2322 HIV RECENT INFECTION TEST-BASED INCIDENCE AS A COUNTERFACTUAL FOR NEW PREP TRIALS

Neil Parkin[1], Fei Gao[2], Eduard Grebe[3,4], Amy Cutrell[5], Moupali Das[6], Deborah Donnell[2], David Glidden[4], James P. Hughes[7], Charu Mullick[8], Jeffrey Murray[8], Michael Robertson[9], Tamar Tchelidze[10], and Veronica Miller[10] for the Forum for Collaborative Research Recency Assay Working Group

[1] Data First Consulting, Sebastopol, CA; [2] Fred Hutchinson Cancer Research Center, Seattle, WA; [3] Vitalant Research Institute, San Francisco, CA; [4] University of California San Francisco, San Francisco, CA; [5] ViiV Healthcare, Research Triangle Park, NC, USA; [6] Gilead Sciences, Foster City, CA; [7] University of Washington, Seattle, WA; [8] US Food and Drug Administration, Silver Spring, MD; [9] Merck & Co., Inc., Kenilworth, NJ, USA; [10] Forum for Collaborative Research, Washington, DC

Background

- Clinical trials of new PrEP agents are challenging because it is not ethical to include a placebo-only group.
- Innovative ways to evaluate new PrEP modalities are needed without impractically large sample sizes required for non-inferiority trials.
- HIV recent infection testing algorithms (RITAs) such as the limiting antigen avidity assay (LAg) plus viral load (VL; **Figure 1**) could be used to derive a "counterfactual" incidence estimate using specimens from untreated, HIV-positive people identified during screening, to which on-PrEP incidence can be compared (**Figure 2**).
- The feasibility of this approach is partly dependent on the sample size needed to ensure adequate power, which is impacted by RITA performance, the number of recent infections identified, the expected efficacy of the intervention, and other factors.

Methods

- Sample sizes (number of persons screened) required to support detection of an 80% reduction in incidence (null hypothesis: 50% reduction) were calculated based on a test statistic of log incidence ratio[1] in different populations, and assuming:
  - a single active intervention arm
  - 4th generation Ab/Ag testing to identify HIV-positives
  - 90% enrollment, 90% recency testing success
  - two years of follow-up on PrEP
  - significance level 0.05 and power 0.8
- Subtype-specific mean durations of recent infection (MDRI) and false recent ratios (FRR) for the LAg + VL RITA were derived from CEPHIA[2] (**Table 1**)
- 7.4% and 14.5% relative standard error on the MDRI for subtype C and subtype B, respectively
- 25% relative standard error on FRR

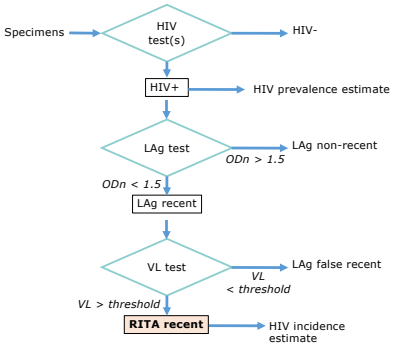
Results and Discussion

- Sample sizes are dependent on several variables (**Table 2**)
- Required sample sizes for four key populations were modeled (**Table 3**)
- These sample sizes are smaller than or comparable to the number of participants in recent phase 3 PrEP trials (e.g. HPTN 083/084, DISCOVER)
- For a given incidence and prevalence, increasing the MDRI from 120 to 200 days, keeping other assumptions constant, leads to a decrease in sample size, especially when FRR is 3% or lower (**Figure 3A**).
- When MDRI is low (e.g. below 150 days) and FRR is 4% or higher, the sample sizes become so high that the trial may not be feasible.
- For a specific combination of MDRI and FRR, samples sizes increase with lower incidence or higher prevalence; when incidence is over approximately 5% per year, the impact of higher prevalence is lessened. At low prevalence (e.g. 10%), sample sizes remain feasible at lower incidence (**Figure 3B**).

Conclusions

- Counterfactual incidence estimates based on recent infection testing can facilitate next-generation PrEP trials, at least in high incidence populations, using RITAs that have been calibrated, and where the efficacy of the intervention is expected to be very high.
- Sample sizes may not be feasible in populations with lower incidence, where the FRR is high (e.g. subtype D), or if PrEP efficacy is not expected to be very high.
- Despite these limitations, generation of a counterfactual incidence estimate based on recency assays appears to be feasible, offers high statistical power, and is nearly contemporaneous with the on-PrEP population.

Figure 1. A Recent Infection Testing Algorithm



LAg: limiting antigen avidity assay; ODn: normalized optical density; VL: viral load; RITA: recent infection testing algorithm

Figure 2. Recent infection testing to estimate counterfactual incidence

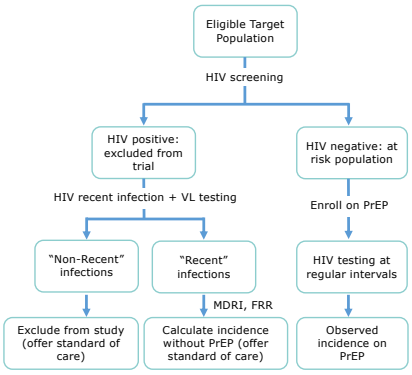


Table 1. MDRI and FRR for LAg (Sedia) and viral load-based RITA

Subtype	VL >	MDRI[a] (days) (95% CI)	Untreated FRR (%) (95% CI)
All	75	202 (180 - 224)	1.7% (0.4 - 4.9)
	1000	171 (152 - 191)	1.1% (0.1 - 4)
A	75	212 (158 - 274)	2.6% (0.1 - 13.8)
	1000	186 (137 - 245)	2.6% (0.1 - 13.8)
B	75	189 (145 - 239)	1.8% (0.1 - 9.6)
	1000	176 (132 - 226)	0% (0 - 6.4)
C	75	194 (169 - 222)	1.4% (0 - 7.6)
	1000	162 (141 - 185)	1.4% (0 - 7.6)
D	75	262 (168 - 375)	NA
	1000	209 (126 - 307)	NA

[a] MDRI based on T=2 years, HIV infection detection using a hypothetical test with a sensitivity of 1 copy/mL[2]. When using 4th generation Ag/Ab tests, the MDRI should be shortened by 11 days. FRR shown is for untreated patients. FRR for treated patients is 0% for all subtypes (95% CI 0-2.8%). A weighted FRR based on the estimated proportion of patients on ART should be used. S/CO: signal to cut-off ratio. NA: not available, ≤10 individuals infected with subtype D for ≥2 years.

Table 2. Key variables that determine required sample sizes

Variable	Rationale	Comments
Desired level of precision	Higher precision requires larger sample sizes which will allow greater opportunity to observe events (i.e. recency assay positive infections and new infections on PrEP)	The level of precision that is required to support the determination of a statistically significant difference in the two incidence estimates can be determined based on accepted levels of type-1 error and power.
Expected prevalence and incidence	For maximum precision and smallest possible sample size, the proportion of HIV-positive cases that meet the recency assay/algorithm criterion should be as high as possible, and the background of cases of long-standing infection as low as possible	High prevalence leads to a reduction in the number of individuals at risk for HIV acquisition, which is a component of the denominator in the incidence rate calculation
Expected efficacy of the intervention	An intervention that is expected to have a very large effect, such as ≥90% reduction in incidence compared to the counterfactual estimate, will require less precision in order to reach statistical significance of the reduction	Minimum acceptable effect of the intervention under the null hypothesis, i.e., the intervention effect we would like to rule out, also impacts the sample size.
Assay calibration (MDRI and FRR)	Longer MDRI and lower FRR will permit the accurate detection of more recent infections compared to shorter MDRI or higher FRR	Calibration parameters are specific to the assay or RITA and the target population (e.g. based on subtype prevalence)
Trial design (e.g. number of different treatments being evaluated)	Evaluation of more than one intervention, or inclusion of an active control arm, will significantly impact required sample sizes	

Table 3. Sample sizes required for clinical trials using RITA-based counterfactual incidence estimates

	Scenario 1	Scenario 2	Scenario 3	Scenario 4
Location(s)	KwaZulu-Natal, South Africa[3]	Mpumalanga province, South Africa[4]	South Africa, eSwatini, Kenya, and Zambia[5]	Peru[6]
Population	AGYW 14-17 years old	MSM > 18 years old	AGYW 16-35 years old	MSM> 18 years old
Survey period	2004-2007	2012-2015	2016-2018	2016-2018
Incidence (annual)	4.7%	12.5%	3.8%	6.1%[b]
Prevalence	27.6%	32.4%	12.1%[a]	29.0%[b]
Subtype	C	C	C	B
MDRI (days)	151	151	151	165
FRR[c]	1.4%	1.4%	1.4%	0%
Number on PrEP[d]	1,369	454	1,469	904
Total number to be screened	2,101	747	1,857	1,414

[a] prevalence based on number of HIV positive individuals identified during screening for the ECHO trial[5]

[b] incidence from placebo group from the AMP study in Lima[6]; prevalence from surveillance study of HIV positive MSM and transwomen; Ministry of Health of Peru, Coordinadora Multisectorial Multisectorial en Salud (CONAMUSA) and Jorge Sanchez, personal communication

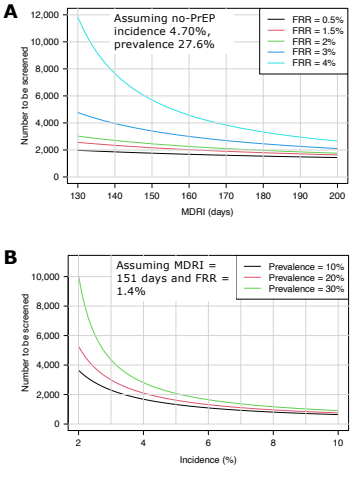
[c] FRR derived from ART naïve patient population. If the proportion of people infected with HIV and on ART can be estimated, a weighted FRR (assuming no false recent results in treated patients) should be used instead.

[d] The number enrolled into the PrEP intervention arm

AGYW: Adolescent girls and young women. MSM: men having sex with men.

Figure 3. Impact of incidence, prevalence, MDRI and FRR on required sample sizes.

Sample sizes shown are the number to be screened, assuming a single arm in the trial.



References

1. Gao F, Glidden DV, Hughes JP, Donnell D. 2020. https://arxiv.org/abs/2011.00725; https://github.com/feigao1/samplesize_RA
2. Sempa JB, Welte A, Busch MP, Hall J, Hampton D, Facente SN, et al. PLoS One. 2019;14(7):e0220345.
3. Abdool Karim Q, Kharsany AB, Frohlich JA, Werner L, Mlotshwa M, Madlala BT, et al. AIDS Behav. 2012;16(7):1870-6.
4. Lane T, Osmand T, Marr A, Struthers H, McIntyre JA, Shade SB. J Acquir Immune Defic Syndr. 2016;73(5):609-11.
5. Evidence for Contraceptive Options and HIV Outcomes (ECHO) Trial Consortium. Lancet. 2019;394(10195):303-13.
6. Corey L, Gilbert PB, Juraska M, Montefiori DC, Morris L, Karuna ST, et al. N Engl J Med. 2021;384(11):1003-14.

Acknowledgements

We thank Tamar Tchelidze and Joseph Lau (Forum for Collaborative Research) for their organizational support of the Recency Assay Working Group.
Funding: The Forum for Collaborative Research received project funding from the Bill and Melinda Gates Foundation (award number: INV-010392/OPP1199945), and is also supported by AbbVie, Gilead, Merck, Janssen, Monogram Biosciences, ViiV Healthcare, Abbott Diagnostics, and Quest Diagnostics. Deborah Donnell, Fei Gao, and James Hughes received funding support from NIH grant number UM1AI068635.
The Forum for Collaborative Research Recency Assay Working Group has the following members in addition to all co-authors: Yen Pottinger (Columbia University); Regine Lehnert (Federal Institute for Drugs and Medical Device); Timothy D. Mastro (FHI360); Beatriz Grinsztejn (FIOCRUZ); Joseph Lau, Tinkhani Mbichila, Eunice Ndzerem-Shang, and Tamar Tchelidze (Forum for Collaborative Research); Jared Baeten, Christoph Carter, Stephanie Cox, Ramin Ebrahimi, Alex Kintu, James Rooney, and Yongwu Shao (Gilead Sciences); Jessica E. Justman (ICAP at Columbia; Mailman School of Public Health); Zeda Rosenberg (International Partnership for Microbicides); Oliver Laeyendecker and Susan H. Eshelman (Johns Hopkins School of Medicine); Frances Cowan (Liverpool School of Tropical Medicine); Brian Rice (London School of Hygiene & Tropical Medicine); Joe Ma (Maxim Biomedical); Elizabeth Russell and Kathleen Squires (Merck); Ronald Mink (Sedia Biosciences); Alex Welte (Stellenbosch University); Kimberly Struble and Wen Zeng (U.S. Food and Drug Administration); David Dunn (University College London); Peter J. Dailey and Sandra McCoy (University of California; Berkeley); George Rutherford and Susie Welty (University of California; San Francisco); Bharat Parekh and Dawn K. Smith (US Centers for Disease Control); Lusine Ghazaryan (USAID); Vani Vannappagari (ViiV Healthcare); Michael P. Busch (Vitalant Research Institute); Sally Hodder (West Virginia Clinical and Translational Science Institute).