THE FORUM FOR COLLABORATIVE RESEARCH

PUBLIC MEETING WEBINAR 3 – PROTOCOL DESIGN CONSIDERATIONS: ANALYSES FOR EFFICACY

IN-DEPTH WEBINAR REPORT

13 – 14 April 2021
Location: Virtual

1608 Rhode Island Avenue NW
Washington, DC 20036
www.forumresearch.org

THE FORUM
For Collaborative Research℠

Berkeley Public Health

CONTENTS

BACKGROUND

With highly effective antiretroviral therapy for treatment of HIV infected individuals and for prevention of HIV acquisition in those at risk for HIV exposure, we still have unacceptably high incidence rates in many communities. Incidence rates as low as 0.16/100 patient-years-of-follow-up (PYFU) in DISCOVER and 0.41/100 PYFU in HPTN 084 demonstrate the immense potential of pre-exposure prophylaxis (PrEP) to controlling the HIV epidemic, but the field is at cross-roads. Adherence to daily regimen will continue to be a challenge. Not everyone eligible to receive PrEP will be able to or be willing to take oral medication. We need to increase the choice of products to meet the needs of everyone, anywhere, who is in need of PrEP.

This report summarized a two-day workshop focused on innovation in trial design to facilitate development of new PrEP products. “We are victims of our success” said Dr. Kenneth H Mayer of the Fenway Institute and Harvard Medical School. “Now we have a challenge: what kind of other ways can we intelligently design trials so we can take advantage of knowing more about the epidemiology, of the spread of HIV in communities, and be able to make robust conclusions from different sorts of data. For that we need input from all key stakeholders, which is what this meeting is set out to do”.

Randomized placebo-controlled trials are the gold standard methodology for estimating the counterfactual and inferring causality. Since placebo arms are no longer ethical for HIV prevention studies, the traditional randomized trial choices are superiority or non-inferiority trials using the standard-of-care as an active control, with the assumption that the original superiority of the standard-of-care to placebo remains constant. Superiority becomes more difficult to demonstrate with each new generation of product, given the high efficacy of the standard-of-care – in this case, emtricitabine and tenofovir disoproxil fumarate (F/TDF) – requiring larger and larger studies to demonstrate small incremental increases in efficacy. Non-inferiority designs depend on reliable data demonstrating superiority of the standard-of-care versus placebo for the population under study to allow establishing appropriate non-inferiority margins. In HIV prevention, such data do not exist for women.

To address these statistical and logistical challenges in the development of new PrEP products, the HIV Forum of the Forum for Collaborative Research initiated a multi-pronged project to discuss alternative pathways for demonstrating efficacy and safety of new PrEP products, the Forum’s HIV Prevention Trial Design Project (PrEP Project). The consensus that evolved is based on the use of a counterfactual estimate of HIV incidence through an *external* “placebo” control. This external control counterfactual estimate can be constructed using data from various sources, as discussed below. Of note, the external control counterfactual estimate-based approach should not be thought of as replacing the gold-standard placebo-controlled randomized clinical trial, rather, to superiority or non-inferiority trials.

The purpose of the two-day webinar was to discuss the external control counterfactual design, primary and secondary analyses, and obtain broad public input on PrEP study

design issues, such as specific approaches to estimate background HIV incidence, including laboratory tools such as HIV recency assays. "Grappling with how we are going to design these studies going forward, which we still desperately need, has to be a global endeavor; we can't do this on a country-by-country level" said Dr. Helen Rees from University of Witwatersrand and PrEP Project Steering Committee co-chair. "These recommendations will be important for all Regulatory Authorities." During the workshop, experts from around the globe deliberated on the choice of external controls to derive a counterfactual estimate for HIV incidence in the communities where the trials are conducted and the role of active control as an additional benchmark. In addition, stakeholders discussed the HIV infection risk tolerance and clinically meaningful threshold for efficacy to help determine stringent parameters for analyzing HIV prevention trials and declare a "win." **Table 1 and Table 2** summarize the questions addressed during the breakout rooms and throughout the workshop.

Table 1: Questions posed to breakout room participants

Acceptable Thresholds for Efficacy **Breakout Rooms 1 & 2 (MSM, TGW, & CGW)**	**Methods for Deriving Counterfactual Estimate Protocol Design: Screening, Inclusion & Exclusion Criteria** **Breakout Room 3**
What is an acceptable efficacy threshold (90%, 80%, etc.) for an investigational product compared to the placebo estimate or background HIV incidence?	Screening: What type of screening criteria for inclusion and exclusion should be applied? • Do we need to continue screening after the trial is fully enrolled to improve our estimates of HIV incidence at the baseline and throughout the duration of the trial? • If yes, then what do we do with individuals who test negative? What are our ethical obligations?
If we have the standard of care arm in the study, how close do we want the new investigational product to be the standard of care? How much precision do we need? How do we power for that level of precision? • What is the clinically meaningful level of comparison?	The recommendation is to use two estimates of background incidence. What other external control possibilities shall we consider, e.g., drug level/adherence correlation with incidence; epi data from trial sites; correlations for other biomarkers; incidence data from contemporaneous or recently completed placebo-controlled trials in similar populations?
Is it essential to differentiate between adherence vs. pharmacologic failure? If so, how do we do that?	How do we analyze estimates from multiple sources of data? How do we weigh these different sets of evidence?

Abbreviations: CGW, cisgender women; MSM, men who have sex with men; TGW, transgender women

Table 2: Questions posed during the workshop

Questions Relating to the Primary Endpoint	Questions Related to Internal Active Control Comparison

• What is the smallest preventive efficacy, compared to no treatment, that you would accept given efficacy of current standards of care? For example, is a 30% or 50% reduction in efficacy acceptable? Does this vary by type of product (ease of use)?
• Would you accept less efficacy for specific product types and why? When analyzing the primary endpoint, how much statistical "discounting" should be done for:
  ○ Preserving efficacy
  ○ Uncertainty surrounding the "placebo" incidence estimate using recency assays or other external controls
• How should one consider adherence vs. pharmacologic ("true") failures when analyzing the data?
	• How comparable should a new drug be to an active control?
	• Is this dependent on product type?
	• How can we quantify acceptable numbers of increased infections compared to an active control? What differences do you want the trial powered to detect?
	• Do adherence failures vs. pharmacologic failures have a role in decision making?

DAY ONE

BUILDING A COUNTERFACTUAL: POSSIBILITIES FOR EXTERNAL CONTROLS

The closest comparison group that can estimate the counterfactual (the ideal) is a placebo. However, since placebo is no longer ethical in PrEP trials, an alternative way, such as using external controls, is under consideration. The International Council for Harmonization (ICH) Guidance E10: Choice of Control Group allows for external controls and discusses limitations and approaches that might make externally controlled trials more persuasive and less biased. The guidance recommends multiple external controls if there is no one optimal external control. Generally, the study groups should be substantially superior to the most favorable control to conclude efficacy. Examples of potentially suitable external controls include HIV surveillance data, other epidemiologic studies, and data from recently completed trials enrolling individuals not on PrEP. For MSM, there may be a correlation of rectal gonorrhea and HIV incidence. A newer approach using a back-calculation of adherence (measured by drug levels) and efficacy has been developed for MSM and may also be applicable to women (1).

"We need to have a method that is available for studies in all population. We also need to understand which method most closely approximates what the true HIV incidence is in the population that we are going to be randomizing into our study", said Dr. Moupali Das, who works on the PURPOSE trials at Gilead Sciences, Inc. Control patients should be as

similar as possible to the population receiving the study drug. If the HIV incidence in the study population compared to the background HIV incidence (external control counterfactual) is considerably lower, one could conclude that the drug is working to prevent HIV infection.

A regulatory precedent, the Pearl Index used for contraception studies, is based on the same approach. Furthermore, the results from the DISCOVER trial of a two-drug combination of emtricitabine and tenofovir alafenamide (F/TAF), a classic non-inferiority study, is proof of concept of how such a method works. In this trial, Gilead Sciences, Inc. calculated the background HIV incidence in the absence of PrEP using multiple approaches.

In the PURPOSE trials that evaluate the safety and efficacy of lenacapavir, a highly potent inhibitor of HIV capsid protein, Gilead Sciences, Inc. will use an alternative clinical trial design approach, specifically a cross-sectional survey using a recent infection testing algorithm (RITA), to determine the background HIV incidence. In addition, secondary analyses will include a comparison to F/TDF as an internal active control that will determine comparable or superior efficacy.

RECENT HIV INFECTION TESTS TO ESTABLISH HIV INCIDENCE "EXTERNAL CONTROL COUNTERFACTUAL ESTIMATE"

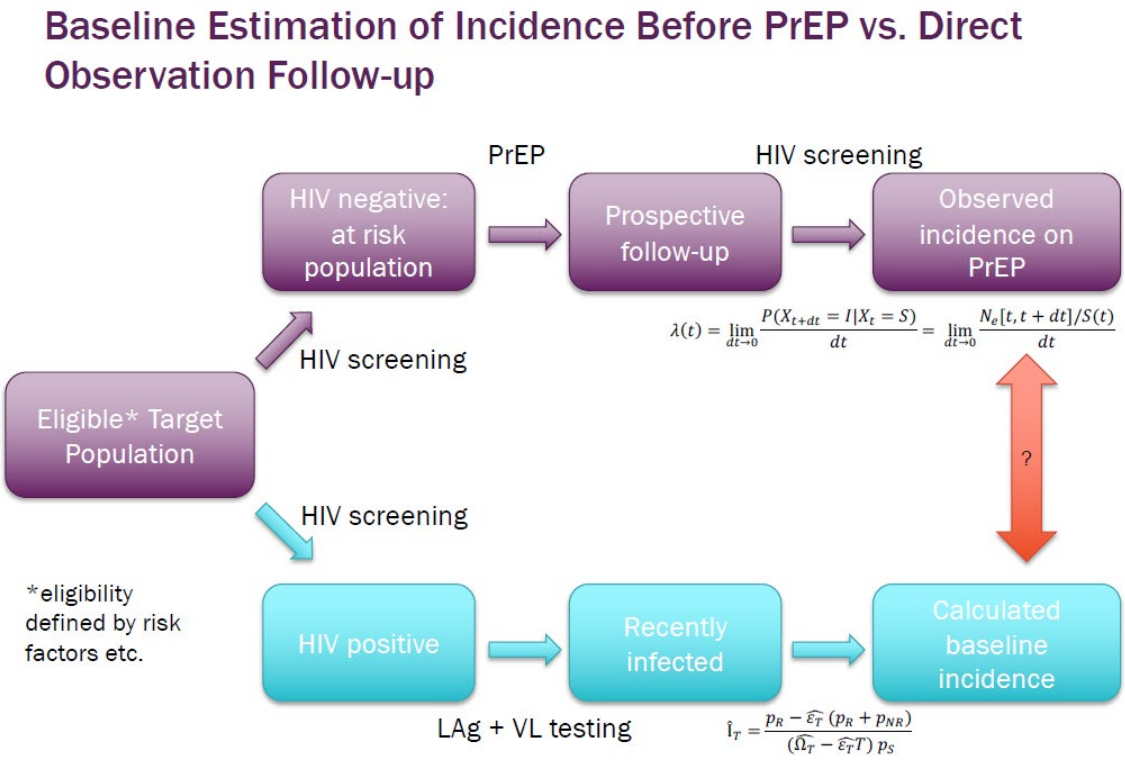
Three well-described methods to measure the HIV incidence are longitudinal cohort studies, inference from serial prevalence data, and cross-sectional survey using an incidence assay. Each method has its limitations and advantages.

A cross-sectional incidence assay (antibody avidity test) is able to distinguish recent from non-recent infection (see below). The Limiting Antigen Avidity Enzyme Imunoassay (LAg), a widely used test originally developed at the CDC, measures tightly binding (i.e., high-avidity) antibodies in a person's blood sample. The results are interpreted based on whether the normalized optical density, a quantitative measure, falls below (recent) or above (non-recent) an arbitrary threshold. Higher normalized optical density reflects more tightly binding antibodies indicating a longer duration of infection. Two key parameters of the test, the mean duration of recent infection (MDRI) and false recency ratio (FRR), must be within specific ranges to estimate HIV incidence with reasonable precision. Ideally, a bigger MDRI and smaller FRR are better. Many people receiving antiretroviral treatment might present on serological tests as "recent" because of low viral load and low-avidity antibodies. "It is recommended to use an algorithm that includes a recency assay and other assays, such as HIV viral load, to reduce the FRR", said Dr. Neil Parkin from Data First Consulting, Inc.

One of the feasibility concerns about this approach has always been the sample size. However, experts concluded that the number of people needed to screen could be as

small as 2000. "This was good news for us because we think that these numbers are certainly very feasible, based on the comparison to the numbers of people included in clinical trials that have been done recently," said Dr. Parkin. An important caveat, when designing the trial using this approach, is to ensure that the characteristics of the trial population, including both HIV-negative and HIV-positive people tested for recent infection, are as similar as possible. Additionally, sample size and the precision of the incidence estimates are dependent on assumptions, such as prevalence and incidence in the respective population, which may not be perfect. In this situation, continued screening of people to collect more data for the external control counterfactual estimate is one way to improve precision. **Figure 1** illustrates how the recency assay can be used to generate a counterfactual incidence estimate to which observed incidence in people on PrEP can be compared.

Figure 1



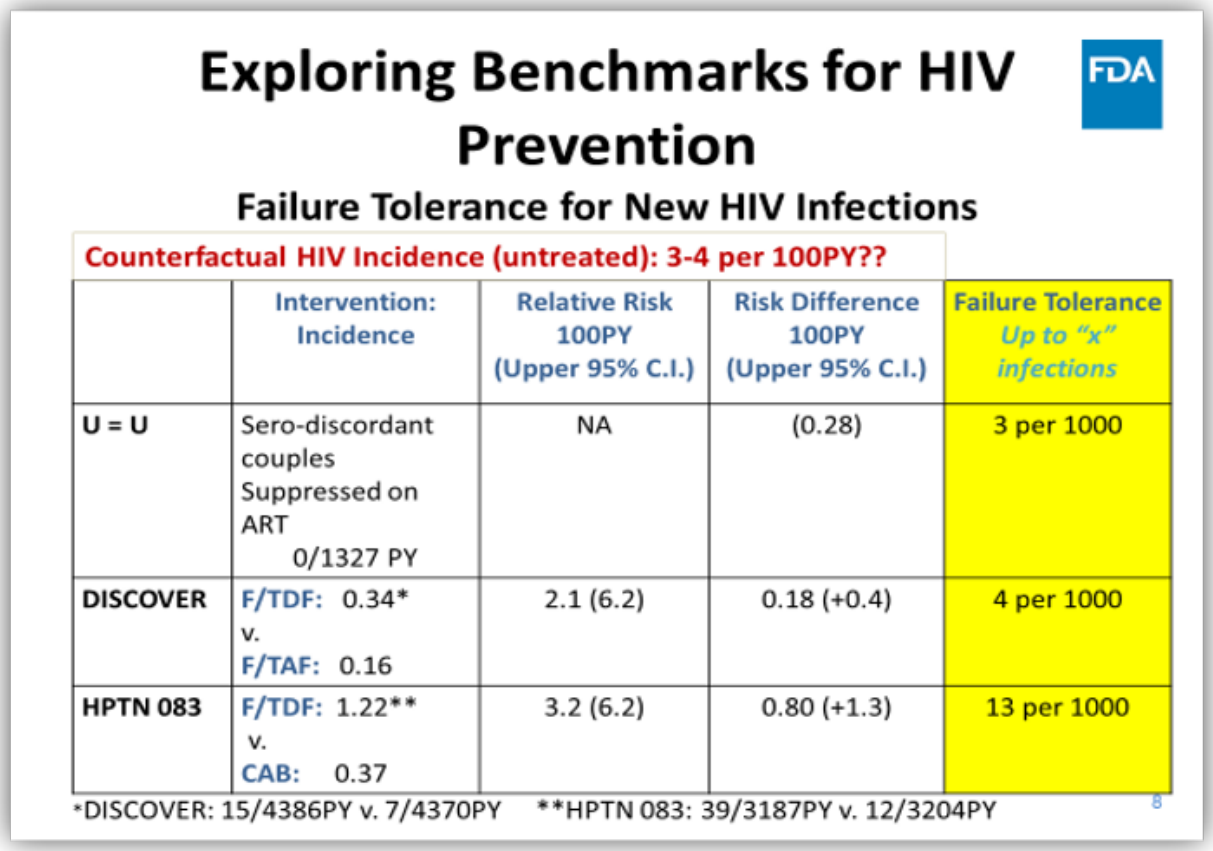
HIV INFECTION RISK TOLERANCE: PRIMARY & SECONDARY ANALYSES

In this novel counterfactual clinical trial design, the primary endpoint compares the HIV incidence observed in subjects randomized to the new intervention to the counterfactual estimate external control. The secondary endpoint assesses how comparable the new treatment is to the standard-of-care (active control). “We need to do decision framing to

qualify our tolerance of HIV infection risk to determine a clinical threshold for efficacy and declare a 'win' ", said Dr. Jeffrey Murray, Deputy Director at the Division of Antivirals of the U.S. Food and Drug Administration.

There are benchmarks for HIV treatment regarding what could be tolerated for clinical differences in viral suppression rates for two HIV treatments or failure rates above approved agents. For example, the statistical margins are much smaller than needed to show that the new treatment was better than the placebo. In treatment-naïve patients, the typical difference between the two treatments is less than 4%, using an upper statistical margin of 10%. In other words, we are willing to tolerate a difference up to 10%. However, in "Switch" trials (already virally suppressed individuals), the difference is much smaller – around 1% with an upper statistical margin of 4% –indicating that we are less tolerant of failure when patients are switched from one successful treatment to another.

Table 3, titled *Exploring Benchmarks for HIV Preventions*, presented by Dr. Murray, shows several data points about what magnitude of HIV infection risk is clinically important. In the first example, the U = U (undetectable equals untransmissible) program was a meta-analysis that demonstrated zero infections in 1327 patient-years-of-follow-up. This translates into a potential infection rate of up to 3 per 1000 persons and we tolerate this low level (essentially zero) of risk. In the DISCOVER trial, the relative risk for F/TDF vs. F/TAF was 2.1. Although F/TAF was numerically and statistically superior to F/TDF, the risk difference was small – 0.18 with an upper bound of 0.4. Most would agree that the two regimens are not *that* different from each other, and that "up to 4" infections per 1000 persons is a clinically acceptable and meaningful outcome. In the HPTN083 study comparing F/TDF to long-acting cabotegravir however, although the relative risk was 3.2 with an upper bound of 6.2 – essentially very similar to DISCOVER – the risk difference of 0.8 with an upper bound of 1.3 was considerably larger. Thirteen infections per 1000 persons may be more than we would be willing to tolerate. Going forward, we might consider a failure tolerance of somewhere between 4 per 1000 and 13 per 1000 as acceptable and power our studies accordingly.



Exploring Benchmarks for HIV Prevention

FDA

Failure Tolerance for New HIV Infections

Counterfactual HIV Incidence (untreated): 3-4 per 100PY??

	Intervention: Incidence	Relative Risk 100PY (Upper 95% C.I.)	Risk Difference 100PY (Upper 95% C.I.)	Failure Tolerance *Up to "x" infections*
U = U	Sero-discordant couples Suppressed on ART 0/1327 PY	NA	(0.28)	3 per 1000
DISCOVER	F/TDF: 0.34* v. F/TAF: 0.16	2.1 (6.2)	0.18 (+0.4)	4 per 1000
HPTN 083	F/TDF: 1.22** v. CAB: 0.37	3.2 (6.2)	0.80 (+1.3)	13 per 1000

*DISCOVER: 15/4386PY v. 7/4370PY **HPTN 083: 39/3187PY v. 12/3204PY

8

A Proposal for a Primary Analysis for an Efficacy Threshold Compared to the External Control Counterfactual

Suppose the external control counterfactual HIV incidence estimate was 3.3 per 100 person-years (95% CI: 3; 3.6). One could then use a fraction of the lower 95% CI (i.e., 3.0) as the upper 95% CI of the observed trial infection rate to define a "win" for the new drug. If one uses half, or 50%, the upper bound for the new drug arm would be 1.5 per 100 patient years of follow-up. Thus, if the true placebo rate were 3.0 per 100 PY, or 30/1000 persons, up to 15 infections/1000 persons on the new drug would be acceptable for it to be considered efficacious. This "discounting" approach is used in vaccine studies. We can decide on the amount of discounting (e.g., 30% or 50% or any other level), depending on our expectations for the product. In other words, we would not be satisfied with a small difference between the new drug and the placebo incidence rates – in the above scenario, we are expecting the new drug to be at least 50% efficacious when compared to the external control HIV incidence estimate. Such "discounting" is essential to preserve a certain amount of preventive efficacy, allow room for error in calculating the incidence estimate (uncertainties in the accuracy of recency assays), and make analysis more rigorous and, thus, persuasive.

Analysis of Secondary Endpoint – the Cost of Certainty

In the proposed counterfactual HIV incidence estimate based clinical trial design, the secondary endpoint compares the incidence of the new treatment to the incidence on the standard-of-care. We need to consider how certain we want to be that the new product is not significantly different than the older product. Dr. Murray presented a scenario in which the external control counterfactual HIV incidence was 4%, and both treatments (standard-of-care and new product) were 90% effective. Whether each arm enrolled 1000 subjects or 2000 subjects and followed for one year, both cases did not substantially change the risk ratio nor the rate difference confidence intervals. In the 1000/arm scenario, we would see up to 6/1000 more infections; in the 2000/arm scenario we would see up to 4/1000 more infections. We need to consider whether this small difference in "certainty" is worth the doubling of the sample size.

Pharmacologic Failure

Pharmacologic (or drug) failures refer to infections occurring in the presence of adequate drug levels – in other words, the drug failed to protect against infection as opposed to failure of protection because of low adherence. Although it is sometimes difficult to determine a "true" drug failure, understanding when and why these happen might help in data interpretation and labeling. The contraception model and the Hormonal Contraceptive Pearl Index[1] serve as an instructive example: the Pearl Index is calculated based on the number of pregnancies occurring *on treatment* – ignoring those after treatment discontinuation – and 28-day cycles during which vaginal intercourse took place without back-up contraception. This rules out adherence-based failure and ensures risk for pregnancy. Looking at data available from DISCOVER, we see that the majority of infections (15/22) occurred in subjects with low drug levels indicating low adherence. All those with high drug levels were presumed infected at baseline. Only one case with "medium" drug levels could potentially be considered a true drug failure. In HPTN083, 4 infections occurred in the cabotegravir arm despite continuous on-time injections, indicating possible "true" drug failures. It will be important to follow these types of evaluations in future trials to better assess the potential of pharmacologic failures to inform trial design, data interpretation, and drug labeling.

[1] FDA. Establishing Effectiveness and Safety for Hormonal Drug Products Intended to Prevent Pregnancy Guidance for Industry July 2019 [Draft Guidance]. Available from: https://www.fda.gov/media/128792/download.

DAY TWO

KEY TAKEAWAYS & RECOMMENDATIONS FROM GLOBAL TOWN HALL

In her summary, Dr. Veronica Miller, Director of the Forum for Collaborative Research, emphasized the importance of such “Global Town Hall” meetings for defining a win: what is a clinically meaningful threshold for efficacy and how precise do we want the comparison to the standard-of-care to be.

The presentations illustrated the different ways of estimating a counterfactual, how some of these principles were applied in earlier studies, and the role the recency assay could play to estimate HIV background incidence. Dr. Miller underlined a few caveats when using a new approach, including the potential difference in risk factors and participant characteristics between HIV-negative and HIV-positive individuals, the importance of screening techniques to minimize bias, and assumptions on prevalence and incidence in a target population.

Dr. Miller also reminded the participants of what Dr. Murray said during his presentation when discussing benchmarks for acceptable risk tolerance in HIV prevention. He said that "we always have to remember that nothing is known with absolute certainty, and even small increments in increasing estimate certainty can be costly in terms of the size or cost of trial, delays in getting data, or disincentives for development of new agents.”

SUMMARY FROM BREAKOUT ROOM 1: ACCEPTABLE THREHOLD FOR EFFICACY FOR MEN WHO HAVE SEX WITH MEN & TRANSGENDER WOMEN

Moderator: Jeffrey Murray, US FDA

Rapporteur: Kimberly Struble, US FDA

Recency assay

The group agreed that although there are some uncertainties about the recency assays and how it will perform in trials, as trying the new methodologies is essential to advance the PrEP field. Using a laboratory test to estimate the background HIV incidence could be a real "win" with the following caveats:

- Need for more validation on recency assay in terms of how the assay performs with different clades and populations.
- Need for a solution on how best to address the recency assay and background incidence "mismatches" if they occur. For example, if the new product is held to the 2% incidence estimated using the recency assays vs. 4% of background incidence that the trial was powered for, it may not show at least 50% efficacy.

- The incidence measured by recency assay is retrospective, whereas it is measured prospectively in the trial. Those who are HIV infected may be systematically different from those who are not infected, even if they were drawn from the same population. Continued discussion on how this can be factored into different scenarios is essential.

Counterfactual Estimate & Acceptable Efficacy Threshold
On the question of what threshold is a "reasonable" standard to control for uncertainty and the minimum amount of efficacy we are willing to accept, participants suggested to discount a lot. For example, using an external control with a fraction (about 50%) of the lower 95% CI as the upper 95% CI of the observed trial infection rate to define a "win".

- The group expressed concerns about heterogeneity over time between populations, within the same city, clinic, etc.

Accounting for external control counterfactual assumptions for different subgroups
The group elaborated on possible solutions to account for counterfactual assumptions that may be less precise in certain subgroups (e.g., low number of events in women in the US, estimates in trans-gendered women, 15-18 year-olds, etc.) Suggestions included:

- Enrich trials for the major risk groups such as MSM and cisgender women
- Ensure high-risk by enrollment criteria
- Not possible to get endpoints in all subgroups, but powering the trial sufficiently and doing some extrapolation on subgroups is possible when looking at trends
- Balance inclusiveness with the ability to show efficacy. This can include low-risk for safety and tolerability concerns but only analyze the highest risk to get the efficacy.

More Choice Data
The discrete choice survey (soon to be published) interviewed 600 women in South Africa, regarding implantable PrEP, revealed that the effectiveness (70% or greater) was the essential characteristic for these women. This comment led to the realization that more choice data from different populations are needed for informed decisions.

Pearl Index
The group cautioned not to overemphasize the Pearl Index when comparing it with counterfactual incidence for HIV, given a clear distinction between sex that leading to pregnancy occurring more frequently vs. sex leading to HIV infection, a life-long condition.

SUMMARY FROM BREAKOUT ROOM 2: ACCEPTABLE THRESHOLD FOR EFFICACY FOR CISGENDER WOMEN

Moderator: Raphael Landovitz, UCLA Center for Clinical AIDS Research & Education

Rapporteur: Charu Mullick, US FDA

The group agreed that given the urgent need to expand HIV prevention modalities portfolio and non-inferiority trials being beyond the realm of possibilities for women, the

counterfactual approach design should be considered a path forward. Participants were also open to some uncertainty, noting the balance/tradeoff between levels of certainty and repercussions with delay or disincentivizing drug development and approval of an effective product.

Recency Assay

While the group acknowledged some uncertainties of recency assays, stakeholders agreed to move forward with the approach and "learn as we go," noting that the multiple methods to assess the placebo or background incidence will help offset or minimize the extent of uncertainty. The recency assay limitations discussed in this group included differential HIV screening standards across regions, inherent variability related to clades and subtypes. The group agreed that multiple methods to assess the counterfactual placebo HIV incidence estimate would off-set or at least minimize the extent of uncertainty.

Acceptable Efficacy Threshold

The efficacy threshold question was embedded in the overall context of the product. In other words, what other advantages does a specific product offer? The group agreed that positioning the choice and willingness to trade off some level of efficacy would be a good way forward because the efficacy metric is only one factor and not the only factor. When doing the overall assessment, stakeholders noted the importance of not thinking of an efficacy threshold as the only determinant and consider other attributes such as ease of use, acceptability advantage, fill a niche, etc. Two possible scenarios for new products were considered:

- Scenario 1: Easy to use, other advantages/attributes AND at least 50% relative reduction (50% or lower excluded from the confidence interval)
    - This scenario would be reasonable
- Scenario 2: Daily oral pill, no other advantages or attributes, AND 50% or lower excluded from the confidence interval
    - This scenario may not be acceptable

Reliance on Critical Ancillary Data

The group discussed the value of additional data to assist in interpreting trial outcomes:

- Knowledge of drug levels
- Deep dive into trial seroconversion narratives
- Knowledge of people's preference and behavior data

Pharmacologic vs. Adherence Failure

The group exhibited some degree of ambivalence about how this would eventually be important from the regulatory and practical perspective and how it can be accomplished within a trial. Participants, however, acknowledged that this is a scientifically interesting and important topic.

SUMMARY FROM BREAKOUT ROOM 3: METHODS FOR DERIVING EXTERNAL CONTROL COUNTERFACTUAL ESTIMATE PROTOCOL DESIGN

Moderator: Sinead Delany-Moretlwe, University of the Witwatersrand

Rapporteur: Amy Cutrell, ViiV Healthcare

The group reached an overarching agreement that counterfactual estimate would provide valuable information, noting that no counterfactual estimate of the HIV incidence in the absence of PrEP can or will be perfect. Given the imprecision of many counterfactual estimates, the group suggested evaluating multiple data sources (consensus was at least two). Although no clear favorites emerged as to which data sources to use:

- The guiding principle should always be that the data be cotemporaneous, from a population as similar as possible to the trial population.
- The trials be conducted in high HIV incidence environments recognizing the ethical considerations of PrEP availability after the trial concludes.
- Stakeholders also suggested putting in place the decision framework for interpreting multiple data sources that provided the selected counterfactual estimates.
- Outlining the pros and cons ahead of time will be very important so that when the data is collected and analyzed, it is possible to weigh the evidence from these various sources.

Screening

The group considered how we need to change the approach to screening participants for the next generation of PrEP clinical trials: for the recency assay-based estimate, for example, participants will need to be screened, irrespective of their HIV status.

Data Sources

The group elaborated on potential data sources, discussing their advantages and disadvantages.

- Observational cohorts may provide longitudinal estimates. However, they can quickly become complicated by participants potentially cycling on and off PrEP during their participation in the cohort.
- Adherence data, which is more advantageous than historical data, but population and products need to be already established.
- Synthetic cohorts could also serve as a potential data source.

Recency Assay

The group agreed that the recency assay is a promising data source, but given its "newness", there is a noticeable learning curve. The group suggested looking into attributes and parameters, including refinement of the viral load threshold, to be explored.

GENERAL DISCUSSION

Is the counterfactual approach for trials testing new PrEP products ready for prime time? In essence, yes. In fact, F/TAF post-marketing commitments included a study in women using at least two methods to construct an external HIV incidence estimate because we are lacking historic trial data on which to base a non-inferiority design. External controls work best for products with high efficacy, which we expect to see with new PrEP products. We need an approach that meets both ethical and scientific rigorous standards. The external control counterfactual design is as re-assuring as a non-inferiority trial in which we are externally referencing something from a different time and place. Even if we were able to do a classic non-inferiority or superiority trial, we would still want to be assured that the trials enrolled the right population to be able to conclude that the low incidence on treatment is due to the protective effect of the drug and not just because the enrolled population happened to be at low risk for HIV infection.

Would this model also work for immune based approaches? Monoclonal antibodies should be considered a "drug" in the sense that they prevent viral entry, so the counterfactual approach would be appropriate for their evaluation. Vaccines studies tend to be much larger than drug studies and appropriate prevention tools need to be accessible for trial participants. The information on cross-sectional incidence gained through drug studies could be very informative for vaccine studies as well. If PrEP moves to a one annual injection option, some might see this as similar to a vaccine product, although the difference between them (a long-acting circulating drug vs. a period of interaction with the immune system leaving a lasting immune effect) remains. As we continue the journey towards more PrEP options, we need to think beyond "efficacy" and "mode of delivery" to a more comprehensive view across different variables.
Looking at 2 years down the road: what would the comparator arm be? Considering that prevention studies will continue to be multi-center/international, the "double-blind, double-dummy" design may not be feasible since different countries may have different standards-of-care (e.g., oral daily PrEP, cabotegravir, dapivirine ring). Open label studies, with standard-of-care based on choice, would be one way to move forward. This approach would generate a complex set of data, but provide a wealth of information, perhaps allowing for some informative sub-group analyses. This is an issue that deserves more discussion in future meetings.

Given that the primary analysis will be against the counterfactual placebo HIV estimate, will we always need an internal active control? If we take adherence out of the picture, we should be able to approach a rabies prophylaxis type of paradigm, where we expect a very low rate of infection. If we enroll a high-risk HIV exposed population and drive the infection rate down close to zero, we may not need an internal active control. It would be a U=U proposition.

ETHICAL CONSIDERATIONS

HIV prevention research is no stranger to ethical challenges and larger problems of social justice and vulnerabilities. We look at these in the context of our moral obligation for scaling up effective means of HIV prevention and range of options. “We have to get this right” said Dr. Jeremy Sugarman from Johns Hopkins Berman Institute of Bioethics “It’s not just about science and statistics; it’s also about how these debates can change the contours of the pandemic and the research paradigm”.

Our ethical considerations are embedded in plenty of existing guidance and policy, from the Nuremburg Code to the recently published UNAIDS and HPTN guidances for HIV prevention research. We have been, and continue to need to be, responsive to successes and failures in research. The HPTN Ethical Guidance for Research[2] lays out guidance points across the continuum of the research timeline – before research, during research and after research (2).

Some have proposed the option of enrolling individuals for whom a product is not appropriate (e.g., medically contra-indicated, strong dislike, or substantial personal barriers), but this approach has its own set of issues, including ensuring that the expression of unacceptability is authentic, and recognizing that the people enrolling into such a placebo arm are qualitatively different from those who have a simple preference for a prevention method under investigation.

The counterfactual estimate presents a potential way forward. Although multiple approaches for the counterfactual estimate are being considered. Dr. Sugarman reminded us that none on their own merit reflect the “gold standard” – the gold standard being the randomized, placebo-controlled trial.
However, as discussed earlier in the meeting, in situations in which the placebo is no longer ethical, we need to rely on superiority or non-inferiority designs, in which the “placebo” is referenced from an earlier trial often carried out in a different place. Although these trial designs have the advantage of randomization, they leave open the question of whether the population enrolled was sufficiently “at risk” for HIV infection without a good understanding of the HIV incidence in the community. No matter what the trial design, there will be issues around what the comparator arm should be as new products are approved and what the standard of prevention is in different regions.

Clinical research must be scientifically sound and provide social value to be considered ethical. Dr. Sugarman encouraged participants to consider the potential limitations of particular counterfactual estimates, address them, and commit to unbiased scientific and regulatory review. To meet the social value criterion, the study should be likely to produce information directly relevant for understanding or intervening on a significant health problem or because of its expected contribution to research likely to promote public health (3). Social value includes outcomes such as producing convincing data for potential

[2] Brown B, Sugarman J. HPTN Ethics Guidance for Research 2003 [updated February 26, 2020. Available from: https://www.hptn.org/sites/default/files/inline-files/HPTNEthicsGuidanceDocument_2.26.20.pdf.

users, potential prescribers, public health authorities, ministries of health and regulators. Research that has high social value will include transparent stakeholder engagement and ethics approval. Given the novelty of the external control counterfactual approach, institutional review boards will need to be involved in discussions to understand the new concepts.

The counterfactual approach brings potential obligations to those who become the "external control". Those who test positive need to be linked to care. The issue of whether recency assay testing results should be disclosed to the individual and/or their caregiver is being debated. Most of all, the informed consent process needs to be robust.

Continuing on the issue of placebo-controlled trials, some participants stated that any placebo arm, including the deferred treatment study design in unethical. The concern is that just not liking to take pills may not be sufficient motivation to participate in complicated trials. Some participants thought that for new products with such different formulations (e.g., 6 monthly injections), a placebo-controlled trial would not be problematic as long as the standard-of-care includes daily oral PrEP be made available, since these "minimal interventions" would open the door to participants who would never agree to daily oral PrEP. The crucial aspect in this approach is that the choice be truly authentic.

Another issue raised is that of equipoise. We talk about "expected efficacy" of new PrEP products. Our expectations are based on pre-clinical data (*in vitro* and animal model) and understanding the mode-of-action of antiretroviral drugs. In some cases, we have a lot of convincing treatment experience to contribute to our expectations. However, as we move away from antiretroviral-based prevention to broadly neutralizing antibodies and vaccines, our expectations are less clear. How certain do we need to be that an investigational agent works in order not to provide effective HIV interventions to all participants? How long will we be able to do placebo-controlled trials for HIV vaccines or broadly neutralizing antibodies?

Participants discussed the issue of giving recency assay results to patients. Giving the individual level results could cause harm, especially for women[3]. By setting a timeline around the time of infection (e.g., in the last year), we are imparting information beyond the individual screening for the trial to those in their networks. In terms of linking HIV-infected individuals to care, the most useful diagnostic is the viral load, rather than the recency test results, since the "recent" status would not change clinical management. On the other hand, should individuals not have the right to this information? Is there an ethical problem with withholding information? The HIV diagnosis in itself can cause harm, and a diagnosis of acute HIV infection would also have to be given to the participant. It is important to remember that the LAg assay is not an approved or licensed diagnostic. It is being used as a surveillance tool for early detection of transmitting networks and to target index testing, and not meant to diagnose individual people as either "recent" or "not recent". The PEPFAR Scientific Advisory Board addressed this issue but did not formulate

[3] For more implications of index and recency testing on HIV criminalization and criminalized populations, see:
New HIV Testing Strategies in PEPFAR COP19: Rollout and Human Rights Concerns 2019 [Issue Brief]. Available from:
https://www.amfar.org/uploadedFiles/_amfarorg/Articles/On_The_Hill/2019/COP19.pdf.

blanket recommendation citing the diversity in epidemic typology, magnitude, populations at risk; preparedness of users and providers; and human rights and ethical considerations(4). The UK has significant experience in returning recency assay results to patient, with no documented adverse events ensuing as a result. RITA is recommended as standard-of-care for all new HIV diagnoses, cautioning that the results need to be interpreted in the context of history of risk. New patients appreciate having a rough idea whether they have been HIV infected for years, or not.

Another option is to engage representatives from the community to share results on a community rather than individual level, since the group comprising the "external control" is contributing to product development. Participants agreed that this issue requires more discussion and deliberation at future meetings.

Community members emphasized the need to better understand the recency assay and how it would be used in clinical research to facilitate their full engagement in these discussions. "We need robust community literacy of all the ins and outs of recency and disclosure… it will be different for different communities". Furthermore, it would be most helpful to have an opportunity to fully review the counterfactual based clinical trial designs being considered and approved, in real time, if sponsors would be willing to engage in such discussions.

CONCLUSION

Bringing new means of HIV prevention products to market has been hampered paradoxically by success of first-generation PrEP products, necessitating innovation in clinical trial design. The goal for the two-day webinar on *Protocol Design Consideration: Analyses for Efficacy* was to discuss the external control counterfactual design and obtain broad public input for PrEP study design issues by answering critical questions, such as what a clinically meaningful threshold for efficacy is and how precise the comparison to the active control should be to declare a "win".

The workshop invited stakeholders (see Appendix) such as regulators, healthcare professionals, academicians, industry, and community representatives across sectors and regions to discuss the counterfactual design. After hearing presentations from experts and participating in panel discussions, participants agreed that the external control counterfactual clinical trial design approach is a reasonable path forward to ensure additional PrEP choices, increase PrEP uptake and adherence, and reduce HIV incidence in the communities with the caveat that we need to collectively "learn as we go" and ensure ethical principles for rigorous science, social value and minimizing risk to participants remain at the forefront.

One method the meeting primarily focused on to construct the external control counterfactual estimate (i.e., background HIV incidence) was the LAg assay discriminating "recent" vs. "non-recent" infection. Although we do not yet have experience on the use of the assay in the context of clinical trials and regulatory approval, this

approach is already being applied in new development programs by Gilead and Merck. The approach appears to be feasible, especially regarding sample size, offers high statistical power, and can produce contemporaneous data that is similar to the population in the study since collecting and testing of samples from HIV-infected individuals can continue throughout the trial duration. Combining this approach with additional external and/or active controls to facilitate the robust interpretation of the study results will be essential.

Two topics stood out for further discussion: 1) recency assay and disclosure of results and 2) evolving standard-of-care. It is clear we need continued stakeholder engagement and link to regulatory authorities in the countries in which these trials are being done.

HIV FORUM SPONSORS

The HIV Forum is funded in part by the:

BILL & MELINDA GATES foundation

and by:

Abbott

abbvie

GILEAD

janssen

MERCK
INVENTING FOR LIFE

Monogram BIOSCIENCES
LabCorp Specialty Testing Group

ViiV Healthcare

APPENDIX

Table 4: List of Participants and Organization

Abraham Johnson	TAG
Adrian Cornejo	UC Berkeley
Aimee Hodowanec	FDA
Alex Kintu	Gilead Sciences, Inc.
Alex Welte	Stellenbosch University
Amy Cutrell	ViiV Healthcare
Andrew Topp	AbbVie
Andrii Chernyshev	ALLIANCE GLOBAL
Angela Snyder	University of Cincinnati
Anja Schiel	Norwegian Medicines Agency
Ann Duerr	Fred Hutchinson Cancer Research Center
Anthony Lamarca	THERAFIRST MEDICAL CENTER
Anthony Mills	Men's Health Foundation
Anusha Govind	UT southwestern
Astrud Reed	UC
Athena Kourtis	CDC
Ayana Elliott	Gilead Sciences, Inc.
Beatriz Grinsztejn	Fundação Oswaldo Cruz - FIOCRUZ
Benjamin Lorenz	FDA
Bernadette Ng'eno	Merck Sharp & Dohme Corp.
Bharat Parekh	CDC
Boitumelo Semete-Makokotlela	South African Health Product Regulatory Authority
Brenda Rodriguez	Forum for Collaborative Research
Brian Palmer	Gilead Sciences, Inc.
Camtu Nguyen	Gilead Sciences, Inc.
Carl Fichtenbaum	University of Cincinnati
Caryn Morse	Wake Forest University Health Sciences
Catherine Slack	UKZN
Cathy Chien	Gilead Sciences, Inc.
Charu Mullick	FDA
Chilufya K Hampongo	Treatment Advocacy and Literacy Campaign
Chris Nguyen	Gilead Sciences, Inc.
Christa Fischer Walker	FHI 360
Christine Heumann	Indiana University
Christoph Carter	Gilead Sciences, Inc.

Claudia Silva	MSD
Connie Celum	University of Washington
Cornelius Van Dam	Cone Health, RCID
Craig Dietz	KC CARE HEALTH CENTER
Cynthia Brinson	CTCR
Dagna Laufer	IAVI
Daisy Ouya	AVAC
Damon Deming	FDA
Danielle Campbell	Community
David Andrist	The Ohio State University ACTG
David Magnuson	Gilead Sciences, Inc.
Dawn Smith	CDC
Daya Moodley	University of KwaZulu Natal
Deborah Donnell	Fred Hutchinson Cancer Research Center
Debra Birnkrant	FDA
Dimitar Tonev	HCV Research UK
Eduard Grebe	Vitalant Research Institute
Elizabeth McGrory	Independent
Elizabeth Russell	Merck & Co, Inc
Erica Crittendon	PCAF
Eunice Ndzerem-Shang	FRC
Fei Gao	Fred Hutchinson Cancer Research Center
Filip Kukulski	Health Canada
Frances Cowan	Liverpool School of Tropical Medicine
Francesca Day	EMA
Garland Lee	Gilead Sciences, Inc.
Grace Kumwenda	Pakachere IHDC
Gus Cairns	NAM / Aidsmap; PrEP in Europe; EATG
Gustavo Doncel	CONRAD
Helen Rees	Wits RHI
Hema Kapoor	Quest Diagnostics
Hengrui Sun	FDA
Ian Frank	University of Pennsylvania
indira brar	Henry Ford Hospital
Jaasiel Chapman	UC Infectious Diseases Research
James McGuire	Merck & Co, Inc
James Rooney	Gilead Sciences, Inc.
Jared Baeten	Gilead Sciences, Inc.
Jason Hindman	Gilead Sciences, Inc.

Javier R. Lama	Asociacion Civil Impacta Salud y Educacion
Jean Lee	Gilead Sciences, Inc.
Jean Marie Arduino	Merck & Co., Inc.
Jeffrey Murray	FDA
Jennifer DeMorin	Merck & Co, Inc
Jeremy Sugarman	Johns Hopkins Berman Institute of Bioethics
Jessica Salzwedel	AVAC
Joerg Zinserling	BfArM
Jorge Gallardo-Cartagena	CITBM
Joseph Lau	Forum for Collaborative Research
Joy West-Blondin	Quest Diagnostics
Jules O'Rear	US Food and Drug Administration
Kagisho Baepanye	HVTN/CoVPN
Karam Mounzer	Philadelphia FIGHT
Karen T.Cuenco	BMGF
Karla Tafur	CITBM
Kate Lawrence	Gilead Sciences, Inc.
Katherine Watson	Quest Diagnostics
Kathleen Squires	Merck Research Labs
Kenneth Mayer	Fenway Health/ Harvard Medical School
Kenneth Mugwanya	University of Washington
Kenyon Farrow	Partners for Dignity & Rights
Kimberly Struble	FDA
Kirk Chan-Tack	FDA
Lara Lewis	CAPRISA
Leila Mansoor	CAPRISA
Li Tao	Gilead Sciences, Inc.
Linda Akunne	US FDA
Linda Fredrick	AbbVie
Linden Lalley-Chareczko	Philadelphia FIGHT Community Health Centers
Lisa Naeger	FDA
Louis Shackelford	HIV Vaccine Trials Network
Luis Javier Hernandez	Forum for Collaborative Research
Lusine Ghazaryan	USAID
Lut Van Damme	BMGF
Lynda Dee	AIDS Action Baltimore
Mallory Rowell	OSU Wexner Medical Center
Manjeetha Jaggernath	Match Research Unit

Marco Pompei	MSD
Maria Jesus Fernandez Cortizo	Spanish Agency on Medicines
Maribel Gonzalez	EMA
Mario Chen	FHI 360
Mario Guerrero	Lundquist Institute at Harbor-UCLA Medical Center
Mark Barnes	Ropes & Gray, LLP
Mark Bernstein	Gilead Sciences, Inc.
Mary Latka	USAID
Mary Singer	FDA
Matshidiso Morolo	Setshaba Research Centre
Matthew Carabasi	Merck Research Labs
Michael Busch	Vitalant Research Institute
Michael Robertson	Merck & Co, Inc
Michel Alary	CHU de Québec - Université Laval
Michelle Rodolph	WHO
Mitchell Warren	AVAC
Moises Huaman	University of Cincinnati
Moupali Das	Gilead Sciences, Inc.
Nandisile (Nandi) Luthuli	AVAC
Navita Jain	AVAC
Neil Parkin	Data First Consulting
Nelly Mugo	Kenya Medical Research Institute
Nicholas Murdock	UC Berkeley
Nikos Dedes	Positive Voice
Nina Russell	BMGF
Nittaya Phanuphak	Institute of HIV Research and Innovation
Nonhlanhla Yende-Zuma	CAPRISA
Ntando Yola	Desmond Tutu Health Foundation
Oliver Laeyendecker	NIAID
Pai-Lien Chen	FHI 360
Pamela Wong	Gilead Sciences, Inc.
Patricia Mayer-Brennan	IPM
Patricia Mendez	Immunocore
Peggy Hwang	Merck & Co, Inc
Pervin Anklesaria	BMGF
Peter Godfrey-Faussett	UNAIDS
Peter Miele	FDA

Peter Ruane	Ruane Clinical Research
Peter Sklar	Merck & Co, Inc
Poonam Mishra	US FDA
Ramin Ebrahimi	Gilead Sciences, Inc.
Ramiro Correa	The Lundquist Institute
Raphael Landovitz	UCLA Center for Clinical AIDS Research & Education
Regine Lehnert	BfArM
Rieke van der Graaf	UMC Utrecht
Robert Allison	FDA
Roger Tatoud	IAS
Romano Baroni	Ararat Research Center
Rosalinda Graci	Gilead Sciences, Inc.
Rosina Phate-Lesihla	Ministry of Health
Roweena Corpuz	Henry Ford Health System
Russell Fleischer	FDA
Sakhile Dube	SADC MRH
Sandra Lovell	AbbVie
Sarah Connelly	FDA
Sarah Read	NIAID
Sean Collins	Gilead Sciences, Inc.
Shannon Allen	USAID
Shapo Pitsi	Setshaba Research Centre
Sharon Hillier	University of Pittsburgh
Sharon Kohrs	UC Infectious Diseases Research
Sheena McCormack	MRC CTU at UCL
Simon Collins	HIV i-Base
Sinead Delany-Moretlwe	Wits RHI
Siobhan O'Connor	CDC
Stanley Wang	AbbVie
Stephanie Buchholz	BfArM
Stephanie Cox	Gilead Sciences, Inc.
Stephanie Troy	FDA
Steven Wakefield	Private citizen
Susan Buchbinder	San Francisco Department of Public Health
Swarup Mehta	Gilead Sciences, Inc.
Takuma Matsuda	Gilead Sciences, Inc.
Tamar Tchelidze	Forum for Collaborative Research
Tamara Ward	University of Cincinnati

Tendai Chipepera	Setshaba research center
Thamban Valappil	FDA
Timothy Jancel	FDA
Timothy Mastro	FHI 360
Tinkhani Mbichila	Forum for Collaborative Research
Toni Sparrow	Gilead Sciences, Inc.
Tulika Singh	Bios clinical research
Vani Vannappagari	ViiV Healthcare
Veronica Miller	Forum for Collaborative Research
Victoria Mason	Forum for Collaborative Research
Virginia Sheikh	FDA
Vladimir Berthaud	MEHARRY MEDICAL COLLEGE
Wayne Duffus	Gilead Sciences, Inc.
Wen Zeng	FDA
Wendy Carter	FDA
YaPei Liu	Gilead Sciences, Inc.
Yodit Belew	FDA
Yohance Whiteside	Merck & Co, Inc
Yongwu Shao	Gilead Sciences, Inc.
Yoshihiko Murata	ViiV Healthcare
Zeda Rosenberg	IPM

REFERENCES

1. Glidden DV. Statistical approaches to accelerate the development of long-acting antiretrovirals for HIV pre-exposure prophylaxis. Curr Opin HIV AIDS. 2020;15(1):56-60.
2. Brown B, Sugarman J. HPTN Ethics Guidance for Research 2003 [updated February 26, 2020. Available from: https://www.hptn.org/sites/default/files/inline-files/HPTNEthicsGuidanceDocument_2.26.20.pdf.
3. International Ethical Guidelines for Health-related Research Involving Humans Geneva: Council for International Organizations of Medical Sciences (CIOMS); 2016 [Fourth Edition:[
4. Karim QA, Gruskin S, Lockett L, Macklin R, Maxwell C, Mayer K, et al. PEPFAR SAB Recency Testing Expert Working Group: Recommendations on Communicating Individual Results from Recency Testing U.S. Department of State2020 [Available from: https://www.state.gov/wp-content/uploads/2020/07/Recency-testing-recommendations-scaled-1.pdf.