DS

.1097/QAD.0000

ocol ef

tephen R. Cole

Hartwig Klinker (8), Haitao Chu

Hernán (1,10,11) on behalf of the

START) study group

han School of Public Health, Boston, MA,

Health, University of Minnesota, Minneapolis,

shospitalet, University of Copenhagen, Denmark;

Population Health, University College London,

lings School of Global Public Health, University of

NC, USA;

6. Infectious Disease Clinical Research Program, Department of Preventive Medicine and Biostatistics, Uniformed Services University of the Health Sciences and Henry M. Jackson Foundation for the Advancement of Military Medicine, USA;
7. Medical Research Council, Clinical Trials Unit in University College London, London, United Kingdom;
8. University of Wuerzburg Medical Center, Germany;
9. The Kirby Institute, Sidney, Australia;
10. Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA;
11. Harvard-MIT Division of Health Sciences and Technology; Boston, MA, USA

Correspondence: Sara Lodi, Department of Epidemiology, Harvard T.H. Chan School of Public Health, 677 Huntington Avenue, Boston, USA. Email: slodi@hsph.harvard.edu

Number of words: abstract (236), manuscript (1798)

Number of tables (1), number of appendices (5)

Funding sources: NIH R01 AI102634, PCORI ME-1503-28119, NIH Grants UM1-AI068641 and UM1-AI120197, Y1-AI-5072 and NIH R01 AI100654. The funding organizations had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript. The Views expressed are those of the author(s) and do not necessarily reflect the official views of the Uniformed Services University of the Health Sciences, the NIH, the Department of Defense, or the Departments of the Army, Navy or Air Force.

Conflicts of interest: AP, Consultancy with GSK Biologicals, Speaker fees for Gilead Sciences, Advisory Board for Abbvie. No other conflict of interest to report.

Abstract

Objective: The START trial found a lower risk of a composite clinical outcome in HIV-positive individuals assigned to immediate initiation of antiretroviral therapy (ART) compared with those assigned to deferred initiation. However, 30% of those assigned to deferred initiation started ART earlier than the protocol specified. To supplement the published intention-to-treat effect estimates, here we estimate the per-protocol effect of immediate versus deferred ART initiation in START.

Design: The START trial randomized 4685 HIV-positive participants with CD4 counts > 500 /mm^3 to start ART immediately after randomization (immediate initiation group) or to wait until the CD4 count dropped below 350 cells/mm^3 or an AIDS diagnosis (deferred initiation group).

Methods: We used the parametric g-formula to estimate and compare the cumulative 5-year risk of the composite clinical outcome in the immediate and deferred initiation groups had all the trial participants adhered to the protocol.

***Results*:** We estimated that the 5-year risk of the composite outcome would have been 3.2% under immediate ART initiation and 7.0% under deferred initiation. The difference of 3.8% (95% confidence interval 1.5,6.5) was larger than the intention-to-treat effect estimate of 3.1%, corresponding to a difference in effect estimates of 0.72% (-0.35,2.35).

***Conclusions*:** The intention-to-treat effect estimate may underestimate the benefit of immediate ART initiation by 23% . This estimate can be used by patients and policy makers who need to understand the full extent of the benefit of changes in ART initiation policies.

Keywords: per-protocol effect, g-formula, antiretroviral treatment, HIV.

Introduction

For almost 30 years after the introduction of antiretroviral therapy (ART), when to start treatment has been a key decision in the care of HIV-positive people. To address this question, the recent START (Strategic Timing of AntiRetroviral Treatment) randomized trial compared the effect of immediate ART initiation with deferred initiation until a confirmed CD4 count <350 cells/mm^3 or an AIDS diagnosis in individuals with CD4 count >500 cells/mm^3 at randomization. The primary outcome was a composite of any serious AIDS event, serious non-AIDS event, or death from any cause. An intention-to-treat (ITT) analysis estimated a 57% lower incidence of the primary outcome in the immediate treatment group than in the deferred treatment group (hazard ratio: 0.43, 95% confidence interval [0.30,0.62]) over an average of 3 years of follow-up [1].

While participant generally adhered well to the protocol, some deviations occurred. In particular, 30% of the participants randomized to the deferred arm started ART with a latest CD4 ≥350 cells/mm^3 [1]. Hence, the ITT effect estimate may have underestimated the benefits of immediate initiation [2]. In addition, a small proportion (4.3%) of people were lost to follow-up [1].

To complement the published ITT effect estimates in the START trial and to understand the potential underestimation of the benefit of immediate initiation resulting from premature treatment in the deferred group, we estimated the per-protocol effect. This is the effect had all participants in the trial adhered to the treatment initiation strategy they were assigned to and, unless they had experienced the primary endpoint, remained under follow-up for the duration of the study. These estimates will assist health care planners quantify the potential impact of immediate ART initiation for all HIV-positive individuals, which is now recommended by national and international guidelines [3-5].

Methods

The START trial has been described elsewhere[1]. The study randomized adult HIV-positive participants with two CD4 counts>500 cells/mm^3 to start ART immediately after randomization (immediate initiation group) or to wait until CD4 count dropped below 350 cells/mm^3 or an AIDS diagnosis (deferred initiation group). Overall median [IQR] time since HIV diagnosis was 1 year [0.4,3.1]. We estimated the 3-year and 5-year risks after randomization of the primary outcome that would have been observed in each group if all participants had fully adhered to the protocol. That is, if they had initiated ART at the assigned time and had remained under follow-up until diagnosis of the primary outcome, death, or the administrative end of follow-up (May 26 2015). As measures of per-protocol effect, we estimated the risk difference.

Estimating the per-protocol effect requires a precise definition of what constitutes a protocol deviation. The maximum time window (grace period) between eligibility for ART initiation and ART initiation was not explicitly specified in the protocol. Therefore we defined immediate initiation as starting ART within 1 month of randomization and deferred initiation as starting

ART within 1 month of a confirmed CD4<350 cells/mm^3. We conducted a sensitivity analysis using a grace period of 2 months. While the protocol did not mandate ART initiation at pregnancy, it was allowed and 49 women in the deferred group started ART because of pregnancy. We therefore conducted a sensitivity analysis in which failure to initiate ART because of pregnancy was considered a protocol deviation.

Follow-up began at randomization and ended at the first of: a primary endpoint, administrative end of follow up (at 60 months or 26 May 2015), or loss to follow-up (12 months after the latest CD4 or HIV-RNA measurement). The latter form of censoring was unnecessary in the published ITT analysis, which did not rely on post-randomization information on ART and prognostic factors.

Our estimates were adjusted for the following baseline and post-randomization variables defined a priori: age (<35, ≥35 years), CD4 count (≤650, >650 cells/mm^3), and HIV-RNA at randomization (<5000, ≥5000 copies/mL), sex, geographical area (high-income regions versus low-mid income regions), the square root of the latest CD4 value, and the natural logarithm of HIV-RNA value, the number of months since the last CD4 and HIV-RNA measurements, ART initiation status (never started ART versus initiated ART) and months since ART initiation.

To adjust for the above variables, we used the parametric g-formula [7, 8] which, in contrast to traditional methods, appropriately adjusts for baseline and post-randomization factors associated with ART initiation and the outcome. The estimation procedure has been described elsewhere [9]. The procedure has two steps. First, we fitted parametric regression models to estimate the joint distribution of the outcome, treatment, and time-varying covariates conditional on previous treatment and covariate history. Second, using the parameter estimates from these models, we

simulated a dataset of 100,000 individuals under each of the two per-protocol initiation strategies. Finally, we computed and compared the outcome risk at 3 and 5 years in the simulated data. The immediate and deferred initiation groups were analyzed separately because the predictors of ART initiation (hazard ratios of initiation estimated using pooled logistic regression models), and hence the confounding structure, varied between groups. We used a nonparametric bootstrap procedure based on 1000 samples to obtain percentile-based 95% confidence intervals (CIs). More specifically, for each bootstrap repetition, we estimated the regression models, repeated the simulation procedure and estimated the risk of the outcome.

To explore the goodness-of-fit of our parametric models, we simulated a dataset under the same degree of adherence to the assigned strategies that was observed in START, and compared the estimates with those from the observed data. For comparison with the published ITT analysis, in which the median follow-up was approximately 3 years, we estimated the 3-year average per-protocol hazard ratio by fitting Cox models to our simulated data. We compared this estimate with that from a naïve per-protocol analysis that censored participants when they deviated from the protocol without adjusting for any covariates and adjusting for baseline covariates only.

All analyses were conducted with the publicly available SAS macro GFORMULA (http://www.hsph.harvard.edu/causal/software/).

Results

Of the 4685 START participants, 9 were excluded from this analysis due to missing baseline HIV-RNA, resulting in a cohort of 4676. Median [IQR] follow-up was 34 [25,46] and 33 [25,45] months in the immediate and deferred initiation groups, respectively; 365 (8%) participants were defined as lost to follow-up due to there being at least a 12 month period in which no CD4 or HIV-RNA was measured. The distribution of participants' characteristics at randomization is shown in Appendix Table 1. During 13,831 person-years of follow-up, 38 participants experienced the primary outcome in the immediate initiation group and 90 in the deferred initiation group. Ten outcome events in the original paper were excluded because they occurred after the date the participants were lost to follow-up according to our definition (see methods).

Of 2,321 participants assigned to immediate initiation, 2061 and 2213 started ART by month 1 and 2 after randomization, respectively and 45 never started ART. Of 2,355 participants assigned to deferred initiation, 1045 initiated ART during follow-up and 650 started ART with a latest CD4 count ≥350 cells/mm^3. The predictors for ART initiation differed in the two initiation groups (Appendix Table 2). Predictors of initiation in the deferred initiation group were younger age, being in a high-income country, higher baseline and latest HIV-RNA, lower latest CD4 count and shorter time since the last CD4 count or HIV-RNA measurement.

The Table shows the ITT and per-protocol risk estimates. The estimated per-protocol 5-year risk (95% CI) was 3.2% (1.9,4.5) for immediate initiation and 7.0% (5.3,9.4) for deferred initiation. The 5-year risk difference was -3.8% (-6.5,-1.5), compared with an ITT difference of -3.1% (-5.3,-0.9), corresponding to a a difference between the per-protocol and ITT effect estimates of 0.72% (-0.35, 2.35). The per-protocol hazard ratio (95% CI) for immediate versus deferred

initiation was 0.34 (0.21,0.52). This was stronger than the originally reported ITT hazard ratio of 0.43 (0.30,0.62). The naïve per-protocol analysis included 109 events (37 and 72 in the immediate and deferred group). The total follow-up while participants adhered to the protocol was 6,516 (94%) and 5,375 (78%) person-years in the immediate and deferred initiation group, respectively. The estimated hazard ratio was 0.41 (0.28,0.61) with no adjustment for baseline covariates and 0.41 (0.27,0.61) with adjustment for baseline covariates, similar to the ITT hazard ratio.

The results did not materially change in the sensitivity analyses (Appendix Table 3). The time-varying means predicted by our models under the observed adherence were similar to the observed means in both initiation groups (Appendix Figure 1).

Discussion

Our per-protocol analysis suggests that the potential benefits of immediate initiation compared with the deferral strategy are about 20% larger than previously suggested by the ITT analysis, although the confidence intervals for our estimates were wide. Compared with deferred initiation until CD4 count dropped below 350 cells/mm^3, immediate initiation reduced the risk of the composite outcome encompassing serious AIDS, serious non-AIDS or death events by 2.7% at 3 years and by 3.8% at 5 years. This stronger protective effect of immediate initiation was anticipated because approximately 30% of participants in deferred ART group initiated ART with a CD4 above 350 cells/mm^3, while protocol deviations were uncommon in the immediate group.

The hazard ratio estimates from the naïve per-protocol analyses (with and without adjustment for baseline variables) were similar to that from the ITT analysis. These estimates underestimated

the per-protocol effect of immediate versus deferred initiation compared with our per-protocol analysis, which further adjusted for post-randomization variables via the parametric g-formula [2, 8, 10]. Our findings show that approaches based on the parametric g-formula can be used for the per-protocol analysis of clinical trials with protocol deviations.

Our analysis has limitations. First, our adjusted per-protocol analysis assumes that all prognostic factors that predict protocol deviations are identified and accurately measured. While this condition cannot be guaranteed, we adjusted for the main factors used in clinical treatment decisions. Second, all models need to be correctly specified. Again this condition cannot be guaranteed, but it seems plausible because our models resulted in simulated data sets with distributions of outcome and time-varying covariates similar to those in the original data. Finally, we could not obtain per-protocol effect estimates for each component of the primary endpoint (i.e., serious AIDS events, serious non-AIDS events and mortality) because the small number of the events led to unstable outcome models. Semiparametric methods, such as inverse-probability weighting, may be an alternative to estimate the per-protocol effect for those outcomes.

In conclusion, our estimates of the per-protocol effect of immediate versus deferred ART initiation provide additional support to recent changes in clinical recommendations. Per-protocol effect estimates are especially relevant to patients and clinicians, and can be used by modelers and health care planners to estimate an upper bound of the impact of changes in recommendations. While there are good reasons for ITT analyses to remain the primary analyses of many randomized trials, appropriately adjusted per-protocol analyses help extract additional important information from clinical trial data.

References

1. Insight Start Study Group. Initiation of Antiretroviral Therapy in Early Asymptomatic HIV Infection. *N Engl J Med* 2015.
2. Hernan MA, Hernandez-Diaz S. Beyond the intention-to-treat in comparative effectiveness research. *Clin Trials* 2012,**9**:48-55.
3. DHHS panel on antiretroviral guidelines for adults and adolescents. Guidelines for the use of antiretroviral agents in HIV-1-infected adults and adolescents. In. Edited by Department of Health and Human Services (DHHS); 2015.
4. European AIDS clinical society (EACS). European guidelines for treatment of HIV infected adults in Europe. In; 2015.
5. World Health Organization (WHO). Guideline on when to start antiretroviral therapy and on pre-exposure prophylaxis for HIV. In. Switzerland; 2015.
6. Royston P, Parmar MK. Restricted mean survival time: an alternative to the hazard ratio for the design and analysis of randomized trials with a time-to-event outcome. *BMC Med Res Methodol* 2013,**13**:152.
7. Robins J, Hernan M. Estimation of the causal effects of time-varying exposures. In: *Advances in longitudinal data analysis.* Edited by Fitzmaurice GD, M. Verbeke, G. Molenberghs, G. Boca Raton: Chapman and Hall/CRC Press; 2009. pp. 553-599.
8. Robins JM. A new approach to causal inference in mortality studies with a sustained exposure period: application to the healthy worker survivor effect. *Mathematical Modelling* 1986,**7**:1393-1512.

9. Young JG, Cain LE, Robins JM, O'Reilly EJ, Hernan MA. Comparative effectiveness of dynamic treatment regimes: an application of the parametric g-formula. *Statistics in biosciences* 2011,**3**:119-143.

10. Murray EJ, Hernan MA. Adherence adjustment in the Coronary Drug Project: A call for better per-protocol effect estimates in randomized trials. *Clin Trials* 2016.

Principal contributions made by the authors

Data collection: Shweta Sharma, Jens Lundgren, Andrew Phillips, Brian Agan, Abdel Babiker, Hartwig Klinker,, Haitao Chu, Matthew Law, James Neaton; Statistical analyses: Sara Lodi, Miguel Hernan, Roger Logan,; Interpretation of results: All authors; Read and approved the manuscript: All authors; Revised the work for important intellectual content: All authors; Drafted the manuscript: Sara Lodi, Miguel Hernan. Sara Lodi, the corresponding author, had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Table. Intention-to-treat (ITT) and per-protocol estimates of risk and risk difference, at 3 and 5 years after randomization, START trial 2009-2015.

Follow-up	Analysis	Risk, % (95% CI)		Risk difference of immediate vs deferred, % (95% CI)
		Immediate ART initiation (N= 2321 §)	Deferred ART initiation (N=2355 §)	
3 years	Intention-to-treat	1.5 (0.9,2.1)	3.9 (3.1,4.7)	-2.4 (-3.4,-1.4)
	Per-protocol*	1.5 (1.0,2.0)	4.1 (3.3,5.1)	-2.7 (-3.8,-1.7)
5 years	Intention-to-treat	3.2 (1.9,4.7)	6.2 (4.7,8.0)	-3.1 (-5.2,-0.8)
	Per-protocol*	3.2 (2.0,4.6)	7.0 (5.2,9.6)	-3.8 (-6.7,-1.5)

§ 9 participants in START were excluded from this analysis due to missing baseline HIV-RNA, 5 in the immediate ART initiation group and 4 in the deferred ART initiation group.

*Estimates under complete adherence to the protocol are adjusted, via the parametric g-formula, for sex, age, CD4 count and HIV-RNA at randomization, geographical area, the most recent CD4 count and HIV-RNA values, the number of months since the last CD4 and HIV-RNA measurements, ART initiation status and months since ART initiation.