Pharmacological Measures of Treatment Adherence and Risk of HIV Infection in the VOICE Study

James Y. Dai, Craig W. Hendrix, Barbra A. Richardson, Cliff Kelly, Mark Marzinke, Z. Mike Chirenje, Jeanne M. Marrazzo, and Elizabeth R. Brown

1Vaccine and Infectious Diseases Division, Fred Hutchinson Cancer Research Center, and 2University of Washington, Seattle; 3School of Medicine, Johns Hopkins University, Baltimore, Maryland; and 4University of Zimbabwe–University of California, San Francisco, Research Programme, Harare

(See the editorial commentary by Vermund and Walker on pages 329–31.)

Background. None of the 3 regimens tested in the VOICE study showed protection against human immunodeficiency virus (HIV) infection in intent-to-treat analyses. Plasma tenofovir concentrations demonstrated poor adherence to the study product among study subjects. Statistical analyses to explore the causal treatment effect on the prevention of HIV infection among adherent individuals are needed.

Methods. We developed an analytical strategy to evaluate whether conventional covariate adjustment removes confounding and thereby reveals a prevention effect among adherent individuals. We applied this strategy to the VOICE study, using 2 dichotomized proxy measures of product use: detection of tenofovir in plasma at least once during follow-up and detection of tenofovir in plasma at the 3-month follow-up visit.

Results. After adjustment for a set of baseline predictors of the risk of HIV transmission, the confounding associated with comparison of adherent individuals in the tenofovir gel arm to placebo recipients was nearly eliminated. The relative risk for a prevention effect among those ever having tenofovir detected was 0.53 (P = .038); the relative risk among those having tenofovir detected at 3 months was 0.40 (P = .045).

Conclusions. A novel regression approach was proposed for causal as-treated analyses in the VOICE study. While intent-to-treat analyses yield null results, this exploratory approach presented evidence suggesting a prevention effect among gel users.

Clinical Trials Registration. NCT00705679.

Keywords. causal inference; HIV prevention; preexposure prophylaxis; prevention efficacy; product adherence; tenofovir detection.

Use of antiretroviral (ARV) drugs in preexposure prophylaxis (PrEP) has been shown to be protective against sexual transmission of human immunodeficiency virus (HIV) in 4 clinical trials, including CAPRISA 004 [1], iPrEx [2], TDF2 [3], and Partners in Prevention [4]. On the basis of these results, daily use of oral emtricitabine and tenofovir disoproxil fumarate (FTC/TDF) was approved by the Food and Drug Administration in July 2012 for prevention of HIV transmission. Unexpectedly, 2 PrEP trials that evaluated the same agent in African women, FEM-PrEP and VOICE (MTN 003), reported futile intent-to-treat results for FTC/TDF in the prevention of HIV transmission [5, 6]. These seemingly contradictory findings are very likely explained by different degrees of treatment adherence achieved in the trials and by differences in the level of adherence to daily oral dosing that is required for protection between heterosexual women and men who have sex with men [7], emphasizing the critical relationship between adherence and PrEP efficacy in HIV prevention.

Accurate measurement of adherence to the study regimens used in these studies presents a challenge. Self-reported product use and counts of returned pills or gel applicators generally overestimated adherence [5, 6]. Pharmacological measures of adherence, such as tenofovir (TFV) concentrations in plasma, tissue, or body fluid, have played a pivotal role in the objective detection of adherence in PrEP trials. For instance, the VOICE study used a prespecified case-cohort sampling scheme for measuring TFV concentrations in plasma samples collected quarterly [6]. Strikingly, <40% of women had a detectable level of TFV in plasma, which is in stark contrast to the self-reported adherence of >90% to pill or gel use.

In addition to providing a more accurate measure of adherence, pharmacological evidence of product use offers the possibility of estimating the causal effect of adherence on the prevention of HIV infection (hereafter, the “prevention effect”) among adherent individuals under suitable assumptions, even if the primary intent-to-treat analysis yields null results on effectiveness. The purpose of such causal as-treated analyses is not to overthrow the primary null intent-to-treat result, but rather to explore the biological effect of the product and to provide guidance in planning for next-phase PrEP trials, if the same product is to be investigated in different populations. As we elaborate later, the validity of as-treated analyses is typically subject to strong and sometimes untestable assumptions, even when rigorous statistical inference frameworks such as potential...
outcomes are used. Therefore, interpretation of as-treated analyses requires extra caution.

Strictly speaking, adherers are participants who consumed the prescribed prevention regimen, such as daily use of ARV pills or gels in the VOICE study. However, it is difficult, if not impossible, to ascertain the individual trajectory of daily adherence during the entire follow-up period. In VOICE, collection of blood specimens for measuring the TFV concentration was scheduled every 3 months, and the TFV concentration in plasma only traces back to product use within several days. In this work, we adopted some coarse grouping rules to classify participants loosely into adherers and nonadherers, perhaps more appropriately described as product users and nonusers, respectively, and we used the term “adherers” as a working reference to product users, based on some pharmacological measure.

The main challenge to estimate the prevention effect among adherers is that ARV drug levels can be measured only for participants in the active drug arm of such trials, not for those in the placebo control arm. Therefore, without placebo tracers, it is not feasible in the VOICE trial to estimate prevention effect among adherers by directly comparing the incidence of HIV infection among adherers in the active product arm, identified by pharmacological measures, to that of a comparable group of adherers in the placebo arm. In most existing analyses, the association between drug levels and HIV infection in the active product arm only [3, 6], or a comparison between adherers in the active product arm and all participants in the placebo arm [8, 9], has been used to infer the prevention effect among adherers. However, as we previously described [10], these association analyses cannot be interpreted causally; investigators must be cautious about confounding and selection bias when comparing active product-arm participants with detectable levels of ARV to participants without any drug detection, since the 2 groups may have distinct sexual-risk-taking profiles. In essence, confounding in this setting means that characteristics associated with adherence can also be directly related to the risk for HIV infection itself. For example, in the VOICE study, young and unmarried women had lower adherence, yet these women also showed a higher risk of HIV infection in the placebo arm [6]. Thus, even if the ARV regimen offers no prophylactic effect, adherers in the active arm can still exhibit a lower incidence of HIV infection than nonadherers in the active arm or placebo recipients as a whole.

The conventional approach to selection bias and confounding is to adjust for confounding variables in regression models, as has been conducted in prior analyses of PrEP trials [3, 6, 8, 9]. However, it is difficult to assess whether confounding has been effectively removed by covariate adjustment and whether there is residual confounding from unmeasured risk variables. In this analysis, using the VOICE study as an example, we present a novel approach to assessing whether covariate adjustment is adequate for removing confounding, so that the resulting association between adherence and HIV infection risk can be interpreted causally under suitable assumptions. We used this strategy to explore the prevention effect among adherers for the 3 active products in the VOICE study: oral TDF, oral FTC/TDF, and topical TFV.

**METHODS**

**Study Design and Pharmacological Measures of Adherence**

The study population, study procedures, and study end points of the VOICE trial have been previously described [6]. Written informed consent was obtained from all VOICE participants, and the study was approved annually by pertinent institutional review and ethics committees [6]. Plasma samples were collected and stored for all participants, a portion of which from the oral TDF, the oral FTC/TDF and the TFV gel arm were selected for TFV measurements by a prespecified case-cohort sample scheme [11]. Specifically, in the 3 active product arms, all HIV seroconverters and approximately 3 times as many nonseroconverters were randomly selected for measuring TFV concentrations in plasma, which were primarily drawn from samples collected at quarterly study visits and sometimes from monthly HIV testing samples, if available. Plasma TFV concentrations were determined using a validated ultra-performance liquid chromatography–tandem mass spectrometry method. The limit of TFV quantification was 0.31 ng/mL. Two dichotomous proxy measures of adherence were investigated: (1) a detectable level of TFV (≥0.31 ng/mL) in at least 1 plasma sample during follow-up and (2) a detectable level of TFV (≥0.31 ng/mL) in the plasma sample collected at the 3-month follow-up visit.

**Statistical Analyses**

In the random subcohort for measuring TFV concentrations in plasma samples, descriptive statistics were computed for the 2 proxy measures of adherence. To examine the temporal change of individual adherence, the transition pattern from the detection status at the 3-month follow-up visit to that at the 6-month follow-up was summarized by classification trees. A Markov chain model was used to estimate the probabilities of TFV detection in one quarterly visit conditional on detection in the previous quarterly visit, with the assumption that such probabilities remain constant over time.

To assess prevention efficacy of the 3 active products, a Poisson regression model with the logarithm of individual follow-up time as the offset was used to compare the HIV infection incidence among adherers in the active arm, defined by either of the 2 proxy measures of adherence, to that in the corresponding placebo arm, with adjustment for a number of HIV infection risk predictors identified a priori [12]. Such Poisson models assume a constant rate of HIV infection and are commonly used in HIV prevention trials to compute annualized incidences of HIV infection. The choice of a simple dichotomized measure of adherence in the Poisson model over time-varying adherence...
principal strata in a Cox model is primarily driven by the intrinsic difficulty of modeling time-varying principal strata in risk sets after randomization [13, 14]. Furthermore, over half of women in the active arms did not have any TFV detection, so the dichotomized measure may present a succinct summary and deliver comparable power performance relative to more-sophisticated continuous measures of adherence. The predictors of HIV infection risk included age (<25 years or ≥25 years), marital status (yes or no), partner having other sex partners (yes, no, or don’t know), having a curable sexually transmitted infection (yes or no), having herpes simplex virus type 2 (HSV-2) infection (yes or no), and alcohol consumption frequency (never, less than once per week, or more than once per week). The case-cohort sampling scheme was accounted for by the inverse probability weighting method [15]. The data from the TDF arm and the matching placebo arm were censored on 3 October 2011, the date of stopping the TDF arm owing to futility, to ensure comparability. This analysis strategy may be subject to residual confounding; that is, after adjustment for the known predictors of HIV infection incidence, the placebo recipients may still have different risk-taking behaviors than adherers in the active arm. Therefore, the estimated incidence ratio for Poisson regression may not be solely attributed to the ARV and cannot be interpreted as a causal efficacy estimate [10].

We propose a strategy to evaluate whether there is residual confounding toward inferring causal prevention efficacy in the aforementioned association analysis. This strategy hinges on the assumption that the nonadherers in the active arm, if classified properly by pharmacological data, should not receive any protection from the randomized treatment assignment; this assumption is referred to as the “exclusion restriction” assumption in the causal inference literature [10, 16–18]. This assumption is reasonable if the nonadherers use an inadequate amount of products for conferring the benefit of PrEP. This is particularly plausible for the proxy adherence measure of never having TFV detection. If the exclusion restriction condition is satisfied, as shown in the Appendix, we can inspect the HIV infection incidence ratio between the nonadherers in the active arm and the placebo recipients as a whole, after adjustment for the HIV infection risk predictors, for evidence of residual confounding. If the adjusted incidence ratio is close to 1, there is little evidence of residual confounding, and so the incidence ratio between adherers in the active arm and the placebo recipients can be roughly attributed to the effect of the product and therefore can be cautiously interpreted as the causal prevention effect among adherers when the untestable exclusion restriction condition is satisfied. Using the potential outcomes and principal stratification notation [19–21], we formally define the causal effect and the confounding, and we provide the justification for the aforementioned analytical strategy. Owing to its technical nature, derivation of the method is relegated to the Appendix.

RESULTS
Summary of Pharmacological Measures of Adherence in VOICE
At the first quarterly visit, <40% of participants had a detectable level of TFV in plasma, and the rate of TFV detection decayed over time during follow-up [6]. The median number of quarterly plasma samples collected from 647 women in the case-cohort study was 3 (or 4, if nonquarterly visits were included). There was no correlation between the percentage of visits with detectable TFV and the number of plasma samples available for a woman (correlation coefficient, −.006; \( P = .89 \)). In the random subcohort of the case-cohort sample, approximately half of the women never had TFV detection in their plasma samples (50% in the FTC/TDF arm, 58% in the TDF arm, and 57% in the TFV gel arm; Figure 1). The percentage of women who had 100% TFV detection in their plasma samples was only 12% in the FTC/TDF arm, 20% in the TDF arm, and 10% in the TFV gel arm. The remaining women had some evidence of using study product, albeit intermittently. Note that a detectable level of TFV means different adherence levels for different routes of product application: for the gel arm, a TFV concentration below the lower limit of quantitation (0.31 ng/mL) in plasma suggests no dosing within the past 2–3 days on average; for the oral arms, it suggests no dosing in the past 7 days on average [22, 23].

Approximately one third of participants had a detectable level of TFV at their 3-month follow-up visit (36% in the FTC/TDF arm, 37% in the TDF arm, 32% in the TFV gel arm). Figure 2 shows the classification trees from the 3-month TFV detection to the 6-month TFV detection for the 2 oral product arms and the gel arm. The majority of women without TFV detection at 3 months had no TFV detected at 6 months (86% in the oral arms and 85% in the gel arm), while a substantial percentage of women with TFV detected at 3 months had no TFV detected at 6 months (37% in the oral arms and 52% in the gel arm). When TFV detection in all quarterly visits was modeled by a homogeneous Markov chain, the transition probability from no detection at one visit to no detection at the next visit was 0.85 for the 2 oral arms and 0.86 for the gel arm. The transition probability from having detection at one visit to no detection at the next visit was 0.39 for the 2 oral arms and 0.56 for the gel arm. These observations suggest that adherence decreased steadily over time and that no detection at 3 months was highly predictive of no or low adherence in later follow-up.

Exploratory Regression Analysis to Infer Causal Prevention Effect Among Adherers
The Poisson regression analyses of HIV infection incidence and adherence, defined by ever or never having TFV detection, are presented in Table 1. We adjusted the relative risk estimate for a comprehensive set of HIV infection risk predictors identified previously, some of which were also associated with TFV detection. For the comparison between the TFV gel and the placebo gel, adjustment for the set of confounding variables largely
removes the selection bias, since the adjusted relative risk for the comparison of drug nonadherers—those who never have plasma detection—in the active arm to placebo recipients is nearly 1. The relative risk estimate for the comparison of product users to placebo recipients is therefore close to the estimate of the causal effect: the prevention efficacy among those who have at least 1 sample with a detectable level of TFV is a 47% reduction in the risk of HIV infection. The associations of adjusting covariates as specified in the “Methods” section were shown in Supplementary Table 1. On the other hand, analyses for oral TDF and oral FTC/TDF did not show adequate confounding control; that is, the adjusted relative risk for the comparison of nonadherers in the active arm to placebo recipients deviates from 1 (1.29 in the FTC/TDF arm and 1.16 in the TDF arm), although not significantly. Both TDF and FTC/TDF failed to show any sizable prevention efficacy among adherers when ever having TFV detection was used as the rule to define adherence.

The results from regression analyses conducted for adherence defined by TFV detection in the plasma sample collected at 3 months are shown in Table 2. The results for the gel arm comparison are consistent with those in Table 1: those who did not have detection at 3 months in the gel arm had a similar HIV incidence as placebo recipients (adjusted relative risk, 0.94; \( P = .76 \)). Thus, the reduction in the HIV infection incidence among those who had detection of TFV at 3 months, compared with that for placebo recipients, can be interpreted approximately as the prevention efficacy (60% reduction; \( P = .05 \)). The associations of adjusting covariates as specified in the “Methods” section were shown in Supplementary Table 2. The 2 oral arms did not demonstrate evidence of a prevention effect among those who had TFV detected at 3 months.

Previous data suggested that a plasma TFV level of 10 ng/mL achieved by oral dosing is comparable to the dosing behavior associated with a plasma TFV level of 0.31 ng/mL achieved by
the vaginal route [20, 21]. In the random subcohort, 54% of women in the FTC/TDF arm and 63% of women in the TDF arm never had TFV concentrations of >10 ng/mL in their plasma samples. On the other hand, 24% of women in the FTC/TDF arm and 31% of women in the TDF arm had a TFV concentration >10 ng/mL in their plasma samples at 3 months. In Table 3, we show the results when 10 ng/mL was used as the cutoff to define adherence for the oral arms. Neither adherers defined by ever having a TFV concentration of ≥10 ng/mL nor adherers defined by having a TFV concentration of ≥10 ng/mL at 3 months showed a significant reduction in the HIV infection risk when compared to the placebo arm, although the change in estimates when moving from a definition using 0.31 ng/mL to one using 10 ng/mL points in the direction of protection.

**DISCUSSION**

In epidemiology, selection bias refers to the bias of an exposure effect on an outcome resulting from the sample selection process, such that the association in the selected samples is not entirely attributable to the true causal effect [24]. It is often difficult to assess whether confounding has been effectively dealt with by covariate adjustment. This analysis exploited the exclusion restriction condition that nonadherers should

<table>
<thead>
<tr>
<th>Variable</th>
<th>HIV Infection Incidencea</th>
<th>Adjusted RRb (95% CI)</th>
<th>P Valueb</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TFV gel</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placebo arm</td>
<td>6.8</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Detection</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>8.2</td>
<td>1.07 (0.73–1.56)</td>
<td>.54</td>
</tr>
<tr>
<td>Ever</td>
<td>3.3</td>
<td>0.76 (0.47–1.22)</td>
<td>.34</td>
</tr>
<tr>
<td><strong>FTC/TDF</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placebo arm</td>
<td>4.6</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Detection</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>6.2</td>
<td>1.21 (0.72–2.03)</td>
<td>.56</td>
</tr>
<tr>
<td>Ever</td>
<td>3.5</td>
<td>0.95 (0.57–1.58)</td>
<td>.87</td>
</tr>
<tr>
<td><strong>TDF</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placebo arm</td>
<td>4.2</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Detection</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>5.1</td>
<td>0.83 (0.50–1.40)</td>
<td>.57</td>
</tr>
<tr>
<td>Ever</td>
<td>2.9</td>
<td>0.80 (0.46–1.37)</td>
<td>.52</td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; FTC, emtricitabine; HIV, human immunodeficiency virus; RR, relative risk; TDF, tenofovir disoproxil fumarate.

* Data are cases/100 person-years.


data were calculated with respect to the placebo arm, with adjustment for HIV infection risk predictors.

---

Figure 2. Classification tree based on plasma tenofovir (TFV) detection from the 3-month follow-up visit to the 6-month follow-up visit in the 2 oral treatment arms (A) and the TFV gel arm (B). Abbreviations: FTC, emtricitabine; NA, not available; TDF, tenofovir disoproxil fumarate.
not receive any protection [10, 16–18] to devise a regression approach to assess the adequacy of control for confounding. While we used the VOICE study as an illustrative example, this approach should be generally applicable to placebo-controlled PrEP trials with drug detection as a proxy measure of adherence in the active product arm. While the primary intent-to-treat analyses showed no effect for either of the 3 active products, this exploratory analysis adds evidence for the association we detected for TFV gel when assessed by the conventional regression analysis previously reported [6], which compared adherers to nonadherers in the TFV gel arm. These results in the 1% TFV gel arm are consistent with those from the recent report from VOICE that TFV gel use is associated with a lower HSV-2 risk [25], together suggesting that the subgroup of gel users may have received some level of protection. Nonetheless, these results require careful interpretation owing to inherent limitations of as-treated secondary analyses, that sample sizes for adherence subgroups generally do not provide adequate power for testing no residual confounding or for testing causal effect among adherers. Furthermore, the untestable exclusion restriction condition, while highly plausible, can be a bit simplistic, as plasma samples collected at month 3 or other visits do not necessarily cover the entire follow-up.

Both detection of TFV at 3 months and ever having TFV detection appear to be good classification rules for separating non-adherers from full adherers and perhaps partial adherers in the gel arm. The true prevention effect among full adherers might be bigger than what were presented in Tables 1 and 2. The HIV infection risk predictors we controlled for seemed to largely remove the confounding. The primary modified intent-to-treat analysis for VOICE showed a 14.5% reduction of HIV infection risk in the gel arm, likely due to low or imperfect adherence. The same 1% TFV gel used pericoitally showed 39% reduction of HIV infection in the CAPRISA 004 trial among sexually active women in South African [1], yet recent results from the FACTS 001 trial did not show protection [26]. These results highlight the importance and complexity of adherence and prevention efficacy to study conduct and analyses.

There is little evidence that oral TDF or FTC/TDF offered protection against HIV in the product user group as defined here. The differing results between the oral arms and the gel arm may be partially explained by different pharmacological properties of plasma TFV detection for different routes. The MTN-001 study showed that serum concentrations after vaginal dosing were 56-fold lower than after oral dosing [22]. Because of this pharmacological property and similar rates of TFV detection observed in oral and gel arms, plasma TFV detection in the gel arm indicates more-recent dosing. Biologically, imperfect or low adherence in the oral arms might not confer any protection from HIV transmission in women having vaginal sex. Recent mathematical modeling suggested that nearly daily adherence is required for oral TDF/FTC dosing to achieve protection [7]. Furthermore, the plasma TFV threshold adjustment presented above and based solely on plasma TFV half-life does not take account of the half-life differences between TFV in plasma and TFV diphosphate (active moiety) half-life in cervicovaginal tissue CD4⁺ T cells [27]. Finally, the possibility of the so-called white-coat effect cannot be ruled out, as in 3 separate oral PrEP studies with paired plasma TFV and PBMC TFV-DP data, between 6% and 33% of participants had plasma concentrations consistent with daily dosing or dosing in the past 24 hours, but low TDF concentrations indicated poor recent dosing adherence (unpublished data on file). These results and discussions underscore the complexity of pharmacological measures of adherence and the difficulty of identifying a group of consistent product users based on these measures.

The key for inferring prevention efficacy is that the pharmacological definition of adherers has to align with an adequate local concentration of the drug that confers protection against HIV infection. For example, in our study, a plasma TFV concentration of 0.31 ng/mL appears to be informative regarding the effectiveness of TFV vaginal gel but not of the oral products. If the concentration level cutoff is not immediately available from prior pharmacological studies, it is informative to examine several different thresholds of drug concentration and corresponding prevention effects, as we did for the VOICE data. It is also useful to model the continuous spectrum of drug concentrations in relation to the risk of HIV infection,

### Table 3. Regression Results for Human Immunodeficiency Virus (HIV) Infection Incidence and Adherence, Defined by Whether the TFV Concentration Is >10 ng/mL in Plasma, for the 2 Oral Arms

<table>
<thead>
<tr>
<th>Variable</th>
<th>HIV Infection Incidencea</th>
<th>Adjusted RRb (95% CI)</th>
<th>P Valueb</th>
</tr>
</thead>
<tbody>
<tr>
<td>FTC/TDF</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TFV level ≥10 ng/mL</td>
<td>Ever</td>
<td>3.0</td>
<td>0.85 (.47–1.54)</td>
</tr>
<tr>
<td>tfv level at 3-mo follow-up</td>
<td>Never</td>
<td>6.5</td>
<td>1.23 (.76–1.98)</td>
</tr>
<tr>
<td>TFV level ≥10 ng/mL</td>
<td>Ever</td>
<td>2.4</td>
<td>0.81 (.35–1.85)</td>
</tr>
<tr>
<td>Never</td>
<td>6.3</td>
<td>1.38 (.80–2.37)</td>
<td>.25</td>
</tr>
<tr>
<td>TFV level &gt;10 ng/mL</td>
<td>≥10 ng/mL</td>
<td>2.7</td>
<td>0.93 (.38–2.27)</td>
</tr>
<tr>
<td>&lt;10 ng/mL</td>
<td>5.7</td>
<td>1.26 (.74–2.16)</td>
<td>.40</td>
</tr>
</tbody>
</table>

Abbreviations: CI, confidence interval; FTC, emtricitabine; RR, relative risk; TDF, tenofovir disoproxil fumarate; TFV, tenofovir.

a Data are cases/100 person-years.
b Data were calculated with respect to the placebo arm, with adjustment for HIV infection risk predictors.
so that one does not have to choose an arbitrary cut point as exemplified in Table 3.

**Supplementary Data**

Supplementary materials are available at http://jid.oxfordjournals.org. Consisting of data provided by the author to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the author, so questions or comments should be addressed to the author.

**Notes**

**Acknowledgments.** We thank all participants in the VOICE study; the VOICE study team; and Ariane van der Straten, Jeanna Piper, and Roberta Black, for helpful comments.

MTN-003 was sponsored by the National Institutes of Health (NIH) and cosponsored by CONRAD and Gilead Sciences. The study was designed and implemented by the Microbicide Trials Network (MTN). The MTN is funded by the National Institute of Allergy and Infectious Diseases (UM1AI068633, UM1AI068615, and UM1AI106707), with cofunding from the Eunice Kennedy Shriver National Institute of Child Health and Human Development and the National Institute of Mental Health, all of which are components of the NIH.

**Financial support.** This work was supported by the NIH (AI068615 and AI029168).

**Potential conflicts of interest.** All authors: No reported conflicts. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

**References**


**APPENDIX**

Denote by \( Y(z) \) the potential outcome for acquisition of HIV under either of the 2 treatment options \( Z = z \), where \( z = 1 \) if a participant is assigned to the ARV arm or 0 if assigned to the placebo arm. For a participant, the counterfactuals contain a pair of potential outcomes \( (Y(1),Y(0)) \), among which only 1 potential outcome is realized. Similarly, define by \( D(1) \) the potential outcome of the pharmacological measure of adherence had a participant been assigned to the drug arm. In the VOICE trial, \( D(1) \) is either ever having TFV detection or TFV detection at 3 months. By definition, \( D(0) = 0 \) because drug concentrations are zero for all placebo recipients. Two principal strata are therefore formed by \( D(1) \), participants who have \( D(1) = 1 \), named “drug adherers,” and those who have \( D(1) = 0 \), named “drug nonadherers” [18,9]. A causal estimand for the efficacy of the PrEP regimen among drug adherers can be generally defined as the risk ratio

\[
\Pr(Y = 1|D(1) = 1, Z = 1) = \frac{\Pr(Y = 1|D(1) = 1, Z = 0)}{\Pr(Y = 1)}.
\]  

Often we have baseline predictors of HIV risk \( (W) \), such that one would also be interested in the conditional causal estimand

\[
\Pr(Y = 1|D(1) = 1, Z = 1, W) = \frac{\Pr(Y = 1|D(1) = 1, Z = 0, W)}{\Pr(Y = 1|D(1) = 1, Z = 0, W)}.
\]  

Neither (1) nor (2) is identifiable because \( D(1) \) is not observed in the placebo arm \( (Z = 0) \).
Formally, the selection bias when comparing drug adherers and drug nonadherers for inferring efficacy can be defined as

\[
\frac{\Pr(Y(0) = 1|D(1) = 1, Z = 0)}{\Pr(Y(0) = 1|D(1) = 0, Z = 0)},
\]  

which is the ratio of HIV infection risk for drug adherers and drug nonadherers in the placebo arm. In the absence of active product, if there is a difference in HIV risk between drug adherers and drug nonadherers, then they likely engage in differential sexual risk behaviors associated with HIV acquisition. If \(W\) contains an adequate set of risk predictors, then

\[
\Pr(Y = 1|D(1) = 0, W, Z = 0) = \Pr(Y = 1|D(1) = 1, W, Z = 0),
\]

which states that conditioning on \(W\) has effectively removed confounding. Furthermore, if we assume that the restriction exclusion condition is satisfied, as shown by

\[
\Pr(Y = 1|D(1) = 0, W, Z = 1) = \Pr(Y = 1|D(1) = 0, W, Z = 0),
\]

then, by combining equations (4) and (5), we obtain the following equation

\[
\frac{\Pr(Y = 1|D(1) = 0, W, Z = 1)}{\Pr(Y = 1|D(1) = 0, W, Z = 0)} = \frac{\Pr(Y = 1|D(1) = 1, W, Z = 1)}{\Pr(Y = 1|D(1) = 1, W, Z = 0)},
\]

which states that if \(W\) contains all relevant risk predictors, the HIV infection risk among drug nonadherers conditional on \(W\) is equal to the risk among the placebo recipients conditional on \(W\). Moreover, it follows that the conditional causal estimand (2) is now estimable,

\[
\frac{\Pr(Y = 1|D(1) = 1, W, Z = 1)}{\Pr(Y = 1|D(1) = 1, W, Z = 0)} = \frac{\Pr(Y = 1|D(1) = 1, W, Z = 1)}{\Pr(Y = 1|W, Z = 0)}.
\]

This implies that causal inference can be conducted by a regression model with one indicator variable for drug adherers and the other for drug nonadherers, both of which are compared to the placebo recipients after adjustment for the baseline predictors of HIV infection risk (\(W\)).