International Journal of Epidemiology, 2016, 140–150
doi: 10.1093/ije/dyv352
Advance Access Publication Date: 15 January 2016
Original article

Infectious Diseases

The effect of antiretroviral therapy on all-cause mortality, generalized to persons diagnosed with HIV in the USA, 2009–11

Catherine R Lesko,[1,2]* Stephen R Cole,[1] H Irene Hall,[3] Daniel Westreich,[1] William C Miller,[1,4] Joseph J Eron,[4] Jianmin Li[3] and Michael J Mugavero;[5] for the CNICS Investigators

[1]Department of Epidemiology, Gillings School of Global Public Health, University of North Carolina, NC, USA, [2]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD, USA, [3]National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Centers for Disease Control and Prevention, Atlanta, GA, USA, [4]Department of Medicine, Division of Infectious Diseases, University of North Carolina, NC, USA and [5]Department of Medicine, Division of Infectious Diseases, University of Alabama at Birmingham, Birmingham, AL, USA

*Corresponding author. Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, 615 North Wolf Street, Baltimore, MD 21205, USA. E-mail: clesko2@jhu.edu

Accepted 7 December 2015

Abstract

Background: Although antiretroviral therapy (ART) is known to be protective against HIV-related mortality, the expected magnitude of effect is unclear because existing estimates of the effect of ART may not directly generalize to recently HIV-diagnosed persons.

Methods: In this study, we estimated 5-year mortality risks for immediate versus no ART initiation among patients ($n = 12\,547$) in the Centers for AIDS Research Network of Integrated Clinical Systems (CNICS) using the complement of adjusted Kaplan–Meier survival functions. We subsequently standardized estimates to persons diagnosed with HIV in the USA between 2009 and 2011, who were enumerated using national surveillance data.

Results: The 5-year mortality, had all patients in the CNICS immediately initiated ART, was 10.6% [95% confidence interval (CI): 9.3%, 11.9%] compared with 28.3% (95% CI: 19.1%, 37.5%) had ART initiation been delayed at least 5 years. The 5-year mortality risk difference due to ART among patients in the CNICS was −17.7% (95% CI: −27.0%, −8.4%). Based on methods for generalizing an estimate from a study sample to a different target population, the expected risk difference due to ART initiation among recently HIV-diagnosed persons in the USA was −19.1% (95% CI: −30.5%, −7.8%).

Conclusions: Immediate ART initiation substantially lowers mortality among persons in the CNICS and this benefit is expected to be similar among persons recently diagnosed with HIV in the USA. We demonstrate a method by which concerns about generalizability can be addressed and evaluated quantitatively.

Key words: HIV, antiretroviral therapy, survival analysis, mortality, effect modification, external validity, generalizability

Key Messages

- Existing estimates of the effect of antiretroviral therapy on survival may not generalize directly to persons newly diagnosed with HIV in the USA between 2009 and 2011. Concerns about the generalizability of an estimate should be addressed whenever the study sample is not representative of the target population and the effect of interest is heterogeneous across subgroups.
- With additional information about the distribution of baseline covariates for the target population, generalizability can be quantitatively assessed through the use of inverse probability weights.
- Immediate initiation of antiretroviral therapy following engagement in care among HIV infected adults in the Center for AIDS Research Network of Integrated Clinical Systems cohort was associated with a risk of mortality that was 18% lower after 5 years; immediate initiation of antiretroviral therapy was expected to be associated with a 5-year risk of mortality that was 19% lower among persons diagnosed with HIV in the USA from 2009 to 2011.

Introduction

There is no question that effective combination antiretroviral therapy (ART) dramatically improves health and survival among HIV-infected persons and reduces transmission to susceptible sexual partners.[1–5] However, existing estimates of the effect of ART may not directly generalize to recently HIV-diagnosed persons. Lack of generalizability has long been recognized as a limitation of clinical trials[6,7] but, until recently,[8] it has received less attention in the context of observational studies and has seldom been quantified.[9]

The objective of this analysis was to estimate the anticipated absolute reduction in all-cause mortality across 5 years of follow-up if all persons recently HIV-diagnosed in the USA from 2009 to 2011 (the target population) had been prescribed ART immediately after diagnosis versus if treatment had been delayed. Given new treatment guidelines that suggest immediate ART initiation and public health efforts to link newly diagnosed persons to care and treatment immediately following diagnosis, curves describing the cumulative incidence of mortality associated with delays in ART initiation that have been generalized to the characteristics of persons recently diagnosed with HIV may be particularly compelling to motivate future newly HIV-diagnosed persons to engage in HIV care and treatment. Because data on the effect of ART in the target population are not available, we generalized data from a study sample that was not strictly representative of the target population; our approach is applicable to other research questions.

Informally, generalizability problems arise when the effect of a treatment is heterogeneous across patient subgroups and the study sample is not representative of the target population to whom we would like to apply its results.[10,11] Estimates of the magnitude of the survival benefit due to ART vary across the clinical cohorts and subgroups defined by patient demographics and clinical characteristics.[1,12] Furthermore, none of the clinical cohorts in which the effect of ART has been estimated is representative of persons recently diagnosed with HIV in the USA.[1,12] The Multicenter AIDS Cohort Study and Women's Interagency HIV Study (MACS/WIHS) are long-standing US-based interval cohorts of gay men and women at high risk for HIV infection, respectively; the hazard ratio for AIDS or death due to ART in the MACS/WIHS was 0.54 (95% CI: 0.38, 0.78).[13] The HIV-CAUSAL Collaboration is a predominantly European cohort; the hazard ratio for death due to ART was 0.48 (95% CI: 0.41, 0.57) overall. Among the 3730 US veterans included in the HIV-CAUSAL Collaboration, the hazard ratio was 0.62 (95% CI: 0.48, 0.82).[1] In the Swiss HIV Cohort Study, whose patients also contributed person-time to the HIV-CAUSAL analysis, the hazard ratio for AIDS or death was 0.14 (95% CI: 0.07, 0.29).[12] To our knowledge, no other studies of the effect of ART on mortality have included patients receiving care in the USA in the study sample, and no study has included persons with AIDS diagnoses at baseline, despite the reality that 26% of persons newly diagnosed with HIV receive an AIDS diagnosis either concurrently, or within 3 months of their HIV diagnosis.[14]

To obtain an estimate of the effect of ART that was generalizable to the target population, we first described the effect of ART on survival in the Centers for AIDS Research Network of Integrated Clinical Systems (CNICS) clinical cohort. We then described subgroup effects in the CNICS and reweighted patients in the CNICS to obtain an

estimate of the average effect of ART generalized to the target population. This approach to addressing the generalizability of an estimate allows for a quantitative assessment of the impact that a non-representative study sample may have on inference for a specific target population.

Methods

Study population

CNICS is an open clinical cohort of HIV-infected patients, 18 years of age, who have attended at least two HIV primary care visits at any of eight CNICS sites after 1 January 1995 (or the site-specific CNICS inception date). Patient demographics, HIV-related diagnoses, laboratory measurements and medications are abstracted from point-of-care electronic medical records. Mortality information is obtained from clinic sources, death certificates and the Social Security Death Index, which is queried regularly by CNICS sites.[15] Institutional review boards at each site approved participation in CNICS, and this analysis of de-identified data was determined not to constitute human subjects research by the institutional review board of the University of North Carolina at Chapel Hill.

The study population for this analysis included ART-naïve patients who enrolled in CNICS and had both CD4 cell count (cells/mm^3) and HIV-1 RNA viral plasma concentration (viral load, $\log_{10}$ copies/ml) measured at least once between 1 January 1998 (assumed to approximately represent the start of the modern ART era) and 30 December 2011 (the last full year for which complete mortality data were available). ART initiation was defined as having three or more antiretroviral medications prescribed on the same day, each for at least 30 days. Exclusion criteria included evidence of previous exposure to any antiretroviral medication (i.e. mono- or dual therapy or having an undetectable viral load). Of 12 995 ART-naïve patients who enrolled in CNICS, 448 (3%) were excluded because they were missing race/ethnicity ($n = 113$) or transmission risk category ($n = 349$). The final study population included 12 547 patients.

Target population

Our aim was to generalize the estimate of the impact of ART initiation on survival in CNICS to all persons diagnosed with HIV infection in the USA from 2009 to 2011. The number of HIV-diagnosed persons in categories defined by race/ethnicity, sex, age group, transmission risk and AIDS diagnosis within 3 months of HIV diagnosis, was provided by the Centers for Disease Control and Prevention from national HIV surveillance data and accounts for delays in reporting new infections.[16]

Statistical analysis

Let T_i be a random variable denoting survival time in days from CNICS enrolment to death from any cause. Let O_i be a random variable denoting time in days from CNICS enrolment to administrative censoring on 31 December 2011 (patients from one site were administratively censored on 15 September 2010), at 5 years of follow-up (to maintain adequately sized risk sets), or at loss to follow-up (defined as having no contact with the CNICS clinic, including therapy initiation or laboratory tests, over a 12-month period). Then let $T_i^* = min(T_i, O_i)$ and let the indicator D denote an observed death, $T_i^* = T_i$.

We pooled time into months to adjust for confounding and possibly differential loss to clinic. Let: A_k be an indicator of having initiated treatment by month k; j denote the number of completed months since the start of follow-up; $\overline{A}_{(k-1)}$ denote the history of exposure through month $k-1$; and $\overline{A}_{(-1)} = 0$ by definition. To mimic an intent-to-treat analysis, once patients initiated ART, they were assumed to remain on treatment for the rest of their time on study.

We estimated exposure effects based on contrasts in mortality under the interventions 'initiate ART immediately' versus 'delay ART initiation for at least 5 years following engagement in care'. That is, we were interested in contrasts of $P\left(T^{\overline{a}=1} < t\right)$ and $P\left(T^{\overline{a}=0} < t\right)$, where superscripts on T denote the potential value of T under exposure $\overline{a}$. In the CNICS, immediate ART initiation was operationalized as initiating ART within the first month following enrolment in the CNICS, which occurs at the second clinic visit. 'Immediate' ART initiation was thus consistent with routine clinical care (in which physicians would typically gather information and laboratory measurements on a patient before initiating them on ART). However, we did not explicitly build a grace period in our analysis (for example, see Cain *et al.* 2010).[17] 'Immediate' ART initiation in the target population is assumed to mean ART initiation within a similarly short window (i.e. 1 month). Thus, inherent in our presentation of the counterfactual cumulative incidence cure for immediately treated persons in the target population is the assumption that following diagnosis, a person would enter care and start ART within 1 month. We estimated mortality risks, risk differences and risk ratios using the complement of inverse probability weighted Kaplan–Meier survival functions.[18] For the sake of completion and comparability with previously published estimates, we calculated the relative hazard of mortality associated with ART use using an inverse probability

weighted marginal structural Cox proportional hazards model with Efron's approximation for tied death times.[19] We checked for violations of the proportional hazards assumption by checking the statistical significance of an interaction term between ART initiation and time. We a priori set alpha = 0.1 for this test.

We controlled for confounding and potentially informative loss-to-clinic in our estimates of effect in the CNICS with weights that are the product of: (i) stabilized inverse probability of treatment weights;[20,21]; and (ii) stabilized inverse probability of censoring weights.[22] We standardized estimates to the target population by applying weights that are the product of the previous two weights and scaled inverse probability of sampling weights.[9,10]

Let $\overline{L}_k$ be the history of measured time-varying confounders through month k, and V be the vector of time-fixed confounders measured at baseline. Finally, let C_k be an indicator of censoring in month k and $\overline{C}_{(-1)} = 0$ by definition. Inverse probability of treatment weights were then defined:

$$w_j^A = \frac{\prod_{k=0}^{j} f\left(A_k|\overline{A}_{(k-1)}, \overline{C}_k = 0\right)}{\prod_{k=0}^{j} f\left(A_k|\overline{A}_{(k-1)}, \overline{L}_k, V, \overline{C}_k = 0\right)}$$

where $f\left(A_k|\overline{A}_{(k-1)}, \overline{L}_k, V, \overline{C}_k = 0\right)$ is, by definition the conditional probability mass function $f_{\left(A_k|\overline{A}_{(k-1)}, \overline{L}_k, V, \overline{C}_k = 0\right)}\left(a_k|\overline{a}_{(k-1)}, \overline{l}_k, v, \overline{c}_k = 0\right)$ with $\left(a_k, \overline{a}_{(k-1)}, \overline{l}_k, v\right)$ evaluated at the random argument $\left(A_k, \overline{A}_{(k-1)}, \overline{L}_k, V\right)$. Similarly, inverse probability of censoring weights were defined:

$$w_j^C = \begin{cases} \dfrac{\prod_{k=0}^{j} f\left(C_k = 0|\overline{C}_{(k-1)} = 0, \overline{A}_{(k-1)}\right)}{\prod_{k=0}^{j} f\left(C_k = 0|\overline{C}_{(k-1)} = 0, \overline{A}_{(k-1)}, \overline{L}_k, V\right)}, & C_j = 0 \\ 0, & C_j = 1 \end{cases}$$

We estimated treatment and censoring weights using pooled logistic models.

Time-fixed covariates included in the models for the denominator of the treatment and censoring weights (the vector V) were: race/ethnicity; sex; age; calendar year; CD4 cell count and viral load most proximate to CNICS enrollment, measured up to 6 months before CNICS enrolment; previous AIDS diagnosis; history of injection drug use (IDU); history of male-to-male sexual contact (MSM); and study site. We combined race and ethnicity into one variable, classifying patients as Hispanic or Latino if indicated, regardless of race. Non-Hispanic patients were classified as Black, White or other race.

Time-varying covariates included in the models for the denominator of the treatment and censoring weights (the vector L) were: CD4 cell count, detectable viral load, AIDS diagnosis,[23,24] hepatitis C virus infection and interactions between CD4 cell count and detectable viral load. Time-varying covariates were updated whenever a patient was seen, with intervals determined by medical providers and by patients' care-seeking behaviour [median number of months between measurements = 3, interquartile range (IQR): 2, 4]. Laboratory values were carried forward in time from the most recent observed value until new values were reported. All continuous variables were modelled using restricted quadratic splines with knots at the 5th, 35th, 65th and 95th percentiles.[25]

When generalizability assumptions are met, inverse probability of sampling weights re-weight the study sample to have the same distribution of effect-measure modifiers as the target population. The inverse probability of sampling weights were defined:

$$w^S = \begin{cases} \dfrac{P(S=1)}{P(S=1|G)}, & S = 1 \\ 0, & S = 0 \end{cases}$$

where S is an indicator of inclusion in the CNICS study sample and G is a vector of time-fixed covariates measured in both the study sample and the target population. The covariates in G should include all covariates that: (i) predict membership in the study sample; and (ii) modify the effect of interest. The vector of covariates, G, included race/ethnicity, sex, age group and indicators for MSM, IDU and AIDS diagnosis at baseline. We additionally included all second-order interactions between these covariates. CD4 cell count and HIV RNA viral load were expected a priori to be strong modifiers of the effect of ART on mortality. However, not all HIV-infected persons have laboratory tests drawn immediately proximate to their diagnosis, and not all states currently collect prognostic laboratory results as part of their HIV surveillance. Because we did not have CD4 cell count and viral load measured on persons in the target population, our main analysis assumed that the distribution of these two variables match their distribution in the CNICS sample at baseline. In a sensitivity analysis, we imputed CD4 cell count and viral load values for members of the target population under varying assumptions about how much the true distribution of these variables in the target population differed from the CNICS sample. We estimated the inverse probability of sampling weights using logistic regression on combined data from the CNICS and the target population. In the model for sampling weights, we assumed that the specific set of individuals diagnosed with HIV in the USA between 2009 and 2011, and reported to the CDC, represented one possible random realization of the target population,

arising from a super-population of all persons who could have been diagnosed with HIV in the USA in that time period. To account for this in our analysis, we weighted persons in the target population by $1/[m/(N-n)]$, where n is the size of the study sample, m is the size of our target population and N is the size of the hypothetical super-population, which is arbitrarily large.[26] We set N to be 1.1 million; the choice of N did not influence our results. Persons in the CNICS received a weight of 1.

When the inverse probability weights for the final structural model exceeded 40, we explored the potential influence of large weights on our estimates by modifying the functional form of covariates in our weight models (e.g. using quadratic or cubic functions instead of splines) and by truncating (e.g. interval censoring or winsorizing) weights at the 0.1st and 99.9th percentile of the distribution of the weights, and observing the change in the final estimates.

Finally, to check for the existence of effect measure modification (a prerequisite for a lack of generalizability), we examined subgroup-specific effects of ART on mortality by stratifying the entire analysis according each of the covariates a priori expected to be associated with the outcome (those listed as being included in the vector G above, as well as CD4 cell count and viral load at baseline).

We calculated 95% confidence intervals (CI) using a standard error estimated by the standard deviation from 200 non-parametric bootstrap random samples drawn from the study sample and the target population with replacement.[27] All analyses were carried out using SAS 9.3 (Cary, NC).

Results

Overall, 12 457 patients in the CNICS met the inclusion criteria; most were male and of White race (Table 1). The median age at CNICS initiation was 38 years. The median CD4 cell count and viral load were 304 cells/μl and 46 276 copies/ml, respectively. Nearly one-fifth of CNICS patients had a history of an AIDS-defining illness at enrolment (18.7%). Injection drug use (18.7%) and hepatitis C virus co-infection were also common (14.0%).

Patients were followed for a median of 32 months [IQR: 17, 60]. During 437 892 person-months of follow-up, 8703 patients (69%) initiated ART, 5390 patients (43%) were lost to clinic and 918 patients died. Overall, 5-year mortality in the cohort was 11.3% (95% CI: 10.5%, 12.0%).

The median number of months to ART initiation was 4 (IQR: 1, 16). Strong predictors of ART initiation included transmission risk factor (IDU, MSM), calendar time of CNICS enrolment, race/ethnicity, CD4 cell count at baseline, hepatitis C infection, AIDS diagnosis and time on study. Strong predictors of loss to clinic included ART initiation, male sex, MSM transmission risk, calendar time of CNICS enrolment, CD4 cell count (at baseline and time-updated), hepatitis C infection, AIDS diagnosis, detectable viral load and time on study. All covariates listed in the Methods section were used to estimate treatment and censoring weights regardless of their predictive ability. Treatment and censoring weights had a mean of 0.98 (range: 0.06, 144.11) and 1.00 (range: 0.32, 12.27), respectively. Using different functional forms of covariates and truncating weights at 0.1st and 99.9th percentiles (0.04 and 18.66, respectively) yielded similar results for the effect of ART on survival.

Figure 1 shows curves for the cumulative incidence of death over 5 years of follow-up for the CNICS cohort and

Table 1. Characteristics of persons enrolled in the Center for AIDS Research Network of Integrated Clinical Systems (CNICS) during 1998–2012, and persons diagnosed with HIV in the USA during 2009–11

	CNICS study sample at enrolment	Recently HIV-diagnosed persons in the USA, 2009–11
Total N	12 457	12 8945
Sex		
Male	10 265 (82)	100 819 (78)
Female	2282 (18)	28 126 (22)
Age		
Median age (IQR)	38 (31, 45)	
18–24 years	980 (8)	25 535 (20)
25–34 years	3731 (30)	35 625 (28)
35–44 years	4766 (38)	31 153 (24)
45–54 years	2471 (20)	25 030 (19)
$\geq$ 55 years	599 (5)	11 602 (9)
Race/ethnicity		
White	5539 (44)	36 635 (28)
Black	4789 (38)	60 516 (47)
Hispanic	1635 (13)	26 079 (20)
Other	584 (5)	5715 (4)
History of AIDS at baseline	2343 (19)	32 896 (26)
HIV risk category		
History of injection drug use only	1440 (12)	6324 (5)
Male-to-male sexual contact only	6606 (53)	77 802 (60)
Injection drug use and male-to-male sexual contact	903 (7)	4105 (3)
CD4 cell count		
Median CD4 cell count (IQR)	304 (111, 488)	
$\leq$ 200 cells/μl	4481 (36)	
201–350 cells/μl	2650 (21)	
351–500 cells/μl	2432 (19)	
>500 cells/μl	2984 (24)	
Median $\log_{10}$ virus copies/ml	4.7 (4.1, 5.2)	

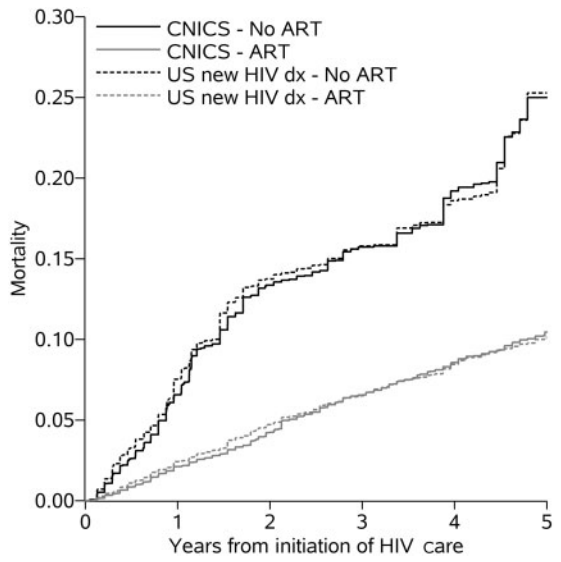

Figure 1. 5-year all-cause mortality under two potential interventions: always treat versus never treat with three or more antiretroviral medications (ART) among: (i) persons enrolled in the Center for AIDS Research Network of Integrated Clinical Systems (CNICS) during 1998–2012; and (ii) persons diagnosed with HIV in the United States during 2009–11. All estimates were adjusted for: race/ethnicity; sex; age at engagement in care; calendar year of engagement in care; CD4 cell count and viral load most proximate to CNICS enrolment; history of injection drug use; history of male-to-male sexual contact; study site; and time-varying CD4 cell count, viral load, AIDS diagnosis and hepatitis C virus infection. Estimates for persons newly diagnosed with HIV in the USA were further standardized to the distribution of sex, age group, race/ethnicity, male-to-male sexual contact, injection drug use and AIDS at baseline in the target populationdx, diagnosis.

target population if ART initiation were immediate versus delayed. In the CNICS cohort, the adjusted risk difference (RD) in mortality due to ART was −17.7% (95% CI: −27.0%, −8.4%; Table 3). The hazard ratio (HR) was 0.33 (95% CI: 0.25, 0.43). The proportional hazards assumption appeared reasonable (*P*-value for interaction between ART initiation and time = 0.6). The effect of ART on survival varied across subgroups of patients, although most subgroup effects were not statistically significantly different. RD due to ART use was −14.4% (95% CI: −23.2%, −5.7%) among patients with no reported IDU, compared with −18.1% (95% CI: −31.8%, −4.4%) among patients with a history of IDU. ART was strongly protective against 5-year mortality for patients with lower CD4 cell counts at baseline and less protective as baseline CD4 increased [RD = −29.3% (95% CI: −43.4%, −15.2%) for those with baseline CD4 ≤ 200 cells/mm^3 versus RD = −2.5% (95% CI: −6.0%, 1.0%) for those with baseline CD4 > 500 cells/mm^3]. ART was also more strongly protective for patients with a previous AIDS diagnosis [RD = − 22.8% (95% CI: −38.3%, −7.3%)] than for those with no previous AIDS diagnosis [RD = −13.4% (95% CI: −21.3%, −5.5%)]. Finally, among White patients, the RD for AIDS initiation was −17.3% (95% CI: −28.9%, −5.8%), compared with −11.3% (95% CI: −20.4%, −2.2%) among Black patients and −10.1% (95% CI: −21.8%, 1.6%) among Hispanic or Latino patients (Table 2).

Compared with CNICS patients, more people in the target population were women, Hispanic or Latino, or Black. The target population had more younger and older people compared with the age distribution of CNICS patients at enrolment. More CNICS patients were diagnosed with AIDS at baseline and fewer had a history of IDU (Table 1). All covariates were strong predictors of inclusion in the CNICS. Inverse probability of sampling weights had a mean of 0.98 (range: 0.17, 14.36).

The estimated RD due to immediate versus delayed ART initiation that we would have expected among recently HIV-diagnosed persons in the USA was −19.1% (95% CI: −30.5%, −7.8%) and the HR was 0.32 (95% CI: 0.23, 0.45) (Table 3). These estimates both assume that the distributions of CD4 cell count and viral load in the target population are the same as the distributions of these variables in the study sample. If, instead, the target population were substantially healthier at HIV diagnosis than the CNICS patients at entry to care (i.e. if the average CD4 cell count in the target population was 200 cells/μl higher and, among people with 'high' viral loads, the viral load was on average 10 000 copies/ml lower in the target population than in the CNICS patients), the HR for the effect of immediate versus delayed ART initiation among recently HIV-diagnosed persons in the USA would have been 0.45 (95% CI: 0.28, 0.71). Other scenarios are presented in Table 4.

Discussion

ART initiation substantxially decreased mortality over 5 years of follow-up among patients in the CNICS. The effect of ART on mortality was heterogeneous across subgroups of patients defined by patient characteristics that were also predictors of inclusion in the CNICS. Despite the effect heterogeneity and differences between the study sample and the target population, the expected magnitude of the survival benefits of ART among recently HIV-diagnosed persons in the USA were similar to those seen in the CNICS.

The HR for death due to ART estimated in the CNICS (HR = 0.33, 95% CI: 0.25, 0.43) was higher than the HR for AIDS or death in the Swiss HIV Cohort Study (HR = 0.14, 95% CI: 0.07, 0.29)[12] but lower than the HR for AIDS or death in the MACS/WIHS (HR = 0.54, 95% CI: 0.38, 0.78)[13] and lower than the HR for death in the HIV-CAUSAL Collaboration (HR = 0.48, 95% CI: 0.41, 0.57).[1] Aside from differences in the source populations

Table 2. Modification of the effect of three or more antiretroviral medications (ART) on all-cause mortality for 12 547 HIV-positive patients receiving care at CNICS sites, 1998–2011

	ART-exposed, 5-year mortality risk (%)	ART-unexposed, 5-year mortality risk (%)	Risk difference, %, (95% confidence interval)	Risk ratio (95% confidence interval)	Hazard ratio (95% confidence interval)
Overall	10.6 (9.3, 11.9)	28.3 (19.1, 37.5)	−17.7 (−27.0, −8.4)	0.37 (0.26, 0.53)	0.33 (0.25, 0.43)
Sex					
Male	10.1 (8.7, 11.5)	20.6 (15.2, 26.0)	−10.5 (−16.1, −4.8)	0.49 (0.36, 0.67)	0.35 (0.27, 0.45)
Female	11.8 (9.7, 14.0)	29.1 (13.1, 45.1)	−17.3 (−33.5, −1.1)	0.41 (0.23, 0.71)	0.41 (0.26, 0.64)
Age at CNICS start					
18–24 years	2.6 (0.7, 4.5)	1.5 (0.0, 9.4)	1.2 (−7.0, 9.4)	1.79 (0.27, 12.03)	1.77 (0.00, 1140.12)
25–34 years	6.7 (5.3, 8.2)	22.1 (10.2, 34.0)	−15.3 (−27.4, −3.3)	0.31 (0.16, 0.58)	0.38 (0.21, 0.66)
35–44 years	10.9 (8.5, 13.2)	18.1 (12.4, 23.8)	−7.2 (−13.2, −1.2)	0.60 (0.41, 0.87)	0.43 (0.32, 0.57)
45–54 years	15.6 (12.3, 19.0)	44.7 (25.1, 64.2)	−29.0 (−48.8, −9.2)	0.35 (0.21, 0.59)	0.24 (0.16, 0.37)
≥ 55 years	12.2 (15.1, 27.3)	33.5 (17.2, 49.8)	−12.3 (−30.0, 5.4)	0.63 (0.35, 1.14)	0.38 (0.18, 0.78)
Race/ethnicity					
White	8.2 (6.6, 9.8)	25.5 (14.2, 36.9)	−17.3 (−28.9, −5.8)	0.32 (0.19, 0.53)	0.26 (0.18, 0.38)
Black	14.4 (11.9, 16.9)	25.7 (16.7, 34.6)	−11.3 (−20.4, −2.2)	0.56 (0.38, 0.82)	0.44 (0.33, 0.58)
Hispanic	6.1 (4.3, 7.9)	16.2 (4.5, 27.7)	−10.1 (−21.8, 1.6)	0.38 (0.15, 0.92)	0.37 (0.19, 0.72)
Other	11.5 (4.9, 18.1)	22.9 (1.7, 44.1)	−11.4 (−33.8, 10.9)	0.50 (0.15, 1.62)	0.74 (0.35, 1.57)
Baseline CD4 count					
≤ 200 cells/mm^3	17.2 (15.6, 18.9)	46.5 (32.7, 60.4)	−29.3 (−43.4, −15.2)	0.37 (0.27, 0.52)	0.26 (0.21, 0.33)
201–350 cells/mm^3	7.6 (5.6, 9.5)	17.7 (5.1, 30.2)	−10.1 (−22.8, 2.5)	0.43 (0.21, 0.89)	0.42 (0.26, 0.68)
351–500 cells/mm^3	5.5 (2.0, 6.8)	12.6 (0.5, 24.7)	−7.1 (−19.4, 5.1)	0.44 (0.14, 1.32)	0.51 (0.26, 0.99)
> 500 cells/mm^3	4.4 (2.0, 6.8)	6.9 (4.6, 9.2)	−2.5 (−6.0, 1.0)	0.64 (0.32, 1.28)	0.71 (0.37, 1.37)
Baseline viral load					
< 10 000 copies/ml	5.2 (3.2, 7.2)	10.6 (5.5, 15.7)	−5.4 (−10.8, −0.1)	0.49 (0.26, 0.90)	0.55 (0.31, 0.98)
10 000–99 999 copies/ml	9.6 (7.8, 11.4)	23.7 (12.2, 35.3)	−14.1 (−26.2, −2.1)	0.40 (0.22, 0.73)	0.50 (0.35, 0.72)
≥ 100 000 copies/ml	13.8 (12.1, 15.6)	42.5 (26.1, 58.9)	−28.7 (−45.3, −12.1)	0.35 (0.21, 0.51)	0.20 (0.15, 0.26)
AIDS at baseline					
Yes	20.7 (17.9, 23.5)	47.5 (27.9, 59.1)	−22.8 (−38.3, −7.3)	0.48 (0.32, 0.70)	0.39 (0.30, 0.54)
No	7.6 (6.4, 8.7)	20.9 (13.2, 28.7)	−13.4 (−21.3, −5.5)	0.36 (0.24, 0.54)	0.28 (0.20, 0.40)
Injection drug use					
Yes	20.4 (15.6, 25.2)	38.5 (25.4, 51.6)	−18.1 (−31.8, −4.4)	0.53 (0.35, 0.81)	0.49 (0.34, 0.72)
No	7.8 (7.0, 8.7)	22.3 (13.7, 30.9)	−14.4 (−23.2, −5.7)	0.35 (0.23, 0.53)	0.30 (0.23, 0.40)
MSM					
Yes	7.0 (5.6, 8.4)	15.0 (9.3, 20.7)	−8.0 (−13.9, −2.0)	0.47 (0.30, 0.72)	0.32 (0.23, 0.45)
No	15.1 (12.7, 17.4)	32.1 (22.1, 42.1)	−17.1 (−27.4, −6.7)	0.47 (0.33, 0.66)	0.39 (0.29, 0.51)

All estimates were adjusted for: race/ethnicity; sex; age at engagement in care; calendar year of engagement in care; CD4 cell count and viral load most proximate to CNICS enrolment; history of injection drug use; history of male-to-male sexual contact; study site; and time-varying CD4 cell count, viral load, AIDS diagnosis and hepatitis C virus infection.

that contributed to each of these studies, differences in inclusion/exclusion criteria and differences in end points (AIDS or death versus death only), these HRs are difficult to compare because the period of follow-up is different for each study, and thus any violations of the proportional hazards assumption will result in a time-averaged summary HR that may be dependent on the length of follow-up.[28] We have improved on previous studies estimating the effect of ART by presenting cumulative incidence curves and RDs, in addition to HRs, which are arguably more useful to policy makers for planning.[29,30] Furthermore, we (and the most recent study)[1] estimated the effect of ART on death alone, as opposed to the effect of ART on time to AIDS or death. In the ART era, the risk of mortality following AIDS diagnosis is significantly reduced,[31] making death a more relevant clinical outcome. We excluded patients who were not ART naïve, consistent with current treatment standards.[32] Finally, we included people with an AIDS diagnosis at baseline, in line with the reality that many patients are diagnosed with AIDS and HIV almost concurrently and will not have the opportunity to start ART before they are diagnosed with AIDS.[14]

For our estimate of the effect of ART on mortality in the CNICS we assume no unmeasured confounding, positivity, treatment variation irrelevance or measurement error, and correct specification of our models for ART

Table 3. Effect of three or more antiretroviral medications (ART) on all-cause mortality for: (i) persons enrolled in the Center for AIDS Research Network of Integrated Clinical Systems (CNICS) during 1998–2012; and (ii) persons diagnosed with HIV in the USA during 2009–11

	5-year mortality risk % (95% confidence interval)	Risk difference, % (95% confidence interval)	Risk ratio (95% confidence interval)	Hazard ratio (95% confidence interval)
CNICS sample				
Crude				
No ART	12.1 (10.0, 14.3)	0.	1.	1.
ART	11.3 (10.4, 12.1)	−0.9 (−3.2, 1.5)	0.93 (0.76, 1.13)	1.01 (0.87, 1.17)
Weighted				
No ART	28.3 (19.1, 37.5)	0.	1.	1.
ART	10.6 (9.3, 11.9)	−17.7 (−27.0, −8.4)	0.37 (0.26, 0.53)	0.33 (0.25, 0.43)
HIV-diagnosed, USA				
No ART	29.5 (18.2, 40.8)	0.	1.	1.
ART	10.4 (9.2, 11.6)	−19.1 (−30.5, −7.8)	0.35 (0.23, 0.53)	0.32 (0.23, 0.45)

All estimates were adjusted for: race/ethnicity; sex; age at engagement in care; calendar year of engagement in care; CD4 cell count and viral load most proximate to CNICS enrolment; history of injection drug use; history of male-to-male sexual contact; study site; and time-varying CD4 cell count, viral load, AIDS diagnosis and hepatitis C virus infection. Estimates for persons newly diagnosed with HIV in the USA were further standardized to the distribution of sex, age group, race/ethnicity, male-to-male sexual contact, injection drug use and AIDS at baseline in the target population.

Table 4. Sensitivity analysis examining hazard ratios for three or more antiretroviral medications (ART) on all-cause mortality among persons diagnosed with HIV in the USA during 2009–11 (target population), assuming different distributions of CD4 cell count and viral load in the target (unmeasured) as compared with the CNICS study sample

CD4 cell count in target is on average:	Viral load in the target is on average:		
	10 000 lower than the sample for patients with a predicted viral load > 100 000	Equal to sample	10 000 higher than the sample
200 cells/ml higher than the sample	0.45 (0.28, 0.71)	0.38 (0.23, 0.63)	0.40 (0.25, 0.65)
100 cells/ml higher than the sample	0.40 (0.27, 0.60)	0.35 (0.24, 0.50)	0.35 (0.24, 0.53)
50 cells/ml higher than the sample	0.37 (0.25, 0.55)	0.35 (0.25, 0.49)	0.34 (0.23, 0.49)
Equal to sample	0.35 (0.24, 0.52)	0.34 (0.24, 0.47)	0.32 (0.22, 0.46)
50 cells/ml lower than the sample	0.33 (0.23, 0.48)	0.32 (0.23, 0.45)	0.30 (0.21, 0.43)
100 cells/ml lower than the sample	0.31 (0.22, 0.45)	0.30 (0.21, 0.42)	0.29 (0.20, 0.41)
200 cells/ml lower than the sample	0.34 (0.23, 0.50)	0.33 (0.22, 0.50)	0.32 (0.21, 0.47)

initiation and loss to clinic. As with all observational studies, we cannot exclude the possibility of unmeasured confounding, but we have adjusted for what we believe to be the strongest confounders of the effect of ART on mortality, and our adjustment set includes all strong confounders that have been included in other studies of the effect of ART. We had sufficient positivity for our set of confounders, as evidenced by the fact that our weights were generally well behaved, and which we verified by inspection of the data. Although ART regimens have changed from 1998 to 2011, we chose to start follow-up in 1998 because it marked the beginning of the highly effective ART era and we felt that differences among highly effective ART regimens were minor in comparison with differences between treated and untreated patients. If ART regimens have improved substantially, our approach will underestimate the effect of ART for the target population (assuming newer regimens are more efficacious with less potential for toxicity). We believe measurement error is unlikely to have had a substantive impact on our results. The covariates we used in our analysis were all collected for clinical purposes and represent the information available to the physicians as they were deciding when to initiate treatment; as such, even if laboratory values did not reflect true biological values of CD4 cell count or viral load, it was the observed, mis-measured value that actually influenced treatment and should have been included in our model. Having IDU as a risk factor may not be a perfect proxy for current drug use, but drug use is a fairly stable behaviour[33] and ever drug use may be as or more important than current use. The

date of ART initiation is a critical clinical milestone and we think it is likely to have been recorded with little error. Finally, deaths were actively ascertained through clinic sources and by matching with the Social Security Death Index. It is therefore unlikely that we have so misclassified vital status on a sufficient number of patients that it would have madee a difference in our final estimates.

We estimated the effect of ART under the interventions, initiating ART immediately or delaying ART initiation at least 5 years. Given that many HIV-infected persons experience at least some delay linking to care following HIV diagnosis,[34,35] delayed ART initiation might not be a completely unrealistic counterfactual exposure distribution for this target population. Even if the average newly HIV-diagnosed individual does not delay ART initiation a full 5 years, a substantively interesting risk difference could be read off the cumulative incidence curves in Figure 1; for example, at 8 months of follow-up, the median time between HIV diagnosis and entry to care among a cohort of newly HIV-diagnosed persons in Philadelphia,[34] the risk difference comparing ART initiated with uninitiated was approximately 3%. Future work (e.g. using the parametric g-formula[36] or alternative weighting strategies) could provide generalized impact estimates[37] that also account for treatment heterogeneity and the distribution of covariates in the target population (note here, that 'generalized' implies estimating effects of contrasts other than always treated versus never treated).

We were interested in estimating the effect of ART on all-cause mortality among recently HIV-diagnosed persons in the USA. For an estimate from a study sample to directly generalize to a specific target population in expectation, a sufficient set of assumptions include: (i) the study sample is a random sample of the target population or covariates associated with selection into the study sample are not also associated with the outcome;[10,38] (ii) no interference;[39] and (iii) similar versions of treatment or treatment variation irrelevance.[11] A more detailed discussion of sufficient assumptions for generalizability appears in the Appendix (available as Supplementary data at *IJE* online). We believe the latter two assumptions to be plausible. We are aware of no evidence that one person's ART initiation would affect another (already infected) person's mortality risk. Additionally, CNICS patients receive a standard of care that would be expected to be fairly similar to care provided by non-CNICS clinics, including the levels of adherence counselling and supportive services. It is potential violations of the first assumption that we have addressed in this analysis. The CNICS cohort is similar to our target population on many structural factors, including the health care delivery system and social context. However, CNICS patients' characteristics are not identical to the characteristics of persons in our target population. We do not, as yet, have a good understanding of the degree to which non-representativeness and non-significant departures from effect homogeneity across multiple subgroups may interact to produce changes in the final standardized estimate for the target population. Therefore, rather than relegating considerations of external validity of our results to a thought exercise in the discussion of this manuscript, we apply a formal correction for non-random sampling into the study sample (inverse probability of sampling weights). Our formal assessment of generalizability provides confidence in the generalizability of research in the CNICS that uses mortality as an outcome.

When generalizing our estimate, we may not have controlled for all causes of selection and of the outcome. CD4 cell count and viral load at baseline were associated with the risk of all-cause mortality, but neither were available in the national surveillance data for all persons at the time they were HIV-diagnosed. There is some evidence to suggest that the average baseline CD4 is probably slightly higher in the target population than in the CNICS. Among a non-random subset of persons newly diagnosed with HIV who had a CD4 cell count measured within 3 months, the proportion with a CD4 cell count ≥ 500 cells/μl was 29%, whereas the proportion with a CD4 cell count < 200 cells/μl was 33%;[14] in the CNICS, those proportions were 24% and 36%, respectively. Failing to account for differences in the distribution of CD4 cell count and viral load at baseline between the CNICS and target population may be a source of residual bias in our generalized effect estimate.[47] However, sensitivity analyses indicated that even moderate shifts in the average CD4 or viral load of patients in the target population as compared with CNICS failed to appreciably alter the estimated effect of ART. Treatment versions may differ between the CNICS and the target population, as CNICS patients are all treated in academic medical centres, which may influence patient adherence and quality of care.[41] There was a positive probability of inclusion in the CNICS in nearly all strata of covariates, owing to the diversity of the CNICS cohort, which was reflected in the stability of the inverse probability of sampling weights. Finally, generalizability may be threatened in the presence of interference;[39] however, interference is likely negligible for this exposure-outcome relationship.

In this study, we showed that ART reduces mortality in a cohort that is geographically and clinically representative of persons recently diagnosed with HIV in the USA. The estimates obtained in the cohort and in our generalization to the target population of recently HIV-diagnosed persons in the USA were similar. Furthermore, by including persons with previous AIDS diagnosis at baseline, and by excluding people with previous exposure to dual- or monotherapy, we

have estimated an effect that is pertinent to current standards of clinical care. Although we observed heterogeneity in the effect of ART across sub-populations, and differences in the distribution of those sub-populations between the CNICS and the target population, this heterogeneity did not change the overall estimate of the effect of ART among persons recently diagnosed with HIV in the USA.

Supplementary Data

Supplementary data are available at *IJE* online.

Acknowledgements

The authors would like to thank the patients, principal investigators, co-investigators and research staff at participating Centers for AIDS Research Network of Integrated Clinical Studies sites.

Funding

The CNICS is supported by R24 AI067039. C.L., S.R.C. and DW were supported in part by National Institute of Health grant R01 AI100654. D.W. was also supported in part by National Institute of Health grant DP2 HD084070. S.R.C., D.W. and J.J.E. were supported in part by the University of North Carolina at Chapel Hill Center for AIDS Research (CFAR) and by NIH-funded program P30 AI50410. M.J.M. was supported in part by National Institute of Health grant R01 AI103661. The content of this manuscript is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Conflict of interest: M.J.M. reports grants and personal fees from BMS, personal fees from Gilead, personal fees from Merck Foundation, grants from Pfizer and grants from Definicare, outside the submitted work; J.E. reports grants from National Institute of Health during the conduct of the study, grants and personal fees from Merck & Co., grants and personal fees from Bristol-Myers Squibb, grants and personal fees from GlaxoSmithKline/ViiV, personal fees from Gilead, personal fees from Janssen and personal fees from Abbvie, outside the submitted work. C.L. reports personal fees from Gilead outside the submitted work. D.W. engages in occasional ad hoc consulting on epidemiological methods with NIH/NICHD, outside the submitted work. No other authors have anything to declare.

References

1. HIV-CAUSAL Collaboration; Ray M, Logan R, Sterne JA *et al*. The effect of combined antiretroviral therapy on the overall mortality of HIV-infected individuals. *AIDS* 2010;**24:**123–37.
2. Hammer SM, Squires KE, Hughes MD *et al*. A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. AIDS Clinical Trials Group 320 Study Team. *N Engl J Med* 1997;**337:**725–33.
3. Cameron DW, Heath-Chiozzi M, Danner S *et al*. Randomised placebo-controlled trial of ritonavir in advanced HIV-1 disease. The Advanced HIV Disease Ritonavir Study Group. *Lancet* 1998;**351:**543–49.
4. Walensky RP, Paltiel AD, Losina E *et al*. The survival benefits of AIDS treatment in the USA. *J Infect Dis* 2006;**194:**11–19.
5. Cohen MS, Chen YQ, McCauley M *et al*. Prevention of HIV-1 infection with early antiretroviral therapy. *N Engl J Med* 2011;**365:**493–505.
6. Gandhi M, Ameli N, Bacchetti P *et al*. Eligibility criteria for HIV clinical trials and generalizability of results: the gap between published reports and study protocols. *AIDS* 2005;**19:**1885–96.
7. Weiss NS, Koepsell TD, Psaty BM. Generalizability of the results of randomized trials. *Arch Intern Med* 2008;**168:**133–35.
8. Bareinboim E, Pearl J. A general algorithm for deciding transportability of experimental results. *Journal of Causal Inference* 2013;**1:**107–34.
9. Cole SR, Stuart EA. Generalizing evidence from randomized clinical trials to target populations: The ACTG 320 trial. *Am J Epidemiol* 2010;**172:**107–15.
10. Stuart EA, Cole SR, Bradshaw CP, Leaf PJ. The use of propensity scores to assess the generalizability of results from randomized trials. *J R Stat Soc A* 2011;**174:**369–86.
11. Hernán MA, VanderWeele TJ. Compound treatments and transportability of causal inference. *Epidemiology* 2011;**22:**368–77.
12. Sterne JA, Hernán MA, Ledergerber B *et al*. Long-term effectiveness of potent antiretroviral therapy in preventing AIDS and death: a prospective cohort study. *Lancet* 2005;**366:**378–84.
13. Cole SR, Hernán MA, Robins JM *et al*. Effect of highly active antiretroviral therapy on time to acquired immunodeficiency syndrome or death using marginal structural models. *Am J Epidemiol* 2003;**158:**687–94.
14. Centers for Disease Control and Prevention. *Monitoring Selected National HIV Prevention and Care Objectives by Using HIV Surveillance Data – USA and Six Dependent Areas– 2012. HIV Surveillance Supplemental Report*. Dekalb County, GA: Centers for Disease Control and Prevention, 2014.
15. Kitahata MM, Rodriguez B, Haubrich R *et al*. Cohort Profile: The Centers for AIDS Research Network of Integrated Clinical Systems. *Int J Epidemiol* 2008;**37:**948–55.
16. Centers for Disease Control and Prevention. *HIV Surveillance Report, 2011*. Dekalb County, GA: Centers for Disease Control and Prevention, 2013.
17. Cain LE, Robins JM, Lanoy E, Logan R, Costagliola D, Hernan MA. When to start treatment? A systematic approach to the comparison of dynamic regimes using observational data. *Int J Biostat* 2010;**6:**Article 18.
18. Kaplan EL, Meier P. Nonparametric estimation from incomplete observations. *J Am Stat Assoc* 1958;**53:**457–81.
19. Efron B. The efficiency of Cox's likelihood function for censored data. *J Am Stat Assoc* 1977;**72:**557–65.
20. Cole SR, Hernán MA. Constructing Inverse Probability Weights for Marginal Structural Models. *Am J Epidemiol* 2008;**168:**656–64.
21. Cole SR, Hernán MA. Adjusted survival curves with inverse probability weights. *Comput Methods Programs Biomed* 2004;**75:**45–49.
22. Robins JM, Finkelstein DM. Correcting for noncompliance and dependent censoring in an AIDS Clinical Trial with inverse probability of censoring weighted (IPCW) log-rank tests. *Biometrics* 2000;**56:**779–88.
23. Centers for Disease Control and Prevention. 1993 revised classification system for HIV infection and expanded surveillance case

definition for AIDS among adolescents and adults. *JAMA* 1993;**269:**729–30.
24. Centers for Disease Control and Prevention. Revised surveillance case definition for HIV infection – USA, 2014. *MMWR* 2014;**63:**1–10.
25. Howe CJ, Cole SR, Westreich DJ, Greenland S, Napravnik S, Eron JJ Jr. Splines for trend analysis and continuous confounder control. *Epidemiology* 2011;**22:**874–75.
26. Buchanan AL, Hudgens MG, Cole SR *et al*. Generalizing Evidence from Randomized Trials using Inverse Probability of Sampling Weights. Technical Report Series 2015.
27. Efron B, Tibshirani R. *An Introduction to the Bootstrap*. New York, NY: Chapman & Hall, 1993.
28. Hernán MA. The hazards of hazard ratios. *Epidemiology* 2010;**21:**13–15.
29. Greenland S. Interpretation and choice of effect measures in epidemiologic analyses. *Am J Epidemiol* 1987;**125:**761–68.
30. Poole C. On the origin of risk relativism. *Epidemiology* 2010;**21**:3–9.
31. Grabar S, Lanoy E, Allavena C *et al*. Causes of the first AIDS-defining illness and subsequent survival before and after the advent of combined antiretroviral therapy. *HIV Med* 2008;**9:** 246–56.
32. Thompson MA, Aberg JA, Hoy JF *et al*. Antiretroviral treatment of adult HIV infection: 2012 recommendations of the International Antiviral Society-USA panel. *JAMA* 2012;**308:** 387–402.
33. Genberg BL, Gange SJ, Go VF, Celentano DD, Kirk GD, Mehta SH. Trajectories of injection drug use over 20 years (1988-2008) in Baltimore, Maryland. *Am J Epidemiol* 2011; **173:**829–36.
34. Bamford LP, Ehrenkranz PD, Eberhart MG, Shpaner M, Brady KA. Factors associated with delayed entry into primary HIV medical care after HIV diagnosis. *AIDS* 2010;**24:**928–30.
35. Torian LV, Wiewel EW, Liu KL, Sackoff JE, Frieden TR. Risk factors for delayed initiation of medical care after diagnosis of human immunodeficiency virus. *Arch Intern Med* 2008;**168:** 1181–87.
36. Westreich D, Cole SR, Young JG *et al*. The parametric g-formula to estimate the effect of highly active antiretroviral therapy on incident AIDS or death. *Stat Med* 2012;**31:**2000–09.
37. Westreich D. From exposures to population interventions: pregnancy and response to HIV therapy. *Am J Epidemiol* 2014;**179:**797–806.
38. Tipton E. Improving Generalizations From Experiments Using Propensity Score Subclassification: Assumptions, Properties, and Contexts. *J Educ Behav Stat* 2013;**38:**239–66.
39. Hudgens MG, Halloran ME. Toward Causal Inference With Interference. *J Am Stat Assoc* 2008;**103:**832–42.
40. Weisberg HI, Hayden VC, Pontes VP. Selection criteria and generalizability within the counterfactual framework: explaining the paradox of antidepressant-induced suicidality? *Clin Trials* 2009;**6:**109–18.
41. Frangakis C. The calibration of treatment effects from clinical trials to target populations. *Clin Trials* 2009;**6:**136–40.
42. Greenhouse JB, Kaizar EE, Kelleher K, Seltman H, Gardner W. Generalizing from clinical trial data: a case study. The risk of suicidality among pediatric antidepressant users. *Stat Med* 2008;**27:** 1801–13.
43. Hernán MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health* 2006;**60:** 578–86.
44. Westreich D, Cole SR. Invited commentary: positivity in practice. *Am J Epidemiol* 2010;**171:**674–77; discussion 78–81.
45. VanderWeele TJ. Concerning the consistency assumption in causal inference. *Epidemiology* 2009;**20:**880–83.
46. Hernán MA, Cole SR. Invited Commentary: Causal diagrams and measurement bias. *Am J Epidemiol* 2009;**170:**959–62; discussion 63–64.
47. Robins J, Hernán MA, Brumback B. Marginal Structural Models and Causal Inference in Epidemiology. *Epidemiology* 2000;**11:** 550–60.