Article

Modeling the Probability of HIV Infection over Time in High-Risk Seronegative Participants Receiving Placebo in Five Randomized Double-Blind Placebo-Controlled HIV Pre-Exposure Prophylaxis Trials: A Patient-Level Pooled Analysis

Maria Garcia-Cremades [1,*,†] , Craig W. Hendrix [2] , Priya Jayachandran [1] , Natasha Strydom [1] , Leah Jarlsberg [1] , Robert Grant [3] , Connie L. Celum [4] , Michael Martin [5,6] , Jared M. Baeten [4] , Jeanne Marrazzo [7] , Peter Anderson [8] , Kachit Choopanya [9] , Suphak Vanichseni [9] , David V. Glidden [10] and Radojka M. Savic [1]

[1] Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA 94143, USA
[2] Division of Clinical Pharmacology, Department of Medicine, Johns Hopkins University, Baltimore, MD 21287, USA
[3] Department of Medicine, University of California San Francisco, San Francisco, CA 94143, USA
[4] Departments of Global Health, Medicine, and Epidemiology, University of Washington, Seattle, WA 98104, USA
[5] Division of HIV/AIDS Prevention, Centers for Disease Control and Prevention, National Center for HIV/AIDS, Viral Hepatitis, STD, and TB Prevention, Atlanta, GA 30333, USA
[6] Thailand Ministry of Public Health—US CDC Collaboration, Nonthaburi 11000, Thailand
[7] Division of Infectious Diseases, University of Alabama at Birmingham Medical Center, Birmingham, AL 35294, USA
[8] Department of Pharmaceutical Sciences, University of Colorado, Denver, CO 80045, USA
[9] Bangkok Tenofovir Study Group, Bangkok 10600, Thailand
[10] Department of Epidemiology and Biostatistics, University of California San Francisco, San Francisco, CA 94143, USA
* Correspondence: margar77@ucm.es
† Current address: Departamento de Farmacia Galénica y Tecnología Alimentaria, Facultad de Farmacia, Universidad Complutense de Madrid, 28040 Madrid, Spain.

Abstract: The World Health Organization recommends pre-exposure prophylaxis (PrEP) for individ-
uals at substantial risk of HIV infection. The aim of this analysis is to quantify the individual risk
of HIV infection over time, using a large database of high-risk individuals ($n = 5583$). We used data
from placebo recipients in five phase III PrEP trials: iPrEx, conducted in men who have sex with men
and transgender women; VOICE, conducted in young women at high sexual risk; Partners PrEP,
conducted in HIV serodiscordant heterosexual couples; TDF2, conducted in high-risk heterosexual
men and women; and BTS, conducted in persons who inject drugs. The probability of HIV infection
over time was estimated using NONMEM7.4. We identified predictors of HIV risk and found a
substantial difference in the risk of infection among and within trial populations, with each study
including a mix of low, moderate, and high-risk individuals ($p < 0.05$). Persons who were female at
birth were at a higher risk of HIV infection than people who were male at birth. Final models were
integrated in a tool that can assess person-specific risk and simulate cumulative HIV risk over time.
These models can be used to optimize future PrEP clinical trials by identifying potential participants
at highest risk.

Keywords: HIV prevention trials; key and vulnerable populations; risk factors; risk phenotypes;
modeling

1. Introduction

In 2020, the World Health Organization (WHO) estimated that 37.7 million people
were living with human immunodeficiency virus (HIV), 1.5 million had newly acquired

HIV, and 608,000 people with HIV had died [1]. While new infections and HIV-related deaths have considerably declined between 2000 and 2020, the HIV pandemic continues to be a major contributor to the global burden of disease [1].

Antiretroviral therapy (ART) reduces the risk of HIV infection when administered as pre-exposure prophylaxis (PrEP) and post-exposure prophylaxis (PEP) [1]. The effectiveness of PrEP at reducing HIV transmission has been demonstrated in several randomized controlled trials (RCTs) [2–10], and its use has been increasing globally, with more than 600,000 people in 76 countries worldwide having received oral PrEP at least once in 2019 [11]. However, there is still considerable room for improvement in the widespread use of PrEP in those at high risk of infection, and even in settings where access to and use of PrEP is high, adherence is a challenge and varies among different populations [12].

Defining the risk of HIV infection remains a challenge and limits efforts to direct HIV prevention services, including PrEP, to people who would benefit most. The WHO recommends PrEP to populations with a substantial risk for acquiring HIV, which is defined as an HIV incidence > 3 per 100 person-years. However, HIV risk varies within populations based on individual characteristics and behaviors [13]. If the characteristics of high-risk individuals and predictors of HIV infection within key populations can be identified, PrEP provision could be better directed to populations who would benefit most, thereby reducing HIV transmission. Likewise, the design of future PrEP clinical trials could be optimized by selecting individuals at highest risk of HIV infection with greater anticipated incidence.

Previous studies used data from several PrEP randomized controlled trials (RCTs) to identify predictors of HIV infection and develop heterogeneous risk scoring tools in specific populations such as men who have sex with men (MSM), heterosexual serodiscordant couples, and heterosexual women [14–17].

In this study, we aimed to characterize and compare the probability of HIV infection within high-risk populations using data from the placebo arms of five PrEP clinical trials: the Pre-exposure Prophylaxis Initiative (iPrEx) trial [2], the Vaginal and Oral Interventions to Control the Epidemic (VOICE) trial [3], Partners Pre-exposure Prophylaxis (Partners PrEP) trial [4], the Bangkok Tenofovir Study (BTS) [5] and the Botswana TDF/FTC Oral HIV Prophylaxis Trial (TDF2) trial [6]. By quantifying the risk factors within each population, we attempted to identify subgroups at low, medium, and high risk of acquiring HIV in each population. With this quantitative understanding of HIV risk in key populations based on individual characteristics and risk behaviors, we may be able to predict HIV incidence in the absence of intervention and optimize future HIV prevention clinical trials by enrolling the individuals at highest risk.

2. Materials and Methods

2.1. Database: Clinical Trials and Study Design

We included participants enrolled in the placebo arms of five phase III clinical trials that evaluated the efficacy of daily TDF, administered with or without FTC, as PrEP in populations at high risk of HIV infection. These RCTs were conducted, between 2005 and 2013, among: MSM/transgender women (TGW) (iPrEx, NCT00458393) [2], young heterosexual women at high sexual risk (VOICE, NCT00705679) [3], heterosexual men and women in HIV serodiscordant partnerships (Partners PrEP, NCT00557245) [4], people who inject drugs (Bangkok Tenofovir Study [BTS], NCT00119106) [5], and heterosexual men and women at high risk (TDF2, NCT00448669) [6]. Trial protocols were reviewed and approved by local and institutional Ethical Committees, and all participants provided written informed consent. Trials included in this analysis were double-blind, with neither the participants nor the researcher knowing participants were receiving placebo or active agent. The randomization process was specific for each trial and can be found in the trial primary publications [2–6]. Overall, in all the trials, study visits were scheduled every month after enrollment, which included drug dispensation, adherence counseling and monitoring, medical history, and HIV status tests. For this study, we included data from trial participants who were HIV negative at enrollment and had the date of enrollment

and the date of HIV infection or the date their data were censored in the trial database. We collected longitudinal HIV test results (collected monthly), and baseline demographic characteristics, sexual transmitted infections (STIs), and risk behavior (Table 1).

Table 1. Population characteristics and risk factors of participants in five pre-exposure prophylaxis studies.

Study	Population	Geographical Region	Demographics	Sexually Transmitted Infections	Sexual Risk Behavior	Other Risk Factors
iPrEx	MSM/TGW	- Peru - Ecuador - South Africa - Brazil - Thailand - US	- Age - Education - Ethnicity - Sex at birth - Marital status - Transgender	- HSV2 - Gonorrhea - Chlamydia - Syphilis	- No. of partners - No. of intercourses - Protective intercourse - Receptive/insertive partner	- Partner's HIV status
VOICE	Young women at high sexual risk	- South Africa - Uganda - Zimbabwe	- Age - Education - Ethnicity - Financial support (by partner) - Sex at birth - Marital status - N° live birth	- HSV1-2 - Gonorrhea - Chlamydia - Syphilis - Trichomonas	- No. of partners - No. of intercourses - Protective intercourse - Type of intercourse (vaginal/anal)	- Partner's HIV status - Partner circumcised - Partner has other sexual partners
Partners PrEP	HIV negative partner in HIV-serodiscordant heterosexual couples	- Kenya - Uganda	- Age - Education - Sex at birth - Income - Marital status - N° Children - Pregnancy	- HSV2 - Gonorrhea - Chlamydia - Syphilis - Trichomonas	- No. of partners - No. of intercourses - Protective intercourse	- Partner's viral load - Male circumcision
BTS	People who inject drugs	- Bangkok (Thailand)	- Age - Education - Ethnicity - Sex at birth		- No. of partners - No. of intercourses - Protective intercourse - Frequency anal intercourse	- Drug injection frequency - Live with partner
TDF2	Heterosexual men/women at high risk	- Botswana	- Age - Education - Ethnicity - Sex at birth - Marital status	- Gonorrhea - Chlamydia - Syphilis - Trichomonas	- No. of partners - No. of intercourses - Protective intercourse	- HIV positive partner

Sexual risk behavior factors were recorded differently between studies (see Table S1a–e in Supplementary Materials).

2.2. Data Analysis

Primary analysis consisted of characterizing the probability of acquiring HIV infection over time of initially seronegative individuals and identifying predictors of infection through parametric survival analysis [18] using NONMEM7.4 [19] with the LAPLACE estimation method. This probability was treated as a right-hand censored time to event variable. The primary outcome was time to HIV seroconversion of initially seronegative individuals, which is defined as the time from the date of enrollment to the date of seroconversion and characterized using a parametric time to event model. These models are characterized by a hazard (hz) that can be parameterized using different distribution

functions. The link between the hz and the overall survival probability (OS) (defined as the probability of remaining HIV uninfected) is established through the cumulative hazard (HZ) over time (OS = e-HZ) [18]. The identification of the different predictors of HIV infection was performed by covariate analysis in the hz baseline parameters. The covariates included in the analysis were the population characteristics and risk factors of participants collected at study baseline. Hazard ratios (HR) were computed for each covariate included in the final multivariate model to compare the effect magnitude of each predictor on the relative risk. The HR was computed by calculating the ratio of the baseline hazard of group A (e.g., No use of condom) to group B (e.g., Use of condom). For continuous predictors, HR was calculated with respect to the median value of the population (e.g., being 20 years old increases the risk of contracting HIV with respect to being 24 years old in the iPrEx study). Internal model evaluation was performed by simulation-based diagnostics using Kaplan–Meier Visual Predictive Check (VPC) plots [20,21] stratified by significant predictors. Detailed explanation of model building and evaluation is included in the Supplementary Methods. Exploratory analysis by Kaplan–Meier plots stratified by the different covariates and univariate Cox-regression analysis was performed to identify preliminary potential predictors of HIV infection and to support the final model.

2.3. Risk Stratification Using Final Multivariate Model

Risk stratification was performed based on a target probability of remaining HIV uninfected at 1 year of follow up (52 weeks). We defined low risk as <1% probability of HIV infection, moderate risk as 1% to 5% probability, and high risk as >5% probability. A risk score (RS) corresponding to these probabilities was calculated by integrating the survival function (RS1% = $-(\log(OS99\%)/52)$). Individual risk scores were computed based on the significant predictors using the baseline hazard for the final model describing time to HIV infection. These computed risk scores were used to distribute individuals into the risk subgroups (low, moderate, high).

2.4. Interactive Web Application

The final models were coded using the mlxR package (https://cran.r-project.org/package=mlxR, accessed on 1 July 2019) and were incorporated into an interactive web application using the R package shiny (https://shiny.rstudio.com, accessed on 1 July 2019) to perform interactive simulations of individual HIV risk profiles within each key population based on individual predictors.
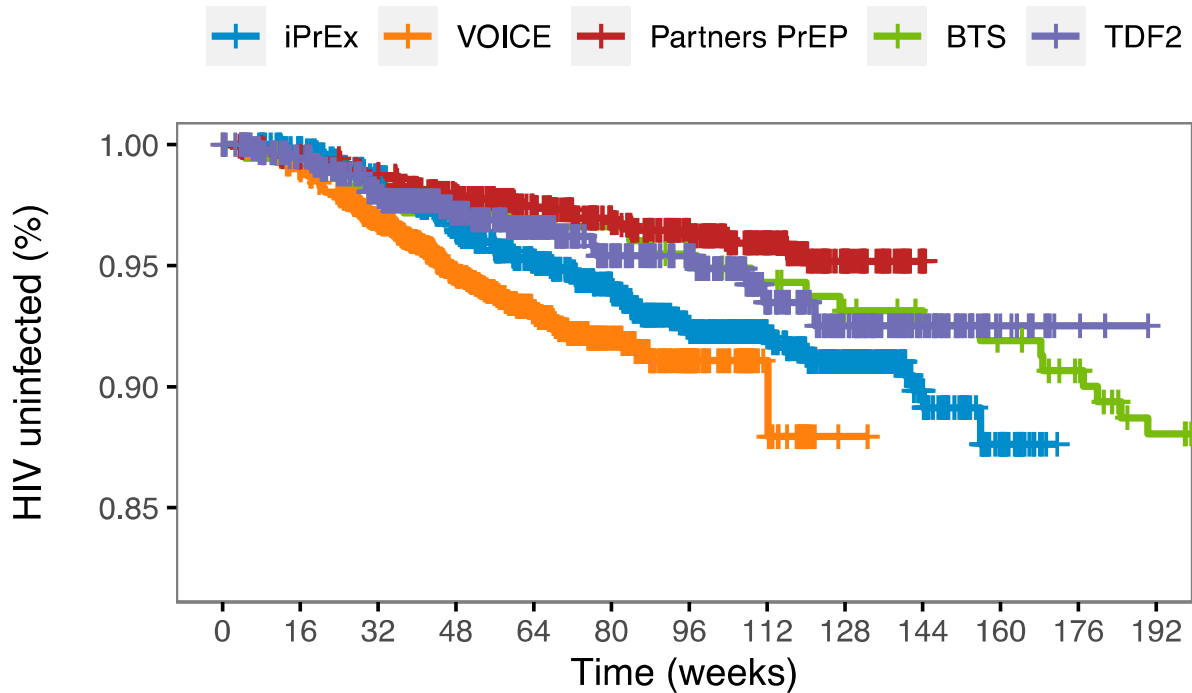
3. Results

3.1. Largest Dataset Regarding HIV Outcome including Different High-Risk Populations

From the placebo arms of the five PrEP phase III clinical trials, 5583 individuals were included in the analysis: 1218 MSM/TGW (iPrEx), 2008 young women at high risk (VOICE), 1574 HIV negative partners in HIV-serodiscordant couples (Partners PrEP), 177 people who inject drugs (BTS), and 606 heterosexual men and women at high risk (TDF2). A detailed description of the baseline characteristics and population-specific risk behaviors collected in the five trials is shown in Supplementary Table S1a–e. Due to the different risk factors and demographic information collected in each study (Table 1 and Table S1a–e), and the large number of individuals enrolled, each study population was analyzed separately. The Supplementary Materials (Preliminary analysis) show the results from the preliminary Kaplan–Meier plot (Supplementary Figures S1–S5) and univariate cox-regression analysis (Supplementary Tables S2–S6) for each study population.

3.2. Substantial Differences in Risk Infection between Different Key Populations

The risk of acquiring HIV over time varied among the studies (Figure 1). Women at high risk (VOICE) showed the highest risk of HIV infection (HIV uninfected < 90% at week 130); and serodiscordant heterosexual partners (Partners PrEP) showed the lowest risk of HIV infection (HIV uninfected ≈ 95% at week 130). The probability of remaining

HIV uninfected over time was characterized by a parametric survival model using an exponential distribution in which the hz over time is equal to a baseline hazard parameter (λ) (hz(t) = λ). Differences in study populations were considered by estimating a different λ value for each population at high risk (Table 2). The hazard function was modified by including significant predictors of HIV risk as covariates of λ using a linear relationship. Predictor estimates and their effect on HIV risk are summarized in Table 2.

Figure 1. Probability of HIV infection over time of the placebo arms of the five PrEP studies. The Bangkok study's follow up was until week 348. Vertical lines represent the censored individuals. Number at risk represents the number of individuals at HIV risk in each study at each time of the *x*-axis. Cumulative number of events represents the number of individuals who contracted HIV infection in each study at each time of the *x*-axis.

Table 2. Final multivariate model by study.

Study	Model Building	Significance	↑ Risk	Parameters	Estimates (RSE%)
iPrEx	Baseline (Exponential)	-	-	θ_λ	0.00103 (14%)
	+Predictor *Condom Receptive Partner*	$p < 0.001$	No use of condom when receptive partner	$\theta_{RECEPTIVE_i} = 1 + \theta_{Receptive\,(\,)}$	$\theta_{Receptive\,(No\,condom)} = 0$ $\theta_{Receptive\,(Condom)} = -0.73\ (12\%)$
	+Predictor *Age*	$p < 0.05$	Younger age	$\theta_{AGE_i} = 1 + \theta_{Age} \times (age - 24)$	$\theta_{Age} = -0.0305\ (34\%)$
	+Predictor *Syphilis Seroreactivity*	Kaplan–Meier analysis	Syphilis positive	$\theta_{SYPHILIS_i} = 1 + \theta_{Syphilis\,(\,)}$	$\theta_{Syphilis(-)} = 0$ $\theta_{Syphilis(+)} = 0.64\ (70\%)$
VOICE	Baseline (Exponential)	-	-	θ_λ	0.001 (15%)
	+Predictor *Herpes2*	$p < 0.05$	Herpes simplex virus 2 Positive	$\theta_{HSV2_i} = 1 + \theta_{HSV2\,(\,)}$	$\theta_{HSV2(+)} = 0.997\ (37\%)$ $\theta_{HSV2(-)} = 0$
	+Predictor *Marital Status*	$p < 0.001$	Not being married	$\theta_{MARITAL_i} = 1 + \theta_{Marital\,(\,)}$	$\theta_{Marital(Married)} = -0.813\ (12\%)$ $\theta_{Marital(No\,Married)} = 0$
	+Predictor *Age*	$p < 0.001$	Younger age	$\theta_{AGE_i} = 1 + \theta_{Age} \times (age - 24)$	$\theta_{Age} = -0.068\ (26\%)$
	+Predictor *Primary Sex Partner has Sex with another Partner (last 3 months)*	$p < 0.05$	Primary sex partner has sex with another partner	$\theta_{PARTNER_i} = 1 + \theta_{Partner\,(\,)}$	$\theta_{Partner(No\,extra\,sex)} = -0.443\ (32\%)$ $\theta_{Partner(Extra\,sex)} = 0$
	+Predictor *Financial Support*	$p < 0.05$	No financial support	$\theta_{FINANCIAL\,SUPPORT_i} = 1 + \theta_{Support(\,)}$	$\theta_{Support(No)} = 0.512\ (60\%)$ $\theta_{Support(Yes)} = 0$
	+Predictor *Number of Children Birth*	Kaplan Meier analysis	Lower number of children birth	$\theta_{CHILDREN_i} = 1 + \theta_{Children} \times (children - 1)$	$\theta_{Children} = -0.162\ (84\%)$
Partners PrEP	Baseline (Exponential)	-	-	θ_λ	0.000166 (43%)
	+Predictor *Age*	$p < 0.01$	Younger age	IF age ≤34 $\theta_{AGE_i} = 1 + \theta_{Age1} \times (age - 34)$ IF age >34 $\theta_{AGE_i} = 1 + \theta_{Age2} \times (age - 34)$	$\theta_{Age1} = -0.253\ (65\%)$ $\theta_{Age2} = 0.0579\ (145\%)$
	+Predictor *Sex at birth*	Kaplan–Meier analysis	Female	$\theta_{SEX_i} = 1 + \theta_{Sex(\,)}$	$\theta_{Sex(Male)} = 0$ $\theta_{Sex(Female)} = 0.742\ (62\%)$
	+Predictor *Number of Children*	Kaplan–Meier analysis	Lower number of children	$\theta_{CHILDREN_i} = 1 + \theta_{Children} \times (children - 3)$	$\theta_{Children} = -0.0828\ (82\%)$
	+Predictor *Partner's Viral Load* *	$p < 0.001$	Partner's viral load detectable	-	-
BTS	Baseline (Exponential)	-	-	θ_λ	0.0005 (23%)
	+Predictor *Sex at birth*	$p < 0.01$	Female	$\theta_{SEX_i} = 1 + \theta_{Sex(\,)}$	$\theta_{Sex(Male)} = 0$ $\theta_{Sex(Female)} = 2.33\ (52\%)$
	+Predictor *Drug Injection Frequency*	-	Daily drug injection	$\theta_{DRUG\,FREQUENCY_i} = 1 + \theta_{Drug\,Frequency\,(\,)}$	$\theta_{Drug\,Frequency\,(\,Weekly)} = 0$ $\theta_{Drug\,Frequency\,(\,Daily)} = 2.22\ (67\%)$
TDF2	Baseline (Exponential)	-	-	θ_λ	0.0006 (20%)
	No available predictor was found significant.				

* Partner's viral load covariate was not included in the final model due to missing information in more than 90% of the individuals of partner's study. $\theta_{PREDICTOR}$ represents the predictor's model relationship, included in the model as covariate of the baseline hazard parameter (λ). RSE: relative standard error, e.g., for iPrEx study: $\lambda = \theta_\lambda \times \theta_{RECEPTIVE_i} \times \theta_{AGE_i} \times \theta_{SYPHILIS_i}$. $\theta_{Predictor()}$ represents the parameter value for the categorical predictors, where ( ) can have one category or other. E.g., for sex at birth: $\theta_{Sex(\,)}$ will take the value of $\theta_{Sex(Male)}$ or $\theta_{Sex(Female)}$ depending on the participant's sex at birth. ↑ Risk: Increase risk.

3.3. Low, Moderate, and High-Risk Profiles Inside Key Populations Based on the Significant Predictors

An RS of 1.93×10^{-4} was associated with a probability of HIV infection of 1% (low risk) and an RS of 9.86×10^{-4} was associated with a probability of 5% (high risk) at year 1 of follow-up. The distribution of individual RSs based on the predictors for each population is shown in Figure 2A.
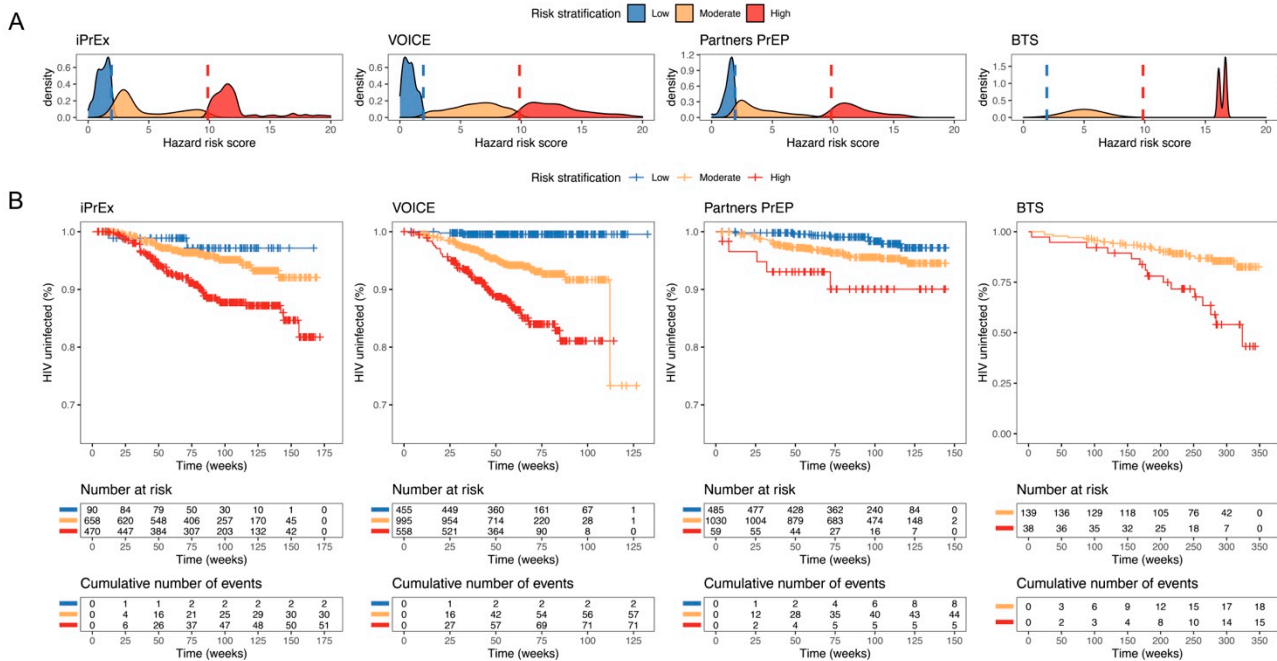

Figure 2. Risk stratification representation. (**A**) Distribution of individual risk scores and (**B**) Observed Kaplan–Meier rates stratified by low, medium, and high-risk group for each study population. Low, medium, and high-risk groups were defined as individuals with a probability of infection over 1 year of <1%, between 1 and 5%, and >5%. No risk stratification is shown for TDF2 study as no available predictor was included in the model.

A subgroup of individuals at low risk (RS $\leq 1.93 \times 10^{-4}$) were identified constituting 8% (MSM/TGW), 23% (young women at high risk) and 31% (HIV negative partners in HIV-serodiscordant heterosexual couples) of their study population. Using this stratification, the population of people who inject drugs in the BTS study did not include low-risk individuals. High-risk individuals (RS $\geq 9.86 \times 10^{-4}$) constituted 36% (MSM/TGW), 28% (young women at high risk), 4% (HIV negative partners in HIV-serodiscordant heterosexual couples) and 21% (people who inject drugs) of their study populations. These high-risk subpopulations show differences across the study populations, with a cumulative probability of acquiring HIV infection in 1 year of 6% (MSM/TGW), 11% (young women at high risk), 7% (HIV negative partners in HIV-serodiscordant heterosexual couples) and 8% (BTS). Moderate risk individuals (1.93×10^{-4} < RS > 9.86×10^{-4}) represented the largest subgroup with individuals constituting 56% (MSM/TGW), 49% (young women at high risk), 65% (HIV negative partners in HIV-serodiscordant heterosexual couples), and 79% (people who inject drugs) of their study populations. We did not find statistically significant predictors of HIV infection in the TDF2 study, so the risk stratification was not applied to that data. The risk stratification based on the individual predictors shows a statistically significant difference in the risk of acquiring HIV infection within and across populations (Figure 2B).

Table 3 shows the distribution of participant characteristics in each risk stratum for each study population. Overall, individuals with high-risk characteristics are most prevalent in the high-risk stratum, in contrast to those with low-risk characteristics (Supple-

mentary Table S7). Interestingly, all the individuals in the high-risk stratum in the Partners PrEP study are women.

Table 3. Distribution of identified risk factors among low, medium, and high-risk groups.

Study	Participant Risk Factors	Number of Participants with Those Risk Factor	Proportion of Participants with Those Risk Factors in Each Risk Strata		
			Low	Moderate	High
iPrEx	No use of condom when receptive partner	734	7/734 (1%)	257/734 (35%)	470/734 (64%)
	Age < 24	570	0/570 (0%)	238/570 (42%)	332/570 (58%)
	Syphilis positive	178	9/178 (5%)	59/178 (33%)	110/178 (62%)
VOICE	Herpes2 positive	918	197/918 (21%)	316/918 (34%)	405/918 (45%)
	Not being married	1582	67/1582 (4%)	957/1582 (61%)	558/1582 (35%)
	Age < 24	894	50/894 (6%)	427/894 (48%)	417/894 (47%)
	Primary sex partner has sex with another partners	231	55/231 (24%)	58/231 (25%)	118/231 (51%)
	No financial support	348	22/348 (6%)	112/348 (32%)	214/348 (62%)
	No children	304	8/304 (3%)	136/304 (45%)	160/304 (53%)
Partners PrEP	Female	621	61/621 (10%)	501/621 (80%)	59/621 (10%)
	Age < 34	755	51/755 (7%)	645/755 (85%)	59/755 (8%)
	Less than 3 children	619	84/619 (14%)	478/619 (77%)	57/619 (9%)
Bangkok	Female	21	0/21 (0%)	0/21 (0%)	21/21 (100%)
	Daily drug injection	17	0/21 (0%)	0/17 (0%)	21/17 (100%)

A comparison of each predictor effect on the HR for each population is shown in Figure 3A. A detailed description of model predictors in each study is included in Supplementary Material (Additional Results). The evaluation of the established models supports the adequacy of the parametric multivariate models and demonstrates good agreement between observed and simulated HIV risk data for each of the identified predictors (Figure 3B). The model evaluation plots support the following statements: (i) there are different levels of HIV risk in each study population (e.g., married women show a lower probability of HIV infection compared to unmarried women in the VOICE trial), and (ii) there is a difference in the high-risk profiles between the study populations (e.g., red profiles of Figure 3B for number of children predictors for VOICE vs. Partners PrEP studies).

3.4. Sex at Birth Effect: Females Appear to Be at Higher Risk of HIV Infection

In the trials that enrolled both men and women, women were at a higher risk of HIV infection. Being female in the Partners PrEP Study increased the relative hazard ratio by 1.74 points, and being female in the BTS increased the relative hazard ratio by 3.33 points. In TDF2, although no predictors were included in the final model, the Kaplan–Meier plot of the raw data shows women at higher risk of HIV infection (Supplementary Figure S5).

3.5. Tool to Estimate Risk of HIV Infection

We developed an online interactive web application to simulate how the risk profiles change for each study population depending on individual predictors. The web app is hosted at http://saviclab.org/hiv-risk/. The tool allows users to compare the probability of HIV infection over time for individuals within and between the key and to determinate the level of risk (low, medium, and high) for each simulated profile. The definition of low, medium, and high risk can be set in the tool by the user by selecting the risk limits (probability of HIV infection) for low and high risk at a given time of follow up.

Supplementary Figure S6 provides a snapshot of the tool comparing the probability of HIV between populations based on the individual predictors.
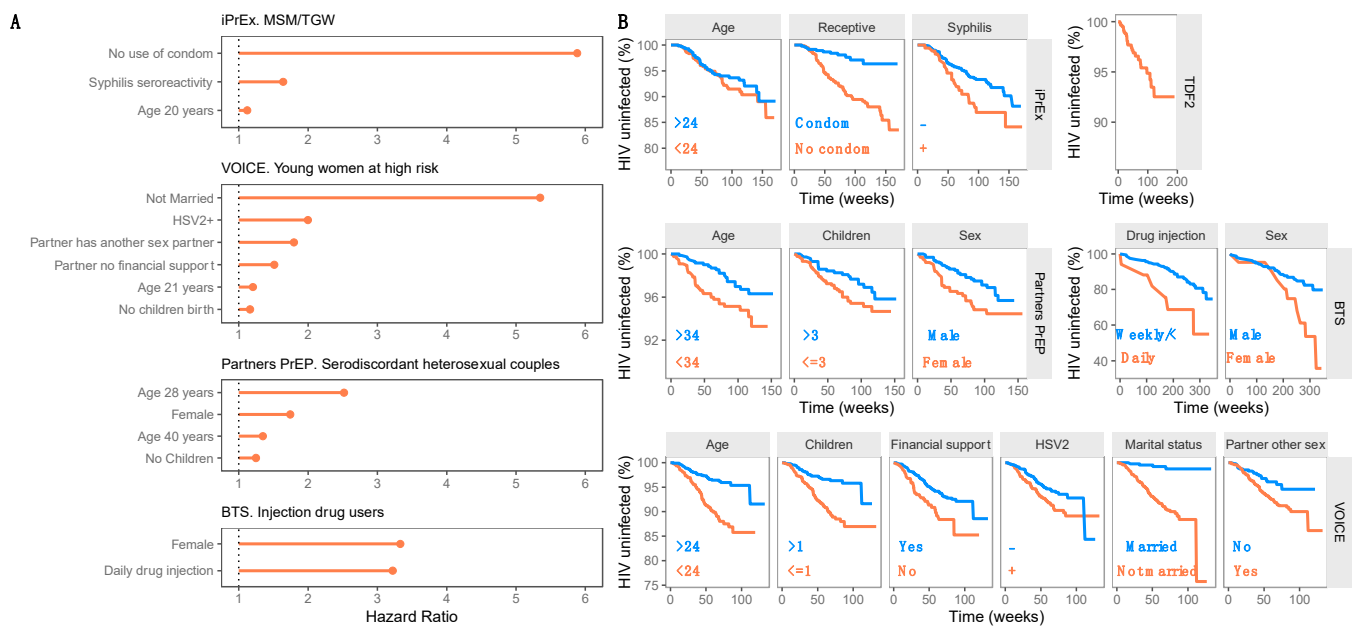

Figure 3. Model predictors and evaluation. (**A**) Effect magnitude of each predictor on the relative risk (hazard ratio) of HIV infection for each population. (**B**) VPCs by studies and by covariates as a Kaplan–Meier plot of HIV probability. Lines represent raw data, and shaded areas cover the 95% prediction interval calculated from 200 simulated studies. No predictor effect is represented for TDF2 study, as no available predictor was included in the model.

4. Discussion

We characterized the probability of HIV infection of five study populations by integrating longitudinal data from the placebo arms of five pivotal PrEP RCTs. The parametric models identified statistically significant predictors of HIV infection in four of the five studies, allowing us to quantify the risk of HIV infection in high-risk individuals not receiving PrEP. We found a substantial difference in the risk of HIV infection among and within the trial populations, with each study population including a mix of low, moderate, and high-risk individuals. Females at birth are associated with a higher risk of infection in populations that included both sexes. Notably, this analysis constitutes the first comparison of substantially different HIV risk levels across key populations.

We developed a parametric multivariate model for each study population due to the heterogeneity of the risk factors across the studies. However, we initially attempted to develop a single parametric multivariate model to quantify the probability of HIV infection in the pooled placebo data from the five studies. This preliminary meta-analysis that included the study populations and their common available covariates (i.e., age, education, ethnicity, sex and high-risk population) identified age, sex and study population as significant predictors of HIV infection ($p < 0.001$). This approach, however, ignored other characteristics defined as high risk (e.g., prisoners) [22]. This analysis also failed to identify the population-specific predictors (e.g., receptive partner condom use among MSM/TGW, or drug injection frequency in people who inject drugs).

Female participants had a higher risk of infection in all studies that included women. While the underlying mechanism for this finding cannot be revealed from this modeling analysis, this result may be influenced by gender inequalities in the populations of women included in this analysis. These gender inequalities could be due to unequal distribution of HIV risk factors (e.g., difference in financial security, sexual behaviors, etc.), as well as the different effect of these risk factors (e.g., male-to-female transmission, concomitant sexually transmitted diseases, etc.) [23]. This finding is consistent with evidence that women account

for more than half of the 36.9 million people living with HIV worldwide (African women comprise the largest component), and that three of five new HIV infections among young people (15–24 years) occur in females [24,25]. Unmarried women in the studies were at a higher risk of HIV infection than married women in the VOICE and Partners PrEP studies. Older age and having had more children lowered the risk of HIV infection by 7% per year increase in age and by 16% per child born to women in the VOICE study and by 25% per year increase in age to 34 years old and by 8% per child born to women in the Partners PrEP study. There is an urgent need to focus HIV prevention on young women at high sexual risk [26], especially in high HIV burden settings.

Our findings are consistent with previous studies that used risk assessment tools to predict HIV infection risk [14,16,17,27–31]. Supplementary Materials (Additional Discussion) includes a comparison of our findings with these studies. These risk assessment tools are largely based on descriptive and semi-parametric Cox-proportional hazard analyses, which are commonly used to quantify risk and identify predictors but cannot be used to simulate HIV risk over time based on individual's characteristics. Our analysis was performed using parametric survival models in which the survival and hazard function are characterized by a distribution expressed in terms of parameters that can be estimated. The main advantages of the parametric survival models we used in this analysis are that, as the underlying hazard over time is fully characterized and estimated, simulations of the cumulative probability of HIV infection over time can be performed in different scenarios, for different individuals, based on the combination of the predictors included in the model (e.g., HIV risk for a 20-year-old man with a syphilis diagnosis, who has condomless receptive anal sex with men while not on PrEP). In addition, we integrated the models in a user-friendly interactive platform that overcomes the associated complexity of model computation. This tool integrates the models for predicting the probability of remaining HIV uninfected over time based on individual characteristics for the four study populations for which significant predictors were identified. It can help identify the combination of variables associated with the highest or lowest risk of infection compared to the median risk profile for a determinant population. Additionally, the tool can be updated by integrating information from ongoing and future PrEP RCTs to continue to refine and improve the estimates [32]. The integration of new data could help to refine and validate the current models, acquiring more precision in the parameters estimates, and it could help to identify new predictors of HIV risk in those population where this analysis was unable to find them (e.g., heterosexual men at high risk of infection, from TDF2 study). The tool serves as a companion interactive app for the manuscript, and it could be used to stimulate conversation about HIV prevention and as an informational tool for assisting on efficient future triapl anning.

Our analysis had several limitations. The studies were conducted among specific populations and geographic locations, and the results may not be generalizable to other populations and sites. Unfortunately, this analysis was unable to find significant predictors of HIV risk in the TDF2 study, which includes data from a population of heterosexual men and women at high risk of infection. In addition, some participant characteristics (e.g., partner viral load (Partners PrEP) or drug use (iPrEx)) that have been found to be associated with HIV infection [14] were not included in our models because of missing data. For example, among the 1574 Partner PrEP participants, only 85 had partner viral load information available. All 49 participants whose partners had quantifiable viral load acquired HIV during follow-up, and none of the 36 participants whose partners had undetectable viral loads became infected. This is consistent with studies indicating essentially no HIV transmission risk when viral load is below detection limits [33]. In addition, the models were study-specific, so it was not possible to include the effect on HIV risk of the combination of characteristics across all of the different study populations. Lastly, the data used are slightly dated, with participants enrolled over ten years ago. Therefore, future plans include the collection of contemporary data and key populations to validate and refine this analysis.

5. Conclusions

In conclusion, the parametric models we used allowed us to identify populations at low, medium, and high risk of HIV infection. Our findings can be used to improve HIV prevention trial recruitment by focusing on individuals at highest risk of HIV infection with greater anticipated incidence. Furthermore, these models can be used to inform risk without bias in a population receiving PrEP as prevention therapy by differentiating between the protective effect of PrEP and the underlying HIV risk based on the identified predictors [34]. Future plans include adding new placebo arms from new clinical trials to the existing framework to promote knowledge integration and data sharing and to increase the predictive power of the analysis.

References

1. WHO. HIV/AIDS Fact Sheet 2020. Available online: https://www.who.int/en/news-room/fact-sheets/detail/hiv-aids (accessed on 28 October 2021).
2. Grant, R.M.; Lama, J.R.; Anderson, P.L.; McMahan, V.; Liu, A.Y.; Vargas, L.; Goicochea, P.; Casapía, M.; Guanira-Carranza, J.V.; Ramirez-Cardich, M.E.; et al. Preexposure chemoprophylaxis for HIV prevention in men who have sex with men. *N. Engl. J. Med.* **2010**, *363*, 2587–2599. [PubMed]
3. Marrazzo, J.M.; Ramjee, G.; Richardson, B.A.; Gomez, K.; Mgodi, N.; Nair, G.; Palanee, T.; Nakabiito, C.; van der Straten, A.; Noguchi, L.; et al. Tenofovir-Based Preexposure Prophylaxis for HIV Infection among African Women. *N. Engl. J. Med.* **2015**, *372*, 509–518. [PubMed]
4. Baeten, J.M.; Donnell, D.; Ndase, P.; Mugo, N.R.; Campbell, J.D.; Wangisi, J.; Tappero, J.W.; Bukusi, E.A.; Cohen, C.R.; Katabira, E.; et al. Antiretroviral Prophylaxis for HIV Prevention in Heterosexual Men and Women. *N. Engl. J. Med.* **2012**, *367*, 399–410. [PubMed]
5. Choopanya, K.; Martin, M.; Suntharasamai, P.; Sangkum, U.; Mock, P.A.; Leethochawalit, M.; Chiamwongpaet, S.; Kitisin, P.; Natrujirote, P.; Kittimunkong, S.; et al. Antiretroviral prophylaxis for HIV infection in injecting drug users in Bangkok, Thailand (the Bangkok Tenofovir Study): A randomised, double-blind, placebo-controlled phase 3 trial. *Lancet* **2013**, *381*, 2083–2090.
6. Thigpen, M.C.; Kebaabetswe, P.M.; Paxton, L.A.; Smith, D.K.; Rose, C.E.; Segolodi, T.M.; Henderson, F.L.; Pathak, S.R.; Soud, F.A.; Chillag, K.L.; et al. Antiretroviral Preexposure Prophylaxis for Heterosexual HIV Transmission in Botswana. *N. Engl. J. Med.* **2012**, *367*, 423–434.
7. Van Damme, L.; Corneli, A.; Ahmed, K.; Agot, K.; Lombaard, J.; Kapiga, S.; Malahleha, M.; Owino, F.; Manongi, R.; Onyango, J.; et al. Preexposure prophylaxis for HIV infection among African women. *N. Engl. J. Med.* **2012**, *367*, 411–422.
8. Molina, J.M.; Capitant, C.; Spire, B.; Pialoux, G.; Cotte, L.; Charreau, I.; Tremblay, C.; Le Gall, J.-M.; Cua, E.; Pasquet, A.; et al. On-demand preexposure prophylaxis in men at high risk for HIV-1 infection. *N. Engl. J. Med.* **2015**, *373*, 2237–2246.
9. Hosek, S.G.; Siberry, G.; Bell, M.; Lally, M.; Kapogiannis, B.; Green, K.; Fernandez, M.I.; Rutledge, B.; Martinez, J.; Garofalo, R.; et al. The acceptability and feasibility of an HIV preexposure prophylaxis (PrEP) trial with young men who have sex with men. *J. Acquir. Immune Defic. Syndr.* **2013**, *62*, 447–456.
10. Karim, Q.A.; Karim, S.S.A.; Frohlich, J.A.; Grobler, A.C.; Baxter, C.; Mansoor, L.E.; Kharsany, A.B.M.; Sibeko, S.; Mlisana, K.P.; Omar, Z.; et al. Effectiveness and safety of tenofovir gel, an antiretroviral microbicide, for the prevention of HIV infection in women. *Science* **2010**, *329*, 1168–1174.
11. WHO. Global State of PrEP. Available online: https://www.who.int/groups/global-prep-network/global-state-of-prep (accessed on 28 October 2021).
12. Sidebottom, D.; Ekström, A.M.; Strömdahl, S. A systematic review of adherence to oral pre-exposure prophylaxis for HIV—How can we improve uptake and adherence? *BMC Infect. Dis.* **2018**, *18*, 581.
13. WHO. Consolidated Guidelines on the Use of Antiretroviral Drugs for Treating and Preventing HIV Infection. 2016. Available online: https://www.who.int/publications/i/item/9789241549684 (accessed on 24 July 2019).
14. Kahle, E.M.; Hughes, J.P.; Lingappa, J.R.; John, G. An empiric risk scoring tool for identifying high-risk heterosexual HIV-1 serodiscordant couples for targeted HIV-1 prevention. *J. Acquir. Immune Defic. Syndr.* **2014**, *62*, 339–347. [CrossRef] [PubMed]
15. Balkus, J.E.; Brown, E.; Palanee, T.; Nair, G.; Gafoor, Z.; Zhang, J.; Richardson, B.A.; Chirenje, Z.M.; Marrazzo, J.; Baeten, J.M. An Empiric HIV Risk Scoring Tool to Predict HIV-1 Acquisition in African Women. *J. Acquir. Immune Defic. Syndr.* **2016**, *72*, 333–343. [CrossRef] [PubMed]
16. Smith, D.K.; Pals, S.L.; Herbst, J.H.; Shinde, S.; Carey, J.W. Development of a Clinical Screening Index Predictive of Incident HIV Infection Among Men Who Have Sex with Men in the United States. *J. Acquir. Immune Defic. Syndr.* **2012**, *60*, 421–427. [CrossRef] [PubMed]
17. Menza, T.W.; Hughes, J.P.; Celum, C.L.; Golden, M.R. Prediction of HIV acquisition among men who have sex with men. *Sex Transm. Dis.* **2009**, *36*, 547–555. [CrossRef] [PubMed]
18. Collett, D. *Modelling Survival Data in Medical Research*; Chapman and Hall/CRC: Boca Raton, FL, USA, 1952.
19. Bauer, R. *NONMEM Users Guide: Introduction to NONMEM 7.4.1.*; ICON Development Solutions: Ellicott City, MD, USA, 2017; Volume 53, p. 128.
20. Nyberg, J.; Karlsson, K.E.; Jönsson, S.; Simonsson, U.S.H.; Karlsson, M.O.; Hooker, A.C. Simulating Large Time-to-Event Trials in NONMEM. PAGE 23 (2014) Abstr. 3166. Available online: www.page-meeting.org/?abstract=3166 (accessed on 1 July 2019).
21. PsN. *VPC and NPC User Guide: PsN 2019, 5.3.0*. pp. 1–46. Available online: https://uupharmacometrics.github.io/PsN/docs.html (accessed on 1 July 2019).
22. WHO. Global Health Sector Strategy on HIV 2016–2021. 2016. Available online: http://apps.who.int/iris/bitstream/10665/2461 78/1/WHO-HIV-2016.05-eng.pdf?ua=1 (accessed on 24 July 2019).
23. Sia, D.; Onadja, Y.; Nandi, A.; Foro, A.; Brewer, T. What lies behind gender inequalities in HIV/AIDS in sub-Saharan African countries: Evidence from Kenya, Lesotho and Tanzania. *Health Policy Plan.* **2014**, *29*, 938–949. [CrossRef]
24. UNAIDS. *Women and HIV—A Spotlight on Adolescent Girls and Young Women*; UNAIDS: Geneva, Switzerland, 2019; pp. 1–20. Available online: http://www.unaids.org/sites/default/files/media_asset/2019_women-and-hiv_en.pdf (accessed on 28 October 2021).

25. UNAIDS. UNAIDS Data 2019. Available online: https://www.unaids.org/sites/default/files/media_asset/2019-UNAIDS-data_en.pdf (accessed on 28 October 2021).

26. The Lancet HIV. For the HIV epidemic to end so must gender inequality. *Lancet HIV* **2019**, *6*, e411. [CrossRef]

27. Krakower, D.S.; Gruber, S.; Hsu, K.; Menchaca, J.T.; Maro, J.C.; Kruskal, B.A.; Wilson, I.B.; Mayer, K.H.; Klompas, M. Development and validation of an automated HIV prediction algorithm to identify candidates for pre-exposure prophylaxis: A modelling study. *Lancet HIV* **2019**, *3018*, e696–e704. [CrossRef]

28. Marcus, J.L.; Hurley, L.B.; Krakower, D.S.; Alexeeff, S.; Silverberg, M.J.; Volk, J.E. Use of electronic health record data and machine learning to identify candidates for HIV pre-exposure prophylaxis: A modelling study. *Lancet HIV* **2019**, *6*, e688–e695. [CrossRef]

29. Padian, N.S.; Buvé, A.; Balkus, J.; Serwadda, D.; Cates, W., Jr. Series HIV Prevention 2 Biomedical interventions to prevent HIV infection: Evidence, challenges, and way forward. *Lancet* **2008**, *372*, 585–599. [CrossRef]

30. Wand, H.; Reddy, T.; Naidoo, S.; Moonsamy, S.; Siva, S.; Morar, N.S.; Ramjee, G. A Simple Risk Prediction Algorithm for HIV Transmission: Results from HIV Prevention Trials in KwaZulu Natal, South Africa (2002–2012). *AIDS Behav.* **2018**, *22*, 325–336. [CrossRef]

31. Buchbinder, S.P.; Glidden, D.V.; Liu, A.Y.; McMahan, V.; Guanira, J.V.; Mayer, K.H.; Goicochea, P.; Grant, R.M. HIV pre-exposure prophylaxis in men who have sex with men and transgender women: A secondary analysis of a phase 3 randomised controlled efficacy trial. *Lancet Infect. Dis.* **2014**, *14*, 468–475. [CrossRef]

32. Kiley, R.; Peatfield, T.; Hansen, J.; Reddington, F. Data Sharing from Clinical Trials—A Research Funder's Perspective. *N. Engl. J. Med.* **2017**, *377*, 1990–1992. [CrossRef] [PubMed]

33. Eisinger, R.W.; Dieffenbach, C.W.; Fauci, A.S. HIV Viral Load and Transmissibility of HIV Infection: Undetectable Equals Untransmittable. *JAMA* **2019**, *321*, 451–452. [CrossRef] [PubMed]

34. Garcia-Cremades, M.; Vučićević, K.; Hendrix, C.W.; Jayachandran, P.; Jarlsberg, L.; Grant, R.; Celum, C.L.; Martin, M.; Baeten, J.M.; Marrazzo, J.; et al. Characterizing HIV-preventive, plasma tenofovir concentrations. A pooled participant-level data analysis from HIV pre-exposure prophylaxis (PrEP) clinical trials. *Clin. Infect. Dis.* **2022**, *26*, ciac313. [CrossRef] [PubMed]

35. Lindbom, L.; Ribbing, J.; Jonsson, E.N. Perl-speaks-NONMEM (PsN)—A Perl module for NONMEM related programming. *Comput. Methods Programs Biomed.* **2004**, *75*, 85–94. [CrossRef] [PubMed]

36. Abdool Karim, S.S.; Richardson, B.A.; Ramjee, G.; Hoffman, I.F.; Chirenje, Z.M.; Taha, T.; Kapina, M.; Maslankowski, L.; Coletti, A.; Profy, A.; et al. Safety and effectiveness of BufferGel and 0.5% PRO2000 gel for the prevention of HIV infection in women. *AIDS* **2011**, *25*, 957–966. [CrossRef]

37. Heffron, R.; Donnell, D.; Rees, H.; Celum, C.; Mugo, N.; Were, E.; de Bruyn, G.; Nakku-Joloba, E.; Ngure, K.; Kiarie, J.; et al. Use of hormonal contraceptives and risk of HIV-1 transmission: A prospective cohort study. *Lancet Infect. Dis.* **2012**, *12*, 19–26. [CrossRef]

38. Lingappa, J.R.; Kahle, E.; Mugo, N.; Mujugira, A.; Magaret, A.; Baeten, J.; Bukusi, E.A.; Cohen, C.R.; Katabira, E.; Ronald, A.; et al. Characteristics of HIV-1 discordant couples enrolled in a trial of HSV-2 suppression to reduce HIV-1 transmission: The partners study. *PLoS ONE* **2009**, *4*, e5272. [CrossRef]