WEB APPENDICES

WEB APPENDIX 1: The overall effect is less than or equal to the composite effect

Below, we demonstrate that the overall causal effect is less or equal to the composite. Let the average network size be $\bar{n} = \sum_{k=1}^{K} n_k / K$. As shown in the *Estimation and Inference for Individual and Disseminated Effects* Section in the manuscript body, γ_3 is the individual effect and γ_2 is the disseminated effect.

Assume γ_2 and γ_3 have the same sign (either both negative or non-negative). The magnitude of the overall risk difference is

$$|RD^{Overall}| = |E[Y(\cdot,1)] - E[Y(\cdot,0)]|$$

$$= |E[Y|X = 1, \mathbf{Z}] - E[Y|X = 0, \mathbf{Z}]|$$

$$= |E_R\{E[Y|X = 1, R, \mathbf{Z}]\} - E_R\{E[Y|X = 0, R, \mathbf{Z}]\}|$$

$$= |\sum_{r=0}^{1}(\gamma_0 + \gamma_2 + \boldsymbol{\gamma_4 Z} + (\gamma_1 + \gamma_3)r)P(R = r)$$
$$- \sum_{r=0}^{1}(\gamma_0 + \boldsymbol{\gamma_4 Z} + \gamma_1 r)P(R = r)|$$

$$= |\gamma_2 P(R = 0) + (\gamma_2 + \gamma_3)P(R = 1)|$$

$$= |\gamma_2 + \gamma_3 E_R\{R\}|$$

$$= |\gamma_2 + \gamma_3 \times \frac{K}{N}|$$

$$= |\gamma_2 + \gamma_3 \times \frac{1}{\bar{n}}|$$

$$\leq |\gamma_2 + \gamma_3| = |RD^{Comp}|.$$

When the sign differs, the overall effect will be smaller than the composite only in certain cases. For example, if the average network size is 3, $RD^I = 1$, and $RD^D = -3$, then $RD^{Comp} = -2$, which is smaller than $RD^{Overall} = -2.67$. It is typically not expected for individual and disseminated effects to be in opposite directions, but it is technically possible.

When there is no disseminated effect, that is, $\gamma_2 = 0$, $|RD^{Overall}| = |\gamma_3 \times \frac{1}{\bar{n}}| \leq |\gamma_3| = |RD^{Comp}|$ because $\frac{1}{\bar{n}} \leq 1$. In addition, when there is no disseminated effect, the overall effect approaches the null value as $\bar{n}$ gets larger, as shown in Appendix B.

When there is no individual effect, i.e., $\gamma_3 = 0$, $RD^{Overall} = RD^{Comp} = RD^D$.

On the risk ratio scale, $RR^I = \exp(\gamma_3)$, $RR^D = \exp(\gamma_2)$, and $RR^{Comp} = \exp(\gamma_2 + \gamma_3)$. Assume γ_2 and γ_3 have the same sign (either both negative or non-negative). The overall risk ratio is

$$RR^{Overall} = \frac{E[Y(\cdot,1)]}{E[Y(\cdot,0)]} = \frac{E[Y|X=1]}{E[Y|X=0]}$$

$$= \frac{E_R\{E[Y|X=1,R,\mathbf{Z}]\}}{E_R\{E[Y|X=0,R,\mathbf{Z}]\}}$$

$$= \frac{E_R\{\exp(\gamma_0 + \gamma_1 R + \gamma_2 + \gamma_3 R + \boldsymbol{\gamma_4 Z})\}}{E_R\{\exp(\gamma_0 + \gamma_1 R + \boldsymbol{\gamma_4 Z})\}}$$

$$= \frac{\sum_{r=0}^1 \exp(\gamma_0 + \gamma_2 + \boldsymbol{\gamma_4 Z} + (\gamma_1 + \gamma_3)r)P(R=r)}{\sum_{r=0}^1 \exp(\gamma_0 + \boldsymbol{\gamma_4 Z} + \gamma_1 r)P(R=r)}$$

$$= \frac{\exp(\gamma_0 + \gamma_2 + \boldsymbol{\gamma_4 Z})P(R=0) + \exp(\gamma_0 + \gamma_1 + \gamma_2 + \gamma_3 + \boldsymbol{\gamma_4 Z})P(R=1)}{\exp(\gamma_0 + \boldsymbol{\gamma_4 Z})P(R=0) + \exp(\gamma_0 + \gamma_1 + \boldsymbol{\gamma_4 Z})P(R=1)}$$

$$= \frac{\exp(\gamma_2)\{1 - 1/\bar{n}[1 - \exp(\gamma_1 + \gamma_3)]\}}{\{1 - 1/\bar{n}[1 - \exp(\gamma_1)]\}}$$

When the average effect of being an index in a control network is null, as anticipated with the HPTN 037 study design, $\frac{E[Y|R=1,X=0,\mathbf{Z}]}{E[Y|R=0,X=0,\mathbf{Z}]} = 1 = \exp(\gamma_1)$. Then, the expression above simplifies to

$$RR^{Overall} = \exp(\gamma_2)\{1 - 1/\bar{n}[1 - \exp(\gamma_3)]\}$$

$$\leq \exp(\gamma_2)\{1 - 1[1 - \exp(\gamma_3)]\} \text{ because } 1/\bar{n} \leq 1$$

$$= \exp(\gamma_2)\exp(\gamma_3) = \exp(\gamma_2 + \gamma_3) = RR^{Comp}$$

When the sign differs, the overall effect will be smaller than the composite only in certain cases. For example, if the average network size is 3, $RR^I = 1.35$, and $RR^D = 0.78$ then $RR^{Comp} = 1.05$, which is smaller in magnitude than $RR^{Overall} = 0.87$. It is typically not expected for individual and disseminated effects to be in opposite directions, but it is technically possible.

Note that similar to as shown with the risk difference, when there is no disseminated effect, that is, $\exp(\gamma_2) = 1$, $RR^{Overall} = \{1 - 1/\bar{n}[1 - \exp(\gamma_3)] \leq \{1 - 1[1 - \exp(\gamma_3)]\} = \exp(\gamma_3) = RR^{Comp}$ because $1/\bar{n} \leq 1$. In addition, when there is no disseminated effect, the overall effect approaches the null value as $\bar{n}$ gets larger, as shown in Appendix B.

When there is no individual effect, i.e., $\exp(\gamma_3) = 1$, $RR^{Overall} = RR^{Comp} = RR^D$.

WEB APPENDIX 2: An illustration of the dependence of the overall effect on the network size

We provide a numerical example of how the overall effect is impacted by the average network size $\bar{n}$. As throughout this paper, we assume only one index per network and allow the average network size to vary. As shown in Web Appendix A, $RD^{Overall} = \gamma_2 + \gamma_3 \times \frac{1}{\bar{n}}$. Based on the illustrative example using the outcome shared "works" in the HPTN data, the disseminated effect was $RD^D = -6$ risk behavior incidents per 100 persons and the individual effect was $RD^I = -5$ risk behavior incidents per 100 persons. In the table below, we examine the impact of the size of the networks on the overall effect assuming one index per network using this formula.

Web Table 1: Overall effect as a function of the individual and disseminated effects in a network-randomized study with average network of size $\bar{n}$ with one index per network

RD^D	RD^I	RD^I as % change from RD^D	RD^{Comp}	$\bar{n}$	$RD^{Overall}$
-5	-12.5	150%	-17.5	2	-11.25
				5	-7.5
				10	-6.25
				20	-5.63
				100	-5.13
				1000	-5.01
-5	-10	100%	-15	2	-10
				5	-7
				10	-6
				20	-5.5
				100	-5.1
				1000	-5.01
-5	-2.5	50%	-7.5	2	-6.25
				5	-5.5
				10	-5.25
				20	-5.13
				100	-5.03
				1000	-5
-5	-5	0%	-10	2	-7.5
				5	-6
				10	-5.5
				20	-5.25
				100	-5.05
				1000	-5.01
0.1	-5	5000%	-4.9	2	-2.4
				5	-0.9
				10	-0.4
				20	-0.15
				100	0.05
				1000	0.01
-5	0	100%	-5	2	-5
				5	-5
				10	-5
				20	-5
				100	-5
				1000	-5

WEB APPENDIX 3: Stratified Model

Assuming a linear model with the identity link fits the data, the stratified model is

$$E[Y_{kij}|R_{ki}, X_k, \mathbf{Z}_{ki}, \mathbf{U}_{ki}] = I(R_{ki} = 0) \times (\beta_0 + \beta_1 X_k + \beta_2 \mathbf{Z}_{ki})$$

$$+I(R_{ki} = 1) \times (\alpha_0 + \alpha_1 X_k + \alpha_2 \mathbf{Z}_{ki}).$$

Then, the effect of being an index member in a control network is

$$E[Y_{kij}|R_{ki} = 1, X_k = 0, \mathbf{Z}_{ki}, \mathbf{U}_{ki}] - E[Y_{kij}|R_{ki} = 0, X_k = 0, \mathbf{Z}_{ki}, \mathbf{U}_{ki}] = \alpha_0 - \beta_0 +$$

$(\alpha_2 - \beta_2)\mathbf{Z}_{ki}$. In this stratified model, the individual rate difference (RD) can be estimated

by $\widehat{RD}_S^I = \hat{\alpha}_1 - \hat{\beta}_1$. Similarly, the disseminated RD can be estimated by $\widehat{RD}_S^D = \hat{\beta}_1$.

The composite RD can be estimated by $\widehat{RD}_S^{Comp} = \hat{\alpha}_1$. For the variance of estimator of the

individual effect, $\widehat{Var}(\widehat{RD}_S^I) = \widehat{Var}(\hat{\alpha}_1) + \widehat{Var}(\hat{\beta}_1) - 2\widehat{Cov}(\hat{\alpha}_1, \hat{\beta}_1)$, where $\widehat{Cov}(\hat{\alpha}_1, \hat{\beta}_1) =$

0 because each parameter is estimated in separate strata.

WEB APPENDIX 4: Sample SAS Code

```
/*SAMPLE SAS CODE*/
/*PURPOSE: Obtain Individual, Disseminated, Composite effect estimates*/
/*Using GEE-based methods obtain estimates of variance*/
/*that account for correlation between subjects within networks*/
/*and across visits within subjects*/
/*EXAMPLE DATASET: ANALYDATA.SAS7BDAT*/
/* mynkid = Network ID */
/* uid = Subject ID */
/* anyrisk = Any report of injection-related risk behavior*/
/* mytreat = Network-level intervention*/
/* myindex = Index status*/
/* basecov = Vector of baseline covariates*/
/*************************************************************/
options ps=60 ls=80 nodate pageno=1 nofmterr ;
dm log "clear;" continue; dm out "clear;" continue;

/*Set library with datasets*/
libname hptn 'M:\HPTN 037\sasdata';
/*load format library*/
options fmtsearch = (hptn.formats);

/*Create temporary SAS dataset*/

data analydata;
set hptn.analydata;
run;

/*Adjusted Aggregate Model to Estimate Rate Ratio*/
proc glimmix data=analydata empirical method=RMPL;
    class  mynkid uid;
      model  anyrisk=   mytreat myindex mytreat*myindex basecov
   /dist=binomial link=log s cl ;
   random intercept/ subject=mynkid;
   random intercept/ subject=uid(mynkid) ;
   estimate 'Individual'  mytreat*myindex 1 / exp cl;
   estimate 'Disseminated' mytreat 1 / exp cl;
   estimate 'Composite'   mytreat*myindex 1 mytreat 1 / exp cl;
run;

/*Adjusted Aggregate Model to Estimate Rate Difference*/
proc mixed data=analydata4  empirical method = REML;
    class uid mynkid;
      model anyrisk =  mytreat myindex  mytreat*myindex basecov/s cl;
    random intercept/ subject=mynkid;
    random intercept/ subject=uid(mynkid);
    estimate 'Individual'   mytreat*myindex 1 / cl;
    estimate 'Disseminated'  mytreat 1 / cl;
    estimate 'Composite'    mytreat*myindex 1 mytreat 1/ cl;
run;

/*Adjusted Stratified Model to Estimate Rate Ratio*/
ods output ParameterEstimates=betas covb=varb0;
proc glimmix data=analydata4 empirical method=RMPL;
    class  mynkid uid;
```

```sas
        /*Among network members R=0*/
        where myindex=0;
        model  anyrisk=  mytreat basecov/dist=binomial link=log s cl;
    random intercept/ subject=mynkid;
    random intercept/ subject=uid(mynkid) ;
     run;
        ods output close;

ods output ParameterEstimates=alphas covb=vara1;
proc glimmix data=analydata4 empirical method=RMPL;
    class  mynkid uid;
        /*Among index members R=1*/
        where myindex=1;
        model  anyrisk=  mytreat basecov/ dist=binomial link=log s cl;
        /*Only one index per network, no network-level random effect needed*/
        random intercept/ subject=uid;
    run;
        ods output close;

/*Organize model output to estimate effects*/
proc sort data=betas;
by effect;
run;
proc sort data=varb0;
by effect;
run;

data beta0 beta1;
merge betas varb0;
by effect;
if effect='Intercept' then do; merge=1; output beta0; end;
if effect='mytreat'   then do; merge=1; output beta1; end;
run;

proc sort data=alphas;
by effect;
run;
proc sort data=vara1;
by effect;
run;

data alpha0 alpha1;
merge alphas vara1;
by effect;
if effect='Intercept' then do; merge=1; output alpha0; end;
if effect='mytreat'   then do; merge=1; output alpha1; end;
run;

data ratio_results;
    merge
      beta0 (rename=(estimate=beta0 col1=varb0 col2=covb01 ))
      beta1(rename=(estimate=beta1  col2=varb1) drop=col1 )
      alpha0(rename=(estimate=alpha0 col1=vara0 col2=cova01))
      alpha1(rename=(estimate=alpha1 col2=vara1) drop=col1 );
          drop  effect;
          by merge;
          /*Compute standard errors for each parameter*/
```

```
            sea0=sqrt(vara0);
            sea1=sqrt(vara1);
            seb0=sqrt(varb0);
            seb1=sqrt(varb1);
            /*Individual effect*/
            indest2=alpha1-beta1;
            indeste2=exp(indest2);
            seind2=sqrt(sea1**2+seb1**2);
            indLCL2=exp(indest2-1.96*seind2);
            indUCL2=exp(indest2+1.96*seind2);
            /*Disseminated effect*/
            disest2=beta1;
            diseste2=exp(beta1);
            disLCL2=exp(disest2-1.96*seb1);
            disUCL2=exp(disest2+1.96*seb1);
            /*Composite effect*/
            compest2=alpha1;
            composte2=exp(compest2);
            compLCL2=exp(compest2-1.96*sea1);
            compUCL2=exp(compest2+1.96*sea1);
            keep indest2 indeste2 indLCL2 indUCL2
                disest2 diseste2 disLCl2 disUCL2
                compest2 composte2 compLCL2 compUCL2;
        run;

/*Rate Ratio Estimates and Confidence intervals based on stratified model*/
proc print data=ratio_results;
run;

/*Adjusted Stratified Model to Estimate Rate Difference*/
ods output SolutionF=betas covb=varb0;
proc mixed data=analydata4 empirical method = reml;
    class uid mynkid;
        where myindex=0;
        model indcotn =  mytreat basecov / cl s ;
        random intercept/ subject=mynkid;
        random intercept/ subject=uid(mynkid) ;
run;
ods output close;

ods output SolutionF=alphas covb=vara1;
        proc mixed data=analydata4 empirical method = reml;
    class uid mynkid;
        where myindex=1;
        model indcotn =  mytreat basecov / cl s;
/*Only one index per network, no network-level random effect needed*/
    random intercept/ subject=uid ;
    run;
ods output close;

/*Compute estimates based on stratified model*/
proc sort data=betas;
by effect;
run;
proc sort data=varb0;
by effect;
run;
```

```
data beta0 beta1;
merge betas varb0;
by effect;
if effect='Intercept' then do; merge=1; output beta0; end;
if effect='mytreat'   then do; merge=1; output beta1; end;
run;


proc sort data=alphas;
by effect;
run;
proc sort data=vara1;
by effect;
run;


data alpha0 alpha1;
merge alphas vara1;
by effect;
if effect='Intercept' then do; merge=1; output alpha0; end;
if effect='mytreat'   then do; merge=1; output alpha1; end;
run;


data difference_results;
   merge
        beta0 (rename=(estimate=beta0 col1=varb0 col2=covb01 ))
        beta1(rename=(estimate=beta1  col2=varb1) drop=col1 )
        alpha0(rename=(estimate=alpha0 col1=vara0 col2=cova01))
        alpha1(rename=(estimate=alpha1 col2=vara1) drop=col1 );
           drop effect;
           by merge;
           /*Compute standard errors for each parameter*/
           sea0=sqrt(vara0);
           sea1=sqrt(vara1);
           seb0=sqrt(varb0);
           seb1=sqrt(varb1);
           /*Individual effect*/
           indest2=alpha1-beta1;
           seind2=sqrt(sea1**2+seb1**2);
           indLCL2=indest2-1.96*seind2;
           indUCL2=indest2+1.96*seind2;
           /*Disseminated effect*/
           disest2=beta1;
           disLCL2=disest2-1.96*seb1;
           disUCL2=disest2+1.96*seb1;
           /*Composite effect*/
           compest2=alpha1;
           compLCL2=compest2-1.96*sea1;
           compUCL2=compest2+1.96*sea1;
       keep indest2 indLCL2 indUCL2
            disest2 disLCl2 disUCL2
            compest2 compLCL2 compUCL2;
       run;

/*Rate Difference Estimates and Confidence intervals based on stratified
model*/
       proc print data=difference_results;
       run;
```