

## Supplementary Digital Content

### **eAppendix 1. Technical details of g-computation approach**

Here, we outline our g-computation approach. We will not go into detail about estimation of the ITT effect in the high incidence African setting because it was a simple implementation of Snowden's adaptation of the g-formula, where our exposure was randomization ( $R$ ) and outcome was HIV seroconversion ( $Y$ ).<sup>1</sup> We did not need to control for any covariates because the randomization ensured exchangeability in expectation.

#### G-computation for the per-protocol parameter

We were first interested in estimating the per-protocol effect in our high incidence African setting, using the risk ratio (RR) as our contrast measure. The per-protocol parameter compares the risk of obtaining the outcome under 2 counterfactual scenarios: one where the entire population adheres to study protocol for the index treatment and the other where the entire population adheres to protocol for the comparator treatment.<sup>2</sup> In our example, we compared the scenario where all women were randomized to a treatment (PrEP) and received that treatment to the one where all women were randomized to the comparator (placebo) and received it. In notation we sought to estimate:

$$RR = P(Y^{r=a=1} = 1) / P(Y^{r=a=0} = 1)$$

where  $Y$  is the outcome,  $R = r$  is the treatment that was randomized, and  $A = a$  is the treatment received. The superscript is intended to highlight that this is a counterfactual outcome where the exposure was set to be  $r = a$ , rather the observed risk of outcome

conditional on  $r = a$ . The two counterfactual outcome probabilities can be obtained via the g-formula:

$$P(Y^{r=a} = 1) = \sum_z E[Y|R = r, A = a, r = a, Z = z]P(Z = z)$$

where  $Z$  is the set of variables necessary to achieve conditional exchangeability between those who took PrEP as designed and those who took placebo as designed.

Using g-computation, we estimated these right hand probabilities through a sequence of parametric models for all variables necessary for the analysis (e.g., exposure, confounders, or outcome). Baseline variables did not need to be modeled. For our simulation (see Figure 1), those baseline variables were the effect measure modifier lack of an STI ( $V$ ), the confounder age ( $M$ ), and randomization. We then ran two logistic regression models. First, we modeled adherence within the stratum where participants were randomized to PrEP because all individuals who were randomized to placebo actually received it. The model took the form:

$$\text{logit}[P(A = 1|R = 1)] = \log \left[ \frac{P(A = 1|R = 1)}{1 - P(A = 1|R = 1)} \right] = \beta_{a0} + \beta_{a1}M$$

Second, we regressed the outcome on all variables which causally affected it:

$$\text{logit}[P(Y = 1)] = \beta_{y0} + \beta_{y1}A + \beta_{y2}M + \beta_{y3}V$$

The next step involved using the estimated coefficients from the above two models to predict for each individual  $i$  their probability of receiving PrEP and contracting HIV under the natural course and under the two counterfactual scenarios of interest. The natural course for receipt of PrEP and HIV seroconversion were obtained as follows:

$$A_i = P(A = 1) = \begin{cases} 0, & R = 0 \\ \text{expit}(\beta_{a0} + \beta_{a1}M), & R = 1 \end{cases}$$

$$Y_i = P(Y = 1) = \text{expit}(\beta_{y0} + \beta_{y1}A_i + \beta_{y2}M_i + \beta_{y3}V_i)$$

where  $\text{expit}(x) = 1/[1 + \exp(-x)]$ . Comparison of this estimated natural course with the observed data served as a check that our parametric models were not misspecified. For the counterfactual scenarios, we intervened to set both randomization and receipt of treatment;  $P(A = a)$  did not need to be predicted (it was known with probability one). We therefore only needed to predict each individual's probability of contracting HIV under the two scenarios, which we did as follows:

$$Y_i^{11} = P(Y^{r=a=1} = 1) = \text{expit}(\beta_{y0} + \beta_{y1}(1) + \beta_{y2}M_i + \beta_{y3}V_i)$$

$$Y_i^{00} = P(Y^{r=a=0} = 1) = \text{expit}(\beta_{y0} + \beta_{y1}(0) + \beta_{y2}M_i + \beta_{y3}V_i)$$

The probabilities were then contrasted to estimate the per-protocol RR.

#### G-computation for generalization

To estimate the per-protocol effect in the target population, we had to adapt the g-formula above to reflect the fact that we were generalizing from our source population to the target:

$$P(Y^{r=a} = 1) = \sum_V E[Y|R = r, A = a, r = a, S = 1, V = v]P(V = v)$$

where  $V$  is the set of variables necessary to achieve conditional exchangeability between those selected into the source population ( $S = 1$ ) and those not selected ( $S = 0$ ).<sup>3</sup> In our case, those selected were the women in the high incidence African setting, and those not selected were women in the target, low incidence US setting. Our set  $V$  included the baseline effect measure modifier lack of an STI, which also differed between the two settings.

Again, a sequence of parametric models were applied, this time using data from the African setting to estimate  $E[Y|R = r, A = a, r = a, S = 1, V = v]$  and from the US setting to estimate  $P(V = v)$ . We accomplished the latter by taking a sample from the African setting,

with sampling weighted by inverse odds of selection.<sup>4</sup> This ensured that, while the sample size and the distribution of all other variables remained the same as in the African setting, the distribution of the modifier age was on average that seen in the US setting.

The weight for each individual  $i$  in both settings took the form:

$$w_i = \begin{cases} \frac{P(S_i = 0|V_i = v)}{P(S_i = 1|V_i = v)} \times \frac{P(S_i = 1)}{P(S_i = 0)} & S_i = 1 \\ 0 & S_i = 0 \end{cases}$$

The conditional and marginal probabilities of selection were estimated using logistic regression in a combined population of women from both the African and US settings. The weighted sample of African women was then used to conduct the same g-computation steps as previously described for the per-protocol parameter.

This same weighted sample was also used to estimate the generalized ITT parameter. As with the per-protocol effect, we ran a series of logistic regression models and then generated for each individual their two counterfactual outcomes. However, there were several important differences in the implementation. For one, we were interested in a different counterfactual comparison, namely:

$$RR = P(Y^{r=1} = 1)/P(Y^{r=0} = 1)$$

Thus, for the ITT effect, we only set the value for randomization and not for whether the woman adhered to protocol. Instead, we were interested in letting receipt of PrEP be as it was in the US setting. Thus, our model to estimate the probability of a woman actually receiving PrEP took the form:

$$\text{logit}[P(A^* = 1|R^* = 1)] = \beta_{a0}^* + \beta_{a1}^* M^*$$

where  $A^*$ ,  $R^*$ , and  $M^*$  were adherence, randomization, and age among US women. These coefficients were then used to estimate the predicted probabilities of being adherent to protocol and the outcome in the weighted sample of African women, as shown below:

Randomization to PreP

$$A_{1i} = P(A^{r=1} = 1) = \text{expit}(\beta_{a0}^* + \beta_{a1}^* M_i)$$

$$Y_{1i} = P(Y^{r=1} = 1) = \text{expit}(\beta_{y0} + \beta_{y1} A_{1i} + \beta_{y2} M_i + \beta_{y3} V_i)$$

Randomization to placebo

$$A_{0i} = P(A^{r=0} = 1) = 0$$

$$Y_{0i} = P(Y^{r=0} = 1) = \text{expit}(\beta_{y0} + \beta_{y1} A_{0i} + \beta_{y2} M_i + \beta_{y3} V_i)$$

Finally, we compared these two counterfactual outcome probabilities to estimate the ITT RR.

## eAppendix 2. Supplementary Table

eTable 1. Intention-to-treat and per-protocol effects in the higher HIV incidence African setting and generalized to the lower incidence US setting, estimated using log binomial models and, for per-protocol effects, inverse probability weighting methods

Parameter	African Trials (trial n=4000)			US Trials (trial n=500)		
	RR	2.5 <sup>th</sup>	97.5 <sup>th</sup>	RR	2.5 <sup>th</sup>	97.5 <sup>th</sup>
Intention to treat	0.65	0.47	0.88			
Per protocol	0.20	0.08	0.35	0.17	0.06	0.31

Abbreviations: HIV, human immunodeficiency virus; US, United States; n: sample size; RR, average risk ratio; 2.5<sup>th</sup>: 2.5<sup>th</sup> percentile of the 2000 simulation RRs; 97.5<sup>th</sup>: 97.5<sup>th</sup> percentile of the 2000 simulation RRs

### eAppendix 3. Simulation and g-computation SAS code

```

DATA a; call streaminit(123);
DO sim=1 TO 2000;
  DO i=1 TO 4500;
    *Create 4000 women enrolled in African trial;
    IF i<=4000 THEN DO; s=1; *Sample;

      *Baseline modifier: no (non-HSV) STI;
      v=rand("bernoulli",0.79);

      *Randomization to PrEP or placebo;
      r=rand("bernoulli",0.5);

      *Confounder of a and y: age >21;
      m=rand("bernoulli",1/3);

      *Receipt of treatment: affected by m and r;
      IF r=1 THEN a=rand("bernoulli",1/(1+exp(-(-log(1/0.5-1)-log(3)*(1/3)+log(3)*m))));
      ELSE a=0;

      *Adherence to protocol;
      IF r=a THEN c=0; ELSE c=1;

      *Outcome: affected by a, m, and v*a;
      y=rand("bernoulli",1/(1+exp(-(-log(1/0.03-1)+log(2.5)*0.25+log(3)*(1/3)+log(3)*0.1975
        -log(2.5)*a-log(3)*m-log(3)*v*a))));

      *ITT counterfactual outcomes;
      a1=rand("bernoulli",1/(1+exp(-(-log(1/0.5-1)-log(3)*(1/3)+log(3)*m))));
      a0=0;
      y1=rand("bernoulli",1/(1+exp(-(-log(1/0.03-1)+log(2.5)*0.25+log(3)*(1/3)+log(3)*0.1975
        -log(2.5)*a1-log(3)*m-log(3)*v*a1))));
      y0=rand("bernoulli",1/(1+exp(-(-log(1/0.03-1)+log(2.5)*0.25+log(3)*(1/3)+log(3)*0.1975
        -log(2.5)*a0-log(3)*m-log(3)*v*a0))));

      *Per-protocol counterfactual outcomes;
      y11=rand("bernoulli",1/(1+exp(-(-log(1/0.03-)+log(2.5)*0.25+log(3)*(1/3)+log(3)*0.1975
        -log(2.5)*1-log(3)*m-log(3)*v*1)))); *r=a=1;
      y00=rand("bernoulli",1/(1+exp(-(-log(1/0.03-)+log(2.5)*0.25+log(3)*(1/3)+log(3)*0.1975
        -log(2.5)*0-log(3)*m-log(3)*v*0)))); *r=a=0;

    OUTPUT;
  END;

  *Create 500 women enrolled in US trial;
  ELSE DO; s=0; *Target;

    *Baseline modifier: no (non-HSV) STI;
    v=rand("bernoulli",0.89);

```

```

*Randomization to PrEP or placebo;
r=rand("bernoulli",0.5);

*Confounder of a and y: age>21;
m=rand("bernoulli",2/3);

*Receipt of treatment;
IF r=1 THEN a=rand("bernoulli",1/(1+exp(-(-log(1/0.75-1)-log(3)*(2/3)+log(3)*m))));
ELSE a=0;

*Adherence to protocol;
IF r=a THEN c=0; ELSE c=1;

*Outcome (lower risk of HIV compared to above);
y=rand("bernoulli",1/(1+exp(-(-log(1/0.003-1)+log(2.5)*0.375+log(3)*(2/3)
+log(3)*0.33375-log(2.5)*a-log(3)*m-log(3)*v*a))));

*ITT counterfactual outcomes (set r=1 and r=0);
a1=rand("bernoulli",1/(1+exp(-(-log(1/0.75-1)-log(3)*(2/3)+log(3)*m))));
a0=0;
y1=rand("bernoulli",1/(1+exp(-(-log(1/0.003-1)+log(2.5)*0.375+log(3)*(2/3)
+log(3)*0.33375-log(2.5)*a1-log(3)*m-log(3)*v*a1))));
y0=rand("bernoulli",1/(1+exp(-(-log(1/0.003-1)+log(2.5)*0.375
+log(3)*(2/3)+log(3)*0.33375-log(2.5)*a0-log(3)*m-log(3)*v*a0))));

*Per-protocol counterfactual outcomes (set r=a=1 and r=a=0);
y11=rand("bernoulli",1/(1+exp(-(-log(1/0.03-1)+log(2.5)*0.375+log(3)*(2/3)
+log(3)*0.33375-log(2.5)*1-log(3)*m-log(3)*v*1))));
y00=rand("bernoulli",1/(1+exp(-(-log(1/0.03-1)+log(2.5)*0.375+log(3)*(2/3)
+log(3)*0.33375-log(2.5)*0-log(3)*m-log(3)*v*0))));

OUTPUT;
END;
END;
END;
RUN;

```

```
*****
```

Estimate true RRs based on individual counterfactual outcomes;

```
DATA africa; SET a; WHERE s=1; RUN;
PROC SORT DATA=africa; BY sim; RUN;
```

```
PROC MEANS DATA=africa noprint;
VAR y11 y00 y1 y0;
OUTPUT OUT=yavg_afr mean=ay11 ay00 ay1 ay0;
RUN;
```



```
DATA truth_afr_pp; SET yavg_afr;
    truth=ay11/ay00;
    param="Africa Per-Protocol";
    KEEP truth param;
```

```
RUN;
```

```
DATA truth_afr_itt; SET yavg_afr;
    truth=ay1/ay0;
    param="Africa ITT";
    KEEP truth param;
```

```
RUN;
```

```
DATA us; SET a; WHERE s=0; RUN;
PROC SORT DATA=us; BY sim; RUN;
```

```
PROC MEANS DATA=us noprint;
    VAR y11 y00 y1 y0;
    OUTPUT OUT=yavg_us mean=ay11 ay00 ay1 ay0;
```

```
RUN;
```

```
DATA truth_us_pp; SET yavg_us;
    truth=ay11/ay00;
    param="US Per-Protocol";
    KEEP truth param;
```

```
RUN;
```

```
DATA truth_us_itt; SET yavg_us;
    truth=ay1/ay0;
    param="US ITT";
    KEEP truth param;
```

```
RUN;
```

```
*****
```

```
ITT and Per-protocol estimated in Africa;
```

```
*G-computation ITT;
```

```
RUN; ODS SELECT NONE; RUN;
```

```
PROC GENMOD DATA=africa desc;
```

```
    BY sim;
```

```
    MODEL y=r / link=logit dist=bin;
```

```
    ODS OUTPUT parameterestimates=betas(keep=sim parameter estimate);
```

```
RUN;
```

```
RUN; ODS SELECT ALL; RUN;
```

```
PROC TRANSPOSE DATA=betas OUT=betas2(drop=beta3) PREFIX=beta;
```

```
    BY sim;
```

```
RUN;
```

```

DATA africa_p;
  MERGE africa betas2;
  BY sim;
  LENGTH param $ 19;
  py_1=1/(1+exp(-(beta1 + beta2*1))); *E[Y(1)];
  py_0=1/(1+exp(-(beta1 + beta2*0))); *E[Y(0)];
  rr=py_1/py_0;
  lnrr=log(py_1/py_0);
  param="Africa ITT";
RUN;

PROC MEANS DATA=africa_p noprint;
  VAR rr; BY sim;
  OUTPUT OUT=afr_p2(keep=sim est) mean=est;
RUN;

PROC UNIVARIATE DATA=afr_p2 noprint; V
  AR est;
  OUTPUT OUT=afr_itt mean=rr pctlpre=p pctlpts=2.5,97.5;
RUN;

*G-computation Per-Protocol;

*Baseline variables don't need to be modeled;
DATA b; SET africa; pv=v; pr=r; pm=m; RUN;

*Model receipt of treatment;
RUN; ODS SELECT NONE; RUN;
PROC GENMOD DATA=b desc;
  WHERE r=1;
  BY sim;
  MODEL a=m / link=logit dist=bin;
  ODS OUTPUT parameterestimates=a_betas(keep=sim parameter estimate);
RUN;
RUN; ODS SELECT ALL; RUN;

DATA a_betas; SET a_betas;
  BY sim;
  RETAIN alnt am;
  IF parameter="Intercept" THEN alnt=estimate;
  IF parameter="m" THEN am=estimate;
  IF last.sim THEN OUTPUT;
  KEEP sim alnt am;
RUN;

*Model the outcome;
RUN; ODS SELECT NONE; RUN;

```

```

PROC GENMOD DATA=b desc;
  BY sim;
  MODEL y=a m v / link=logit dist=bin;
  ODS OUTPUT parameterestimates=y_betas(keep=sim parameter estimate);
RUN;
RUN; ODS SELECT ALL; RUN;

DATA y_betas; SET y_betas;
  BY sim;
  RETAIN yInt ya ym yv;
  IF parameter="Intercept" THEN yInt=estimate;
  IF parameter="a" THEN ya=estimate;
  IF parameter="m" THEN ym=estimate;
  IF parameter="v" THEN yv=estimate;
  IF last.sim THEN OUTPUT;
  KEEP sim yInt ya ym yv;
RUN;
DATA b; MERGE b a_betas y_betas; BY sim; RUN;

*Natural history;
DATA b; SET b;
  IF r=1 THEN pa_n=1/(1+exp(-(aInt+am*pm)));
  ELSE pa_n=0;
  py_n=1/(1+exp(-(yInt+ya*pa_n+ym*pm+yv*pv)));
RUN;

*Per-protocol effect;
DATA b2; SET b;
  LENGTH param $ 19;
  *R=A=1;
  py_1=1/(1+exp(-(yInt+ya*1+ym*pm+yv*pv)));
  *R=A=0;
  py_0=1/(1+exp(-(yInt+ya*0+ym*pm+yv*pv)));
  rr=py_1/py_0;
  lnrr=log(py_1/py_0);
  param="Africa Per-Protocol";
RUN;

PROC MEANS DATA=b2 noprint;
  BY sim;
  VAR rr;
  OUTPUT OUT=b3(keep=sim est) mean=est;
RUN;
PROC UNIVARIATE DATA=b3 noprint;
  VAR est; OUTPUT OUT=afr_pp mean=rr pctlpre=p pctlpts=2.5,97.5;

```

RUN;

\*\*\*\*\*

Generalized ITT and Per-protocol to US population;

\*Create selection weights;

PROC SORT DATA=a OUT=all; BY sim; RUN;

PROC LOGISTIC DATA=all desc noprint;

BY sim;

MODEL s=;

OUTPUT OUT=sn p=sn;

TITLE "Sampling weights numerator";

RUN;

PROC LOGISTIC DATA=all desc noprint;

BY sim;

MODEL s=v;

OUTPUT OUT=sd p=sd;

TITLE "Sampling weights denominator";

RUN;

TITLE;

PROC SORT DATA=all; BY sim i; RUN;

PROC SORT DATA=sn; BY sim i; RUN;

PROC SORT DATA=sd; BY sim i; RUN;

DATA a2; MERGE all sn sd; BY sim i;

IF s=1;

s\_wt=(sn/(1-sn))\*((1-sd)/sd);

RUN;

\*Bootstrap Africa data to get US modifier distribution;

PROC SORT DATA=a2; BY sim; RUN;

PROC SURVEYSELECT DATA=a2 OUT=a\_boot

SEED=1600123

METHOD=pps\_wr

SAMPSIZE=4000

REP=1

OUTHITS;

STRATA sim;

SIZE s\_wt;

RUN;

DATA a\_boot; SET a\_boot;

```

        genrep=replicate;
        DROP replicate expectedhits _level_ samplingweight numberhits;
RUN;
PROC SORT DATA=a_boot; BY sim genrep; RUN;

*G-computation;
*Baseline variables don't need to be modeled;
DATA a_boot; SET a_boot; pv=v; pr=r; pm=m; RUN;

%macro gen_models(n_sims=);
    DATA a_betas; SET _NULL_; RUN;
    DATA a_us_betas; SET _NULL_; RUN;
    DATA y_betas; SET _NULL_; RUN;

    %DO simn=1 %TO &n_sims.;

        *Model receipt of treatment (Africa data);
        RUN; ODS SELECT NONE; RUN;
        PROC GENMOD DATA=a_boot desc;
            WHERE r=1 AND sim=&simn.;
            BY sim genrep;
            MODEL a=m / link=logit dist=bin;
            ODS OUTPUT parameterestimates=betas(keep=sim genrep parameter estimate);
        RUN;
        RUN; ODS SELECT ALL; RUN;
        DATA betas; SET betas;
            BY sim genrep;
            RETAIN alnt am;
            IF parameter="Intercept" THEN alnt=estimate;
            IF parameter="m" THEN am=estimate;
            IF last.genrep THEN OUTPUT;
            KEEP sim genrep alnt am;
        RUN;
        DATA a_betas; SET a_betas betas; RUN;

        *Model receipt of treatment (US data);
        RUN; ODS SELECT NONE; RUN;
        PROC GENMOD DATA=us desc;
            WHERE r=1 AND sim=&simn.;
            BY sim;
            MODEL a=m / link=logit dist=bin;
            ODS OUTPUT parameterestimates=betas(keep=sim parameter estimate);
        RUN;
        RUN; ODS SELECT ALL; RUN;
        DATA betas; SET betas;

```

```

        BY sim;
        RETAIN alnt_us am_us;
        IF parameter="Intercept" THEN alnt_us=estimate;
        IF parameter="m" THEN am_us=estimate;
        IF last.sim THEN OUTPUT;
        KEEP sim alnt_us am_us;
RUN;
DATA a_us_betas; SET a_us_betas betas; RUN;

*Model the outcome (Africa data);
RUN; ODS SELECT NONE; RUN;
PROC GENMOD DATA=a_boot desc;
    WHERE sim=&simn.;
    BY sim genrep;
    MODEL y= a m v / link=logit dist=bin;
    ODS OUTPUT parameterestimates=betas(keep=sim genrep parameter estimate);
RUN;
RUN; ODS SELECT ALL; RUN;

DATA betas; SET betas;
    BY sim genrep;
    RETAIN yInt ya ym yv;
    IF parameter="Intercept" THEN yInt=estimate;
    IF parameter="a" THEN ya=estimate;
    IF parameter="m" THEN ym=estimate;
    IF parameter="v" THEN yv=estimate;
    IF last.genrep THEN OUTPUT;
    KEEP sim genrep yInt ya ym yv;
RUN;
DATA y_betas; SET y_betas betas; RUN;
%END;
%mend gen_models;

%gen_models(n_sims=2000);

DATA a_boot; MERGE a_boot a_betas y_betas; BY sim genrep; RUN;
PROC SORT DATA=a_boot; BY sim; RUN;
DATA a_boot; MERGE a_boot a_us_betas; BY sim; RUN;

*Natural Course;
DATA a_boot; SET a_boot;
    pa_n=1/(1+exp(-(alnt+am*pm)));
    py_n=1/(1+exp(-(yInt+ya*pa_n+ym*pm+yv*pv)));
RUN;
PROC MEANS DATA=a_boot mean std; VAR py_n pa_n; TITLE "Natural Course"; RUN;

```

```

*Per-Protocol;
DATA a_boot2; SET a_boot;
  *R=A=1;
  py_1=1/(1+exp(-(yInt+ya*1+ym*pm+yv*pv)));
  *R=A=0;
  py_0=1/(1+exp(-(yInt+ya*0+ym*pm+yv*pv)));
  rr=py_1/py_0;
  lnrr=log(py_1/py_0);
RUN;

DATA pp; SET a_boot2; KEEP sim genrep lnrr rr param; RUN;
PROC SORT DATA=pp; BY sim genrep; RUN;

*Summarizing over units;
PROC MEANS DATA=pp noprint;
  BY sim; var rr;
  OUTPUT OUT=est_pp(keep=sim est) mean=est;
RUN;

*Summarizing over simulations;
PROC UNIVARIATE DATA=est_pp noprint;
  VAR est;
  OUTPUT OUT=us_pp mean=rr pctlpre=p pctlpts=2.5,97.5;
RUN;

*ITT: set only R and let A be what it was in the US;
DATA a_boot3; SET a_boot;
  *R=1;
  pa_1=1/(1+exp(-(aInt+am_us*pm)));
  py_1=1/(1+exp(-(yInt+ya*pa_1+ym*pm+yv*pv)));
  *R=0 (placebo has perfect adherence in our simulation);
  pa_0=0;
  py_0=1/(1+exp(-(yInt+ya*pa_0+ym*pm+yv*pv)));
  rr=py_1/py_0;
  lnrr=log(py_1/py_0);
RUN;

DATA itt; SET a_boot3; KEEP sim genrep lnrr rr param; RUN;
PROC SORT DATA=itt; BY sim genrep; RUN;

PROC MEANS DATA=itt noprint;
  BY sim;
  VAR rr;
  OUTPUT OUT=est_itt(keep=sim est) mean=est;
RUN;

```

```
PROC UNIVARIATE DATA=est_itt noprint;
  VAR est;
  OUTPUT OUT=us_itt mean=rr pctlpre=p pctlpts=2.5,97.5;
RUN;

DATA prepsim.gcomp_est; SET afr_itt (in=ai) afr_pp (in=ap) us_itt (in=usi) us_pp (in=usp);
  LENGTH param $ 19;
  IF ai THEN param="Africa ITT";
  IF ap THEN param="Africa Per-Protocol";
  IF usi THEN param="US ITT";
  IF usp THEN param="US Per-Protocol";
RUN;
```



## References

1. Snowden JM, Rose S, Mortimer KM. Implementation of G-computation on a simulated data set: demonstration of a causal inference technique. *Am J Epidemiol*. 2011;173(7):731-738.
2. Lodi S, Sharma S, Lundgren JD, et al. The per-protocol effect of immediate versus deferred antiretroviral therapy initiation. *AIDS*. 2016;30(17):2659-2663.
3. Lesko CR, Buchanan AL, Westreich D, Edwards JK, Hudgens MG, Cole SR. Generalizing Study Results: A Potential Outcomes Perspective. *Epidemiology*. 2017;28(4):553-561.
4. Westreich D, Edwards JK, Lesko CR, Stuart E, Cole SR. Transportability of Trial Results Using Inverse Odds of Sampling Weights. *Am J Epidemiol*. 2017;186(8):1010-1014.