. Thus, patients who misuse alcohol tend to see worse clinical HIV outcomes.

PROs can be a powerful way to collect alcohol use data. These self-reported questionnaires are rich sources of timely data and can be implemented in clinical settings with minimal burden on patients and staff (www.acade myhealth.org/files/2012/sunday/crane2.pdf). In addition, they are less subject to social desirability bias than in-person interviews [8, 9]. CNICS uses the validated AUDIT and AUDIT-C surveys to detect problem drinking in their PROs [14, 15]. AUDIT is especially useful because it was designed to catch drinking problems before they progress to severe dependence or alcoholism [21].

However, when selection of patients to complete PROs is self-determined, selection bias may arise. Any measure of association may be biased if factors associated with

Table 5 Weighted difference in PRO-completion proportions,[a] corrected for non-differential misclassification of at-risk alcohol use diagnoses

Sensitivity	Exposed PRO completions[b, c]	Total exposed visits[b, c]	Weighted difference in prop.[c]
1.00	855	4821	−0.032
0.95	900	5075	−0.032
0.90	950	5357	−0.032
0.85	1006	5672	−0.033
0.80	1069	6026	−0.033
0.75	1140	6428	−0.033
0.70	1222	6887	−0.033
0.65	1316	7417	−0.033
0.60	1425	8035	−0.033
0.55	1555	8766	−0.033
0.50	1710	9642	−0.034
0.45	1900	10,714	−0.034
0.40	2138	12,053	−0.034
0.35	2443	13,775	−0.035
0.30	2851	16,071	−0.035
0.25	3421	19,285	−0.036

RD risk difference, *Prop.* proportion

[a] Defined as completion of the required three alcohol questions

[b] Exposure in this analysis was receipt of an at-risk alcohol use diagnosis within the year prior to the eligible CNICS visit

[c] Counts and differences in proportions were weighted by stabilized inverse probability of treatment weights

selection are affected by or have a common cause with the exposure and are affected by or have a common cause with the outcome [12]. In other words, selection bias can be considered a type of collider stratification bias [13].

The bias can also be framed as a missing data problem. If a researcher uses the PRO data, some CNICS patients will not have answered the alcohol questions or will not be included. Bias can occur when, as above, the missingness is a collider [12]. Additionally, we suspect that the alcohol data is "missing not at random," i.e. the missingness is associated with the unmeasured alcohol consumption. Unlike other types of missing data, "missing not at random" typically cannot be controlled for using multiple imputation or inverse probability of missingness weights [12, 22]. Researchers can explore the effects of data missing not at random in sensitivity analyses [23].

There are several strengths of our study. We were able to examine the association between alcohol-related diagnoses and PRO completion in over 130,000 patient-visits contributed by more than 14,000 HIV-positive patients. CNICS collects data in the whole cohort and in a subset, which allowed us to examine the PRO selection process and the potential for selection bias. We were able to control for a variety of baseline and visit-specific patient characteristics. Furthermore, CNICS is fairly representative of the population of newly diagnosed HIV patients in the United States [24].

This study had several limitations. We could not account for mental health diagnoses, which we believe could be confounders of the association between at-risk alcohol use and PRO completion. It is probable that there were other, unmeasured confounders, as it is nearly impossible to measure all components of a "healthy lifestyle." Furthermore, we were limited in the type of sensitivity analyses we could conduct. For example, it would have been informative to use a biomarker as an alternate exposure definition, but there were no available laboratory correlates for recent or chronic heavy drinking. We were also limited by the fact that we only had record of a clinical diagnosis for "at-risk alcohol use," rather than more specific ICD codes. Due to this, we could not look at the differences in completion by different alcohol diagnoses.

Another limitation was that it was impossible to know at which visits patients were actually offered a PRO to complete. We had to use the observed data and fairly strict exclusion criteria to try and select only those visits where a patient was most likely to be considered eligible to complete a PRO. Unfortunately, a limitation of the data was that we could not distinguish whether a patient was considered incapable of completing a PRO at a particular visit, as might occur if they were intoxicated, or if they simply were not offered a PRO to complete. Our choice to exclude patients who were not Black, White, or Hispanic could further limit the interpretation of our results. These

individuals could have been more likely to have alcohol diagnoses and even less likely to complete a PRO. However, in the CNICS, this race/ethnicity group is very small (less than 8% of the cohort), so their exclusion was unlikely to greatly affect our results.

Also, because we looked only at patients who arrived at their visits, our study does not address the issue of lower retention in care among CNICS patients who are heavy drinkers, which is another potential source of selection bias that researchers need to consider when working with the alcohol data [25]. Lastly, there was no way to determine whether a patient filled out the PRO accurately. It is possible that those who misuse alcohol (some of whom may not have a clinical diagnosis for at-risk alcohol use) are more likely to misrepresent their alcohol consumption when completing the PROs. For a researcher using the PRO alcohol data, this misclassification may be as important as any selection bias incurred from heavy drinkers not completing the PROs.

There are several further steps that could be taken to continue examining PRO completion in this cohort. First, our analysis could be repeated for other variables in the PRO data, to assess whether use of those variables as an exposure or confounder could result in bias. Second, one could delve deeper into the relationship between receipt of alcohol diagnoses and PRO completion by considering exposure trajectories (i.e. compare patients who had many diagnoses since enrollment into CNICS to those with few). This might shed light on the relationship between long-term drinking behavior and patterns of PRO completion. Our goal in this paper was to explore the potential selection bias resulting from differences in PRO completion and to quantify that bias to inform sensitivity analyses. We chose the best available marker of recent heavy drinking available for the entire cohort as an example variable that could be associated with differences in PRO completion.

Conclusions

That we observed an association between receipt of an at-risk alcohol use diagnosis and completion of the PROs, even after controlling for a variety of patient characteristics, supports the theory that the data may be missing not at random. However, the strength of association was modest. We are not suggesting that our estimated difference in proportions accurately reflects the magnitude of the bias that will occur if the PRO alcohol data is used as an exposure because that will depend on factors such as the strength of the association between the chosen outcome and PRO completion. Our estimate could, though, provide an upper bound of the bias caused by conditioning on a collider [26] and could be used to inform future sensitivity analyses. More broadly, our findings serve as information to raise awareness that the PRO subset does differ from the entire CNICS cohort and that a researcher may wish to pursue sensitivity analyses for selection bias in studies that estimate effects using the PRO alcohol data.

Appendix

See Tables 6 and 7.

Table 6 Baseline characteristics of CNICS patients at first eligible visit and all eligible visits

Characteristics	First visit		All visits	
	All patients ($n = 14{,}145$)	Alcohol questions completed ($n = 2601$)	All eligible visits ($n = 133{,}036$)	Alcohol questions completed ($n = 27{,}668$)
Male, n (%)	11,984 (84.7)	2185 (84.0)	113,974 (85.7)	23,385 (84.5)
Race/ethnicity, n (%)				

Table 6 continued

Characteristics	First visit		All visits	
	All patients (n = 14,145)	Alcohol questions completed (n = 2601)	All eligible visits (n = 133,036)	Alcohol questions completed (n = 27,668)
White non-Hispanic	7195 (50.9)	1371 (52.7)	71,606 (53.8)	14,261 (51.6)
Black non-Hispanic	4631 (32.7)	933 (35.9)	35,549 (26.7)	9276 (33.5)
Hispanic	2319 (16.4)	297 (11.4)	25,881 (19.5)	4125 (14.9)
HIV risk category, n (%)				
MSM	8997 (64.5)	1625 (63.2)	87,382 (66.5)	17,804 (65.0)
IDU	915 (6.6)	170 (6.6)	9578 (7.3)	1420 (5.2)
Both	541 (3.9)	136 (5.3)	5256 (4.4)	1287 (4.7)
Other	3495 (25.1)	641 (24.9)	29,256 (22.3)	6901 (25.2)

CNICS Center for AIDS Research Network of Integrated Clinical Systems, *HIV* human immunodeficiency virus, *MSM* men who have sex with men, *IDU* injection drug user

Table 7 Visit-specific characteristics of CNICS patients at first eligible visit and all eligible visits

Characteristics	First visit		All visits	
	All patients (n = 14,145)	Alcohol questions completed (n = 2601)	All eligible visits (n = 133,036)	Alcohol questions completed (n = 27,668)
Age, mean (SD)	42.4 (10.8)	42.6 (10.7)	44.7 (10.5)	44.5 (10.5)
Past year alcohol diagnosis, n (%)	557 (3.9)	89 (3.4)	4919 (3.7)	778 (2.8)
Any prior alcohol diagnosis, n (%)	2031 (14.4)	417 (16.0)	24,001 (18.0)	4701 (17.0)
Any prior cocaine diagnosis, n (%)	1790 (12.7)	314 (12.1)	18,313 (13.8)	3206 (11.6)
Any prior amphetamine diagnosis, n (%)	1450 (10.3)	287 (11.0)	18,946 (14.2)	3228 (11.7)
Any prior opiates diagnosis, nn (%)	571 (4.0)	99 (3.8)	8083 (6.1)	1051 (3.8)
On ART, n (%)	10,233 (72.3)	2066 (79.4)	119,343 (89.7)	25,429 (91.9)
CD4 counts, mean (SD)				
Most recent	488.9 (298.0)	500.6 (298.5)	525.5 (300.6)	545.7 (301.7)
Nadir since baseline	307.2 (252.6)	294.1 (245.4)	259.7 (216.2)	255.0 (216.1)
Log_{10} viral load, mean (SD)				
Most recent	2.7 (1.4)	2.5 (1.4)	2.2 (1.1)	2.1 (1.0)
Peak since baseline	4.3 (1.4)	4.3 (1.4)	4.5 (1.3)	4.5 (1.3)
Past PRO completion, n (%)	0 (0.0)	0 (0.0)	38,805 (29.2)	18,836 (68.1)
Years since PRO introduction, mean (SD)	1.5 (1.8)	1.7 (1.8)	2.5 (1.7)	3.1 (1.7)
Years since entry into CNICS, mean (SD)	3.0 (3.7)	3.3 (3.6)	5.0 (4.0)	5.2 (4.0)

CNICS Center for AIDS Research Network of Integrated Clinical Systems, *SD* standard deviation, *ART* antiretroviral therapy, *PRO* patient reported outcomes

References

1. Dingle GA, Oei TP. Is alcohol a cofactor of HIV and AIDS? Evidence from immunological and behavioral studies. Psychol Bull. 1997;122(1):56–71.
2. Galvan FH, Bing EG, Fleishman JA, London AS, Caetano R, Burnam MA, et al. The prevalence of alcohol consumption and heavy drinking among people with HIV in the United States: results from the HIV Cost and Services Utilization Study. J Stud Alcohol. 2002;63(2):179–86.
3. Molina PE, Bagby GJ, Nelson S. Biomedical consequences of alcohol use disorders in the HIV-infected host. Curr HIV Res. 2014;12(4):265–75.
4. Shuper PA, Neuman M, Kanteres F, Baliunas D, Joharchi N, Rehm J. Causal considerations on alcohol and HIV/AIDS–a systematic review. Alcohol Alcohol. 2010;45(2):159–66.
5. Braithwaite RS, Conigliaro J, Roberts MS, Shechter S, Schaefer A, McGinnis K, et al. Estimating the impact of alcohol consumption on survival for HIV+ individuals. AIDS Care. 2007;19(4):459–66.
6. Kozak MS, Mugavero MJ, Ye J, Aban I, Lawrence ST, Nevin CR, et al. Patient reported outcomes in routine care: advancing data capture for HIV cohort research. Clin Infect Dis. 2012;54(1):141–7.
7. Crane HM, Lober W, Webster E, Harrington RD, Crane PK, Davis TE, et al. Routine collection of patient-reported outcomes in an HIV clinic setting: the first 100 patients. Curr HIV Res. 2007;5(1):109–18.
8. Lawrence ST, Willig JH, Crane HM, Ye J, Aban I, Lober W, et al. Routine, self-administered, touch-screen, computer-based suicidal ideation assessment linked to automated response team notification in an HIV primary care setting. Clin Infect Dis. 2010;50(8):1165–73.
9. Kissinger P, Rice J, Farley T, Trim S, Jewitt K, Margavio V, et al. Application of computer-assisted interviews to sexual behavior research. Am J Epidemiol. 1999;149(10):950–4.
10. Kitahata MM, Rodriguez B, Haubrich R, Boswell S, Mathews WC, Lederman MM, et al. Cohort profile: the centers for AIDS Research Network of Integrated Clinical Systems. Int J Epidemiol. 2008;37(5):948–55.
11. Edwards JK, Cole SR, Westreich D, Crane H, Eron JJ, Mathews WC, et al. Multiple imputation to account for measurement error in marginal structural models. Epidemiology. 2015;26(5):645–52.
12. Westreich D. Berkson's bias, selection bias, and missing data. Epidemiology. 2012;23(1):159–64.
13. Hernán MA, Hernández-Díaz S, Robins JM. A structural approach to selection bias. Epidemiology. 2004;15(5):615–25.
14. Bush K, Kivlahan DR, McDonell MB, Fihn SD, Bradley KA. The AUDIT alcohol consumption questions (AUDIT-C): an effective brief screening test for problem drinking. Ambulatory Care Quality Improvement Project (ACQUIP). Alcohol Use Disorders Identification Test. Arch Intern Med. 1998;158(16):1789–95.
15. Bradley KA, Bush KR, Epler AJ, Dobie DJ, Davis TM, Sporleder JL, et al. Two brief alcohol-screening tests From the Alcohol Use Disorders Identification Test (AUDIT): validation in a female Veterans Affairs patient population. Arch Intern Med. 2003;163(7):821–9.
16. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. 2000;11(5):550–60.
17. Howe CJ, Cole SR, Westreich DJ, Greenland S, Napravnik S, Eron JJ Jr. Splines for trend analysis and continuous confounder control. Epidemiology. 2011;22(6):874–5.
18. Bush B, Shaw S, Cleary P, Delbanco TL, Aronson MD. Screening for alcohol abuse using the CAGE questionnaire. Am J Med. 1987;82(2):231–5.
19. Moore RD, Bone LR, Geller G, Mamon JA, Stokes EJ, Levine DM. Prevalence, detection, and treatment of alcoholism in hospitalized patients. JAMA. 1989;261(3):403–7.
20. Chander G, Lau B, Moore RD. Hazardous alcohol use: a risk factor for non-adherence and lack of suppression in HIV infection. J Acquir Immune Defic Syndr. 2006;43(4):411–7.
21. Saunders JB, Aasland OG, Babor TF, de la Fuente JR, Grant M. Development of the Alcohol Use Disorders Identification Test (AUDIT): WHO Collaborative Project on Early Detection of Persons with Harmful Alcohol Consumption–II. Addiction. 1993;88(6):791–804.
22. Greenland S, Finkle WD. A critical look at methods for handling missing covariates in epidemiologic regression analyses. Am J Epidemiol. 1995;142(12):1255–64.
23. Daniel RM, Kenward MG, Cousens SN, De Stavola BL. Using causal diagrams to guide analysis in missing data problems. Stat Methods Med Res. 2012;21(3):243–56.
24. Lesko CR, Cole SR, Hall HI, Westreich D, Miller WC, Eron JJ, et al. The effect of antiretroviral therapy on all-cause mortality, generalized to persons diagnosed with HIV in the USA, 2009–2011. Int J Epidemiol. 2016;45:140–50.
25. Monroe AK, Lau B, Mugavero MJ, Mathews WC, Mayer KH, Napravnik S, et al. Heavy Alcohol Use is Associated with Worse Retention in HIV Care. J Acquir Immune Defic Syndr. 2016;73:419–25.
26. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. Epidemiology. 2003;14(3):300–6.