**Supplementary Material**

**Manuscript:** Analysis of false-negative HIV rapid tests performed on oral fluid in three international clinical research studies.

**Corresponding author:** Marcel E. Curlin. Current contact information: Oregon Health and Sciences University, 3181 SW Sam Jackson Park Rd, Portland, OR 97239. Email: curlin@ohsu.edu. Tel.: 503 494 4971

**Contents**

1. HIV testing in contributing parent studies
2. Statistical methods
3. Figure S1 – HIV testing strategies in TDF2, BTS and BMCS (parental studies)
4. Figure S2 – Testing schedules and interval calculations
5. Figure S3 - Breakdown of study participants and test results in all contributing studies.
6. Supplementary table S1
7. Citations

**1. HIV testing in parent studies**

**TDF2:** Prior to enrolment, HIV screening was performed with two rapid HIV tests on whole-blood samples in parallel, using Determine HIV-1/2 (Abbott Diagnostics) and either Uni-Gold Recombigen HIV (Trinity Biotech) or OraQuick Advance HIV-1/2 (OraSure Technologies). At enrollment and at monthly visits while on study, HIV rapid testing was performed using OraQuick Advance HIV-1/2 test on oral fluid.

**BTS:** At enrolment and monthly visits while on study (every 28 days), oral fluid was tested for HIV antibodies (OraQuick Rapid HIV-1/2 Antibody Test; OraSure Technologies Inc, PA, USA). On Sept 15, 2011, testing with monthly enzyme-immunoassay (Genetic Systems HIV-1/HIV-2 Plus O EIA; Bio-Rad, Redmond, WA, USA) on blood was added to improve detection of early HIV infection. We tested blood samples obtained at the final follow-up visit for HIV with enzyme-immunoassay and nucleic-acid amplification (Aptima HIV-1 RNA Qualitative Assay; Gen- Probe Inc, San Diego, CA, USA).

**BMCS:** HIV-infection was determined at baseline and every 4 months while on study using OraQuick HIV-1/2 Rapid Test (OraSure Technologies, Bethlehem, Pennsylvania, USA). Reactive tests were confirmed according to Thai national guidelines with three rapid tests on blood [Determine HIV-1/2; Abbott Laboratories, Tokyo, Japan; Double- Check II HIV-1&2; Organics, Yavne, Israel (after 02/2011 replaced by SD-Bioline HIV-1/2 3.0; Standard Diagnostics, Kyonggi-do, South-Korea); Capillus HIV- 1/HIV-2; Trinity Biotech, Jamestown, New York, USA (after 11/2008 replaced by Core HIV-1/2, Birmingham, UK)].

**2. Statistical methods**

**Test results and intervals:** Estimated date of infection was calculated retrospectively based on the results of enzyme immunoassay (EIA) and/or nucleic acid amplification (NAAT) gold standard (GS) tests. The estimated date of infection was defined as the midpoint between the last negative and first positive GS test (**Figure S2**). If a last negative GS test was not available, infection was assumed to occur 20 days before the first positive GS test. Nonreactive OFOQ tests were categorized as true-negative (TN) or false-negative (FN), and reactive tests were categorized as either true-positive (TP) or false-positive (FP) with reference to GS test results, or unconfirmed (UNC) if no additional confirmatory testing was available. OFOQ results were considered FN if a GS test from contemporaneous or earlier specimens was positive. The OFOQ conversion delay time was defined as the time between estimated date of infection and estimated date of OFOQ conversion from a negative to a positive result. The OFOQ conversion delay time was calculated as: [date of last negative OFOQ] – [midpoint date between last negative and first positive EIA/NAAT](i.e., midpoint method, **Figure S2**). Lack of operator proficiency was defined as one or more incorrect OFOQ result interpretations during any proficiency assessments. High operator workload was defined as performing more than five additional tests on the day of testing.

**Statistical analysis:** The primary outcome variable was the proportion of FN results among all negative OFOQ responses in newly infected individuals after the estimated time of infection. Continuous variables were summarized using mean, standard deviation, median, and range. Categorical variables were summarized by frequency counts and percentages. The relationships between primary outcomes and potential predictors (participant age, gender, clinic, study arm, HIV subtype, operator workload, and time to test kit expiration) were summarized using prevalence ratios and robust 95% confidence intervals (CI) obtained from a log-binomial regression model using the generalized estimating equations (GEE) approach [[1](#_ENREF_1), [2](#_ENREF_2)]. Individual participants were treated as clusters in GEE analyses. Repeated HIV viral load measures were analyzed using a linear mixed-effects regression model and accounting for lower and upper detection limits [[3](#_ENREF_3), [4](#_ENREF_4)]. Statistical hypotheses tests were interpreted at the 0.05 level of significance.

In secondary analyses we considered; a) what proportion of negative OFOQ tests were obtained after the appearance of HIV-specific (or gp41-specific) antibodies in blood; b) whether there were any significant differences in ARV exposure (treatment arm assignment), HIV PVL, and subtype distribution between participants with prolonged (i.e. > 180 days) and those without a prolonged history of FN results, and; c) whether there were any interactions between participant, test center, test operator, and test kit lot in those with exceptionally frequent FN results.

Geometric mean HIV PVLs were calculated using all available PVL measurements obtained prior to OFOQ conversion. We analyzed log-transformed HIV PVL using the mixed-effect model described above. Test operators and clinics were categorized as having less than, or equal to or above, the corresponding median number of FN results compared with other operators or clinics, for each study. We performed cross-tabulation of FN results by clinic, test operators, test kit lots, and study participants to describe possible associations among them. Analyses were generated using SAS Version 9.3 (SAS Institute Inc., Cary, NC, USA).

**3. Figure S1**

****

**Figure S1 –** **Testing schemes in TDF2, BTS and BMCS studies.** Screening for HIV infection during the study was performed monthly in BTS and TDF2, and every four months in BMCS, using OFOQ. Positive results were confirmed by EIA and/or NAAT in blood. Look-back testing was performed retrospectively on stored blood specimens collected every three months (TDF2 and BTS ) or annually (BMCS).

**4. Figure S2**

**Figure S2.** Testing schemes in TDF2, BTS and BMCS studies (upper panels), and schematic showing calculation of the estimated OF OQ conversion delay time (bottom panel) in a hypothetical BMCS seroconverter. In this example, a hypothetical BMCS seroconverter is found to be HIV-infected by OFOQ at 20 months (black oval), and look-back testing in blood reveals the estimated time of infection to be six months by the midpoint method (midpoint between open and solid diamonds). Key: TDF2: TDF2 study; BTS: Bangkok Tenofovir Study; BMCS; Bangkok MSM Cohort Study; EIA: enzyme immunoassay (diamonds); NAAT: nucleic acid amplification test (diamonds); OFOQ: oral fluid OraQuick test (ovals); non-reactive tests: open symbols; reactive tests: closed symbols.

**5. Figure S3**



**Figure S3 –** Breakdown of OraQuick HIV test results by participant category and test result, in the TDF2 study, the Bangkok Tenofovir Study, and the Bangkok MSM Cohort Study. Bold numbers represent study participants and italics represent tests.

**6. Supplementary Table S1**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | TDF2 | BTS | BMCS | Combined |
| FN results | no FN results | FN results | no FN results | FN results | no FN results | FN results | no FN results |
| Seroconverters | 12 (36.4) | 21 (63.6) [22] | 32 (60.4) | 21 (39.6) | 36 (18.0)  | 164 (82.0) | 80 (28.0) |  206 (72.0) [207] |
| Age (yrs): Median (range) | 24 (22-36) | 25.5 (21-29) | 30 (20-50) | 27 (20-53) | 24 (19-48) | 24 (18-42) | 26 (19-50) | 25 (18-53) |
| Gender: n (%) Female Male | 8 (66.7)4 (33.3) | 13 (61.9) [14]8 (38,1) | 5 (15.6)27 (84.4) | 6 (28.6)15 (71.4) | 0 (0.0)36 (100.0) | 0 (0.0)164 (100.0) | 13 (16.3)67 (83.7) | 19 (9.2) [20]187 (90.8) |
| Study arm: n (%) TDF (TDF/FTC in TDF2) Placebo | 4 (33.3)8 (66.7) | 2 (9.5) [3]19 (90.5) | 11 (34.4)21 (65.6) | 6 (28.6)15 (71.4) | NA | NA | 15 (34.1)29 (65.9) | 8 (19.0) [9]34 (81.0) |
| HIV subtype: n (%) B/BE E C Missing | 0 (0.0)0 (0.0)12 (100.0)0 (0.0) | 0 (0.0)0 (0.0)23 (100.0)0 (0.0) | 3 (9.4)25 (78.1)0 (0.0)4 (12.5) | 2 (9.5)19 (90.5)0 (0.0)0 (0.0) | 5 (13.9)24 (66.7)0 (0.0)7 (19.4) | 26 (15.9)122 (74.4)0 (0.0)16 (9.8) | 8 (10.0)49 (61.3)12 (15.0)11 (13.7) | 28 (13.5 )141 (67.8)23 (11.1)16 (7.7) |
| No. of test results (total) | 280 [351] | 1583 [1726] | 1008 [1351] | 2871 [3428] |
| Tests by category: n (%) | 34 (12.1) | 246 (87.9) | 147 (9.3) | 1436 (90.7) | 52 (5.2) | 956 (94.8) | 233 (8.1) | 2638 (91.9) |
| Clinics: n (%) | 1 (50.0) | 1 (50.0) | 13 (86.7) | 2 (13.3) | NA | NA | 14 (82.4) | 3 (17.6) |
| Operators: n (%) | 6 (46.2) | 7 (53.8) | 23 (62.2) | 14 (37.8) [15] | 10 (76.9) |  3 (23.1)  | 39 (61.9) | 24 (38.1) [25] |
| Test lots: n (%)  | 9 (37.5) | 15 (62.5) [17] | 48 (72.7) | 18 (27.3) [20] | 23 (40.3) | 34 (59.6) [35] | 80 (54.4) | 67 (45.6) [72] |
| Delay time: Median (range)  | 52 (0-236) | NA | 112 (0-547) | NA | 20 (0-384) | NA | 82 (0-547) | NA |
| \*\*Viral load [FN vs. TP] : n (%) Geometric Mean  (95% CI)  | 20 (41.7)24,801(9.808-62,710) | 28 (58.3)19,337(9,754-38,334) | 65 (58.6)16,657(6,629-41,857) | 46 (41.4)33,102(13,933-78,642) | 25 (12.7)46,770(17,828-122,698)  | 172 (87.3)52,009(36,808-73,488) | 110 (30.8)24,409(14,064-42,361) | 247 (69.2)42,978(31,642-58,374) |
|  p-value | 0.652 | 0.229 | 0.834 | 0.067 |

**Table S1.** **OraQuick-related descriptive characteristics for the Oral Fluid test based on data from TDF2 study (Botswana, 2007-2010), the Bangkok Tenofovir Study (Bangkok, 2005-2012) and the Bangkok MSM Cohort Study (Bangkok, 2006-2011).** Numbers in square brackets represent statistics based on all OraQuick tests, including false-negative (FN), true-negative (TN), true-positive (TP), and “unconfirmed” OFOQ test results that cannot be verified due to lack of an available confirmatory blood test. “∥ “ denotes column percentages. “\*\*” denotes geometric mean and 95% confidence interval (95% CI) obtained by back transforming from log scale the results of a mixed-effects model accounting for lower and upper limits of detection. NA indicates “not applicable”

**7. Citations**

1. McNutt LA, Wu C, Xue X, Hafner JP. Estimating the relative risk in cohort studies and clinical trials of common outcomes. American journal of epidemiology 2003 May 15;157(10):940-3.

2. Hardin JW, Hilbe JM. Generalized estimating equations. Boca Raton, Fla.: Chapman & Hall/CRC, 2003.

3. Wiegand RE, Rose CE, Karon JM. Comparison of models for analyzing two-group, cross-sectional data with a Gaussian outcome subject to a detection limit. Statistical methods in medical research 2014 May 5.

4. Thiebaut R, Jacqmin-Gadda H. Mixed models for longitudinal left-censored repeated measures. Computer methods and programs in biomedicine 2004 Jun;74(3):255-60.