Risk ratios for contagious outcomes: Simulation code

Olga Morozova, Ted Cohen, Forrest W. Crawford

J. R. Soc. Interface

In this document we provide the code that allows replication of the simulation results presented in the paper and in the Supplement. Every piece of code presented below calls a function from the source file 'simulation_functions.R', which implements one simulation for β=0 and γ=0 for the set of other parameters used to produce a given figure in the paper and the Supplement. The actual figures were produced by using the same functions for a range of β and γ between -5 and 5 with a step of 0.2. Details regarding the number of simulations per combination of parameters that were used to produce the figures are provided in the Supplement.

```
source('simulation_functions.R')
```

Figures presented in the section 'Simulation results' of the paper

Figure 4: Corresponds to Supplement Figure 3 (see below).

Constant cluster size (4 subjects in each cluster), constant observation time, no infections at baseline, block randomized distribution of covariate x with k=1, k=2 and k=3.

Figure 5: Corresponds to Supplement Figure 5 (see below).

Constant cluster size (4 subjects in each cluster), constant observation time, no infections at baseline, cluster randomized distribution of covariate x with p=0.25, p=0.5, p=0.75

Figure 6: Corresponds to Supplement Figure 6 (see below).

Variable cluster size (Poisson distributed), constant observation time, no infections at baseline, block randomized distribution of covariate x with k=1.

Figure 7: Corresponds to Supplement Figure 7 (see below).

Variable cluster size (Poisson distributed), constant observation time, no infections at baseline, block randomized distribution of covariate x with half of subjects per cluster having x=1.

Figures presented in the Supplement

4.2.1. Distribution of covariate x when cluster size is constant

Supplement Figure 3: Constant cluster size (4 subjects in each cluster), constant observation time, no infections at baseline, block randomized distribution of covariate x with k=1, k=2 and k=3 (corresponds to Figure 4 in the paper).

```
res<-sim.nc.clusters.ef3(nc=500,hh_size=4,tx_prob=0.25,fu_time=450,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] 0.01481509
```

```
res<-sim.nc.clusters.ef3(nc=500,hh_size=4,tx_prob=0.5,fu_time=450,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] 0.05406722
```

```
res<-sim.nc.clusters.ef3(nc=500,hh_size=4,tx_prob=0.75,fu_time=450,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] 0.05781957
```

Supplement Figure 4: Constant cluster size (4 subjects in each cluster), constant observation time, no infections at baseline, simple Bernoulli randomized distribution of covariate x with Pr(x=1)=0.25, Pr(x=1)=0.5 and Pr(x=1)=0.75.

```
res<-sim.nc.clusters.ef4(nc=500,hh_size=4,tx_prob=0.25,fu_time=450,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] -0.1149087
```

```
res<-sim.nc.clusters.ef4(nc=500,hh_size=4,tx_prob=0.5,fu_time=450,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] 0.0441313
```

```
res<-sim.nc.clusters.ef4(nc=500,hh_size=4,tx_prob=0.75,fu_time=450,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] 0.03017088
```

Supplement Figure 5: Constant cluster size (4 subjects in each cluster), constant observation time, no infections at baseline, cluster randomized distribution of covariate x with p=0.25, p=0.5, p=0.75 (corresponds to Figure 5 in the paper).

```
res<-sim.nc.clusters.ef5(nc=500,hh_size=4,tx_prob=0.25,fu_time=450,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] -0.2958454
```

```
res<-sim.nc.clusters.ef5(nc=500,hh_size=4,tx_prob=0.5,fu_time=450,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] 0.04652002
```

```
res<-sim.nc.clusters.ef5(nc=500,hh_size=4,tx_prob=0.75,fu_time=450,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] -0.425874
```

4.2.2. Variable cluster size under different distributions of covariate x

Supplement Figure 6: Variable cluster size (Poisson distributed), constant observation time, no infections at baseline, block randomized distribution of covariate x with k=1 (corresponds to Figure 6 in the paper).

```
res<-sim.nc.clusters.ef6(nc=500,hh_size=2,fu_time=750,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] -0.4218162
```

```
res<-sim.nc.clusters.ef6(nc=500,hh_size=3,fu_time=525,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] -0.1350039
```

```
res<-sim.nc.clusters.ef6(nc=500,hh_size=4,fu_time=450,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] -0.1756788
```

```
res<-sim.nc.clusters.ef6(nc=500,hh_size=5,fu_time=330,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] -0.2687586
```

Supplement Figure 7: Variable cluster size (Poisson distributed), constant observation time, no infections at baseline, block randomized distribution of covariate x with half of subjects per cluster having x=1 (corresponds to Figure 7 in the paper).

```
res<-sim.nc.clusters.ef7(nc=500,hh_size=2,tx_prob=0.5,fu_time=750,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] 0.3220224
```

```
res<-sim.nc.clusters.ef7(nc=500,hh_size=3,tx_prob=0.5,fu_time=525,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] -0.01249816
```

```
res<-sim.nc.clusters.ef7(nc=500,hh_size=4,tx_prob=0.5,fu_time=450,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] 0.03613971
```

```
res<-sim.nc.clusters.ef7(nc=500,hh_size=5,tx_prob=0.5,fu_time=330,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] 0.02868321
```

Supplement Figure 8: Variable cluster size (Poisson distributed), constant observation time, no infections at baseline, simple Bernoulli randomized distribution of covariate x with Pr(x=1)=0.5.

```
res<-sim.nc.clusters.ef8(nc=500,hh_size=2,tx_prob=0.5,fu_time=750,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] 0.2492964
```

```
res<-sim.nc.clusters.ef8(nc=500,hh_size=3,tx_prob=0.5,fu_time=525,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] -0.07500167
```

```
res<-sim.nc.clusters.ef8(nc=500,hh_size=4,tx_prob=0.5,fu_time=450,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] -0.1263711
```

```
res<-sim.nc.clusters.ef8(nc=500,hh_size=5,tx_prob=0.5,fu_time=330,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] 0.01170061
```

Supplement Figure 9: Variable cluster size (Poisson distributed), constant observation time, no infections at baseline, cluster randomized distribution of covariate x with p=0.5.

```
res<-sim.nc.clusters.ef9(nc=500,hh_size=2,tx_prob=0.5,fu_time=750,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] -0.7693637
```

```
res<-sim.nc.clusters.ef9(nc=500,hh_size=3,tx_prob=0.5,fu_time=525,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] -0.253412
```

```
res<-sim.nc.clusters.ef9(nc=500,hh_size=4,tx_prob=0.5,fu_time=450,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] 0.1804806
```

```
res<-sim.nc.clusters.ef9(nc=500,hh_size=5,tx_prob=0.5,fu_time=330,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] -0.1644285
```

4.2.3 Duration and variability of observation time Ti

Supplement Figure 10: Constant cluster size (4 subjects in each cluster), no infections at baseline, block randomized distribution of covariate x with k=2. In first three plots observation time is fixed at the values Ti=50, Ti=450 and Ti=1500; and these three plots are generated using the function ‘sim.nc.clusters.ef3’. The fourth plot uses variable (exponentially distributed) observation time with a mean of 450. The fourth plot is generated using the function ‘sim.nc.clusters.ef10’.

```
res<-sim.nc.clusters.ef3(nc=500,hh_size=4,tx_prob=0.5,fu_time=50,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] 0.1541507
```

```
res<-sim.nc.clusters.ef3(nc=500,hh_size=4,tx_prob=0.5,fu_time=450,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] 0.05244648
```

```
res<-sim.nc.clusters.ef3(nc=500,hh_size=4,tx_prob=0.5,fu_time=1500,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] 0.002333723
```

```
res<-sim.nc.clusters.ef10(nc=500,hh_size=4,tx_prob=0.5,fu_time=450,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] 0
```

Supplement Figure 11: Variable cluster size (Poisson distributed), no infections at baseline, simple Bernoulli randomized distribution of covariate x with Pr(x=1)=0.5. In first three plots observation time is fixed at the values Ti=50, Ti=450 and Ti=1500; and these three plots are generated using the function 'sim.nc.clusters.ef8'. The fourth plot uses variable (exponentially distributed) observation time with a mean of 450. The fourth plot is generated using the function 'sim.nc.clusters.ef11'.

```
res<-sim.nc.clusters.ef8(nc=500,hh_size=4,tx_prob=0.5,fu_time=50,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] 0.8008208
```

```
res<-sim.nc.clusters.ef8(nc=500,hh_size=4,tx_prob=0.5,fu_time=450,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] 0.05236446
```

```
res<-sim.nc.clusters.ef8(nc=500,hh_size=4,tx_prob=0.5,fu_time=1500,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] 0.01417008
```

```
res<-sim.nc.clusters.ef11(nc=500,hh_size=4,tx_prob=0.5,fu_time=450,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] 0.1640715
```

Supplement Figure 12: Constant cluster size (4 subjects in each cluster), no infections at baseline, cluster randomized distribution of covariate x with p=0.5. In first three plots observation time is fixed at the values Ti=50, Ti=450 and Ti=1500; and these three plots are generated using the function ‘sim.nc.clusters.ef5’. The fourth plot uses variable (exponentially distributed) observation time with a mean of 450. The fourth plot is generated using the function ‘sim.nc.clusters.ef12’.

```
res<-sim.nc.clusters.ef5(nc=500,hh_size=4,tx_prob=0.5,fu_time=50,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] 0.6931472
```

```
res<-sim.nc.clusters.ef5(nc=500,hh_size=4,tx_prob=0.5,fu_time=450,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] 0.3262969
```

```
res<-sim.nc.clusters.ef5(nc=500,hh_size=4,tx_prob=0.5,fu_time=1500,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] -0.07267119
```

```
res<-sim.nc.clusters.ef12(nc=500,hh_size=4,tx_prob=0.5,fu_time=450,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] 0.01360565
```

4.2.4 Infections at time zero

Supplement Figures 13 and 14: Constant cluster size (4 subjects in each cluster), constant observation time, block randomized distribution of covariate x with k=2. Conditional probabilities of being infected at baseline: Pr[Y(0)=1|x=1]=p; and Pr[Y(0)=1|X=1]=q.

```
res<-sim.nc.clusters.ef13(nc=500,hh_size=4,tx_prob=0.5,fu_time=450,p=0,q=0,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] 0.02247286
```

```
res<-sim.nc.clusters.ef13(nc=500,hh_size=4,tx_prob=0.5,fu_time=150,p=0,q=0.1,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] 0.5131889
```

```
res<-sim.nc.clusters.ef13(nc=500,hh_size=4,tx_prob=0.5,fu_time=18,p=0,q=0.5,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] 0.5940172
```

```
res<-sim.nc.clusters.ef13(nc=500,hh_size=4,tx_prob=0.5,fu_time=9,p=0,q=0.9,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] 0.5590442
```

```
res<-sim.nc.clusters.ef13(nc=500,hh_size=4,tx_prob=0.5,fu_time=150,p=0.1,q=0,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] -0.5222805
```

```
res<-sim.nc.clusters.ef13(nc=500,hh_size=4,tx_prob=0.5,fu_time=60,p=0.1,q=0.1,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] -0.004405293
```

```
res<-sim.nc.clusters.ef13(nc=500,hh_size=4,tx_prob=0.5,fu_time=16,p=0.1,q=0.5,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] 0.6094989
```

```
res<-sim.nc.clusters.ef13(nc=500,hh_size=4,tx_prob=0.5,fu_time=8,p=0.1,q=0.9,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] 0.6535716
```

```
res<-sim.nc.clusters.ef13(nc=500,hh_size=4,tx_prob=0.5,fu_time=18,p=0.5,q=0,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] -0.5495988
```

```
res<-sim.nc.clusters.ef13(nc=500,hh_size=4,tx_prob=0.5,fu_time=16,p=0.5,q=0.1,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] -0.3226831
```

```
res<-sim.nc.clusters.ef13(nc=500,hh_size=4,tx_prob=0.5,fu_time=11,p=0.5,q=0.5,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] -0.1876387
```

```
res<-sim.nc.clusters.ef13(nc=500,hh_size=4,tx_prob=0.5,fu_time=7,p=0.5,q=0.9,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] -0.03719283
```

```
res<-sim.nc.clusters.ef13(nc=5000,hh_size=4,tx_prob=0.5,fu_time=9,p=0.9,q=0,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] -0.5479994
```

```
res<-sim.nc.clusters.ef13(nc=5000,hh_size=4,tx_prob=0.5,fu_time=8,p=0.9,q=0.1,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] -0.5519346
```

```
res<-sim.nc.clusters.ef13(nc=500,hh_size=4,tx_prob=0.5,fu_time=7,p=0.9,q=0.5,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] -0.133169
```

```
res<-sim.nc.clusters.ef13(nc=500,hh_size=4,tx_prob=0.5,fu_time=5,p=0.9,q=0.9,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] 0.6690496
```

Supplement Figures 15 and 16: Variable cluster size (Poisson distributed), constant observation time, simple Bernoulli randomized distribution of covariate x with Pr(x=1)=0.5. Conditional probabilities of being infected at baseline: Pr[Y(0)=1|x=1]=p; and Pr[Y(0)=1|X=1]=q.

```
res<-sim.nc.clusters.ef15(nc=500,hh_size=4,tx_prob=0.5,fu_time=450,p=0,q=0,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] -0.0480066
```

```
res<-sim.nc.clusters.ef15(nc=500,hh_size=4,tx_prob=0.5,fu_time=150,p=0,q=0.1,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] -0.05323764
```

```
res<-sim.nc.clusters.ef15(nc=500,hh_size=4,tx_prob=0.5,fu_time=18,p=0,q=0.5,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] 0.06740993
```

```
res<-sim.nc.clusters.ef15(nc=500,hh_size=4,tx_prob=0.5,fu_time=9,p=0,q=0.9,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] 0.3814465
```

```
res<-sim.nc.clusters.ef15(nc=500,hh_size=4,tx_prob=0.5,fu_time=150,p=0.1,q=0,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] 0.08687489
```

```
res<-sim.nc.clusters.ef15(nc=500,hh_size=4,tx_prob=0.5,fu_time=60,p=0.1,q=0.1,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] 0.02960138
```

```
res<-sim.nc.clusters.ef15(nc=500,hh_size=4,tx_prob=0.5,fu_time=16, p=0.1,q=0.5,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01
)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] -0.01392258
```

```
res<-sim.nc.clusters.ef15(nc=500,hh_size=4,tx_prob=0.5,fu_time=8,p=0.1,q=0.9,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] -0.02240341
```

```
res<-sim.nc.clusters.ef15(nc=5000,hh_size=4,tx_prob=0.5,fu_time=18,p=0.5,q=0,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] -0.02027824
```

```
res<-sim.nc.clusters.ef15(nc=500,hh_size=4,tx_prob=0.5,fu_time=16,p=0.5,q=0.1,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] -0.1100009
```

```
res<-sim.nc.clusters.ef15(nc=500,hh_size=4,tx_prob=0.5,fu_time=11,p=0.5,q=0.5,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] 0.04434493
```

```
res<-sim.nc.clusters.ef15(nc=500,hh_size=4,tx_prob=0.5,fu_time=7,p=0.5,q=0.9,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] -0.1709538
```

```
res<-sim.nc.clusters.ef15(nc=5000,hh_size=4,tx_prob=0.5,fu_time=9,p=0.9,q=0,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] -0.01722264
```

```
res<-sim.nc.clusters.ef15(nc=1000,hh_size=4,tx_prob=0.5,fu_time=8,p=0.9,q=0.1,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] -0.2352435
```

```
res<-sim.nc.clusters.ef15(nc=500,hh_size=4,tx_prob=0.5,fu_time=7,p=0.9,q=0.5,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] -0.1853396
```

```
res<-sim.nc.clusters.ef15(nc=500,hh_size=4,tx_prob=0.5,fu_time=5,p=0.9,q=0.9,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
res.uninf <- subset(res,Yi0==0)
log.rr <- log(mean(res.uninf$Y[res.uninf$Xi==1]) / mean(res.uninf$Y[res.uninf$Xi==0]))
log.rr
```

```
## [1] -0.1106656
```

Supplement Figure 17: Constant cluster size (4 subjects in each cluster), constant observation time, block randomized distribution of covariate x with k=2. Clusters are selected based on having at least one infection at “baseline”, which are obtained by simulating the infection transmission on the larger number of clusters using the function ‘sim.nc.clusters.ef3’.

```r
res<-sim.nc.clusters.ef3(nc=16722,hh_size=4,tx_prob=0.5,fu_time=75,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
a <- aggregate(res$Y, by=list(HHid=res$HHid), FUN=sum)
bl.inf <- which(a$x>0)
res.bl.inf <- subset(res, HHid %in% bl.inf)
res.bl.inf.split <- split.data.frame(res.bl.inf, as.factor(res.bl.inf$HHid))
dres <- list()
for (l in 1:length(bl.inf)){
cur.hh <- res.bl.inf.split[[l]]
cur.bl.inf <- which(cur.hh$Y==1)
if (length(cur.bl.inf)==1) {k <- cur.bl.inf} else {k <- sample(cur.bl.inf, 1)}
index <- rep(0,length(cur.hh$Xi))
index[k] <- 1
Y.end<-sim.one.cluster(Xi=cur.hh$Xi,Yi0=cur.hh$Y,Ti=10,t.beta=0,t.gam=0,t.alpha=0.0001, t.omega=0.01)
dres[[l]]<-data.frame(HHid=cur.hh$HHid,Ti1=cur.hh$Ti,index=index,Ti2=10,Xi=cur.hh$Xi,Yi0=cur.hh$Y,Y.end=Y.end)
}
dres <- bind_rows(dres)
dres.noindex <- subset(dres, index==0)
dres.uninf <- subset(dres, Yi0==0)
log.rr.noindex <- log(mean(dres.noindex$Y.end[dres.noindex$Xi==1]) / mean(dres.noindex$Y.end[dres.noindex$Xi==0]))
log.rr.uninf <- log(mean(dres.uninf$Y.end[dres.uninf$Xi==1]) / mean(dres.uninf$Y.end[dres.uninf$Xi==0]))
```

```r
log.rr.noindex
```

```
## [1] -0.01519296
```

```r
log.rr.uninf
```

```
## [1] -0.03005147
```

Supplement Figure 18: Variable cluster size (Poisson distributed), constant observation time, simple Bernoulli randomized distribution of covariate x with Pr(x=1)=0.5. Clusters are selected based on having at least one infection at “baseline”, which are obtained by simulating the infection transmission on the larger number of clusters using the function ‘sim.nc.clusters.ef18’.

```
res<-sim.nc.clusters.ef18(nc=8571,hh_size=4,tx_prob=0.5,fu_time=150,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
a <- aggregate(res$Y, by=list(HHid=res$HHid), FUN=sum)
bl.inf <- which(a$x>0)
res.bl.inf <- subset(res, HHid %in% bl.inf)
res.bl.inf.split <- split.data.frame(res.bl.inf, as.factor(res.bl.inf$HHid))
dres <- list()
for (l in 1:length(bl.inf)){
cur.hh <- res.bl.inf.split[[l]]
cur.bl.inf <- which(cur.hh$Y==1)
if (length(cur.bl.inf)==1) {k <- cur.bl.inf} else {k <- sample(cur.bl.inf, 1)}
index <- rep(0,length(cur.hh$Xi))
index[k] <- 1
Y.end<-sim.one.cluster(Xi=cur.hh$Xi,Yi0=cur.hh$Y,Ti=10,t.beta=0,t.gam=0,t.alpha=0.0001, t.omega=0.01)
dres[[l]]<-data.frame(HHid=cur.hh$HHid,Ti1=cur.hh$Ti,index=index,Ti2=10,Xi=cur.hh$Xi,Yi0=cur.hh$Y,Y.end=Y.end)
}
dres <- bind_rows(dres)
dres.noindex <- subset(dres, index==0)
dres.uninf <- subset(dres, Yi0==0)
log.rr.noindex <- log(mean(dres.noindex$Y.end[dres.noindex$Xi==1]) / mean(dres.noindex$Y.end[dres.noindex$Xi==0]))
log.rr.uninf <- log(mean(dres.uninf$Y.end[dres.uninf$Xi==1]) / mean(dres.uninf$Y.end[dres.uninf$Xi==0]))
```

```
log.rr.noindex
```

```
## [1] -0.02538892
```

```
log.rr.uninf
```

```
## [1] -0.1846947
```

4.2.5 Ratio ω/α

Supplement Figure 19: Constant cluster size (4 subjects in each cluster), constant observation time, no infections at baseline, block randomized distribution of covariate x with k=2. All three plots in this figure are generated using the function ‘sim.nc.clusters.ef3’ with different values of ω, α and observation time Ti.

```
res<-sim.nc.clusters.ef3(nc=500,hh_size=4,tx_prob=0.5,fu_time=450,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] 0.03460553
```

```
res<-sim.nc.clusters.ef3(nc=500,hh_size=4,tx_prob=0.5,fu_time=750,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.001)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] 0.05644131
```

```
res<-sim.nc.clusters.ef3(nc=500,hh_size=4,tx_prob=0.5,fu_time=75,t.beta=0,t.gam=0,t.alpha=0.001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] -0.1170566
```

Supplement Figure 20: Variable cluster size (Poisson distributed), constant observation time, no infections at baseline, simple Bernoulli randomized distribution of covariate x with Pr(x=1)=0.5. All three plots in this figure are generated using the function ‘sim.nc.clusters.ef8’ with different values of ω, α and observation time Ti.

```
res<-sim.nc.clusters.ef8(nc=500,hh_size=4,tx_prob=0.5,fu_time=450,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] -0.06623484
```

```
res<-sim.nc.clusters.ef8(nc=500,hh_size=4,tx_prob=0.5,fu_time=750,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.001)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] -0.03244636
```

```
res<-sim.nc.clusters.ef8(nc=500,hh_size=4,tx_prob=0.5,fu_time=75,t.beta=0,t.gam=0,t.alpha=0.001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] 0.02442996
```

Supplement Figure 21: Variable cluster size (Poisson distributed), constant observation time, no infections at baseline, block randomized distribution of covariate x with k=1. All five plots in this figure are generated using the function ‘sim.nc.clusters.ef6’ with different values of ω, α and observation time Ti.

```
res<-sim.nc.clusters.ef6(nc=500,hh_size=4,fu_time=3600,t.beta=0,t.gam=0,t.alpha=0.00001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] -0.2325195
```

```
res<-sim.nc.clusters.ef6(nc=500,hh_size=4,fu_time=450,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] -0.2239663
```

```
res<-sim.nc.clusters.ef6(nc=500,hh_size=4,fu_time=750,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.001)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] -0.2806185
```

```
res<-sim.nc.clusters.ef6(nc=500,hh_size=4,fu_time=130,t.beta=0,t.gam=0,t.alpha=0.001,t.omega=0.001)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] -0.145079
```

```
res<-sim.nc.clusters.ef6(nc=500,hh_size=4,fu_time=160,t.beta=0,t.gam=0,t.alpha=0.001,t.omega=0.0001)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] -0.008620743
```

Supplement Figure 22: Variable cluster size (Poisson distributed), constant observation time, no infections at baseline, cluster randomized distribution of covariate x with p=0.5. All five plots in this figure are generated using the function ‘sim.nc.clusters.ef9’ with different values of ω, α and observation time Ti.

```
res<-sim.nc.clusters.ef9(nc=500,hh_size=4,tx_prob=0.5,fu_time=3600,t.beta=0,t.gam=0,t.alpha=0.00001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] -0.1519638
```

```
res<-sim.nc.clusters.ef9(nc=500,hh_size=4,tx_prob=0.5,fu_time=450,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.01)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] 0.2719212
```

```
res<-sim.nc.clusters.ef9(nc=500,hh_size=4,tx_prob=0.5,fu_time=750,t.beta=0,t.gam=0,t.alpha=0.0001,t.omega=0.001)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] -0.1282791
```

```
res<-sim.nc.clusters.ef9(nc=500,hh_size=4,tx_prob=0.5,fu_time=130,t.beta=0,t.gam=0,t.alpha=0.001,t.omega=0.001)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] -0.2313543
```

```
res<-sim.nc.clusters.ef9(nc=500,hh_size=4,tx_prob=0.5,fu_time=160,t.beta=0,t.gam=0,t.alpha=0.001,t.omega=0.0001)
log.rr <- log(mean(res$Y[res$Xi==1]) / mean(res$Y[res$Xi==0]))
log.rr
```

```
## [1] -0.07895505
```