Journal Title
XX(X):1–30

DOI: 10.1177/ToBeAssigned
www.sagepub.com/
SAGE

Web Appendix for “Sample size estimation for modified Poisson analysis of cluster randomized trials with a binary outcome”

Fan Li[1,2,3] **and Guangyu Tong**[1,3]

A. R Code for Modified Poisson Regression and Bias-corrected Variances

A.1 Independence Working Correlation

Under the independence working correlation, the point estimate can be easily obtained simply using the `glm` function as

```
fit1=glm(y~trt,family=poisson(link="log"),data=wdata)
```

where `y` is the vector of outcome and `trt` is the vector of clustered treatment indicator, and `wdata` is the data set containing this information. Based on the point estimates obtained from the `fit1` object, the bias-corrected variances can then by obtained using the following function:

```
#######################################################
# Bias-corrected Variance of Modified Poisson Analysis
#######################################################

########################################
# Input
```

---

[1] Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA
[2] Center for Methods in Implementation and Preventive Science, Yale University, New Haven, CT, USA
[3] Yale Center for Analytical Sciences, Yale University, New Haven, CT, USA

Corresponding author:
Fan Li, Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA.
Email: fan.f.li@yale.edu

```
# y: N by 1 vector of outcomes
# X: N by p design matrix (including intercept)
# beta: p by 1 mean model parameter estimates
# phi: 1 by 1 dispersion parameter estimate
# id: N by 1 cluster indicator
########################################

LogPoisBCV=function(y,X,beta,phi,id){

  # Creates two vectors that have the start and end points for each cluster
  BEGINEND=function(n){
    last=cumsum(n)
    first=last-n+1
    return(cbind(first,last))
  }

  # Score function
  SCORE=function(beta,phi,y,X,n,p){
    U=rep(0,p)
    UUtran=Ustar=matrix(0,p,p)
    locx=BEGINEND(n)

    for(i in 1:length(n)){
      X_c=X[locx[i,1]:locx[i,2],,drop=FALSE]
      y_c=y[locx[i,1]:locx[i,2]]

      U_c=rep(0,p)
      Ustar_c=matrix(0,p,p)
      mu_c=exp(c(X_c%*%beta))

      C=X_c*mu_c
      A=y_c-mu_c
      INVB=diag(1/mu_c,n[i])/phi

      U_c=t(C)%*%INVB%*%A
      UUtran_c=tcrossprod(U_c)
      Ustar_c=t(C)%*%INVB%*%C
      U=U+U_c
      UUtran=UUtran+UUtran_c
      Ustar=Ustar+Ustar_c
    }
    return(list(U=U,UUtran=UUtran,Ustar=Ustar))
  }
```

```
# Compute (A - mm')^{-1}c without performing the inverse directly
INVBIG=function(ainvc,ainvm,m,c,start,end){
  for(i in start:end){
    b=ainvm[,i]
    bt=t(b)
    btm=bt%*%m
    btmi=btm[,i]
    gam=1-btmi
    bg=b/gam
    ainvc=ainvc+bg%*%(bt%*%c)
    if(i<end){
      ainvm=ainvm+bg%*%btm
    }
  }
  return(ainvc)
}

# Creates bias-corrected covariance matrix of beta
p=ncol(X)
n=as.numeric(table(id))
SCORE_RES=SCORE(beta,phi,y,X,n,p)
U=SCORE_RES$U
UUtran=SCORE_RES$UUtran
Ustar=SCORE_RES$Ustar

# Naive or Model-based estimator
naive=solve(Ustar)

# BC0 or usual Sandwich estimator
robust=naive%*%UUtran%*%t(naive)

# new commands to compute INV(I - H1)
eigenRES1=eigen(naive)
evals1=eigenRES1$values
evecs1=eigenRES1$vectors
sqrevals1=sqrt(evals1)
sqe1=evecs1%*%diag(sqrevals1)

# Bias-corrected variance
Ustar_c_array=UUtran_c_array=array(0,c(p,p,length(n)))
UUtran=UUbc=UUbc2=UUbc3=Ustar=matrix(0,p,p)
```

```
locx=BEGINEND(n)

for(i in 1:length(n)){
  X_c=X[locx[i,1]:locx[i,2],,drop=FALSE]
  y_c=y[locx[i,1]:locx[i,2]]
  mu_c=exp(c(X_c%*%beta))

  U_i=U_c=rep(0,p)
  Ustar_c=matrix(0,p,p)

  # commands for beta
  C=X_c*mu_c
  A=y_c-mu_c
  INVB=diag(1/mu_c,n[i])/phi
  U_i=t(C)%*%INVB%*%A

  # commands for generalized inverse - beta
  ai1=INVB
  mm1=C%*%sqe1
  ai1A=ai1%*%A
  ai1m1=ai1%*%mm1
  ai1A=INVBIG(ai1A,ai1m1,mm1,A,1,p)
  U_c=t(C)%*%ai1A

  Ustar_c=t(C)%*%INVB%*%C
  Ustar=Ustar+Ustar_c
  UUtran_c=tcrossprod(U_i)
  UUtran=UUtran+UUtran_c
  UUbc_c=tcrossprod(U_c)
  UUbc=UUbc+UUbc_c
  UUbc_ic=tcrossprod(U_c,U_i)
  UUbc2=UUbc2+UUbc_ic

  Ustar_c_array[,,i]=Ustar_c
  UUtran_c_array[,,i]=UUtran_c
}

# calculating adjustment factor for BC3
for(i in 1:length(n)){
  Hi=diag(1/sqrt(1-pmin(0.75,c(diag(Ustar_c_array[,,i]%*%naive)))))
  UUbc3=UUbc3+Hi%*%UUtran_c_array[,,i]%*%Hi
}
```

```
  # BC1 or Variance estimator due to Kauermann and Carroll (2001);
  varKC=naive%*%(UUbc2+t(UUbc2))%*%t(naive)/2

  # BC2 or Variance estimator due to Mancl and DeRouen (2001);
  varMD=naive%*%UUbc%*%t(naive)

  # BC3 or Variance estimator due to Fay and Graubard (2001);
  varFG=naive%*%UUbc3%*%t(naive)

  ########################################
  # Output
  # naive: naive or model-based var
  # robust: robust sandwich var
  # varMD: bias-corrected sandwich var due to Mancl and DeRouen (2001)
  # varKC: bias-corrected sandwich var due to Kauermann and Carroll (2001)
  # varFG: bias-corrected sandwich var due to Fay and Graubard (2001)
  ########################################
  return(list(naive=naive,robust=robust,varMD=varMD,varKC=varKC,varFG=varFG))
}
```

Example usage of this function is

```
LogPoisBCV(y=y, X=cbind(1,trt), beta=as.numeric(coef(fit1)),
phi=1, id=id)
```

where `id` is the (sorted) vector of cluster identifiers.

A.2 Exchangeable Working Correlation

We provide a user-written function that implements the modified correlation estimator in Section 2.3 of the main text. The function also provides the bias-corrected variances.

```
############################################################
# Modified Poisson GEE analysis of cluster randomized trials

# INPUT
# y: The binary outcome variable
# X: Marginal mean covariates (design matrix including intercept)
# id: Cluster identifier
# n: Vector of cluster sample sizes
# maxiter: Maximum number of iterations for Fisher Scoring
# epsilon: Tolerance for convergence

# Note that all inputs are required.
# ID's should be integers from 1 to K.
# Data should be sorted by ID before calling.
```

```
###########################################################

binMPreg=function(y, X, id, n, maxiter=50, epsilon=0.00001){
  require(MASS)

  # Creates two vectors that have the start
  # and end points for each cluster
  BEGINEND=function(n){
    last=cumsum(n)
    first=last-n+1
    return(cbind(first,last))
  }

  is_pos_def=function(A){
    return(min(eigen(A)$values)>1e-13)
  }

  # Score function
  SCORE=function(beta, alpha, y, X, n, p){
    U=rep(0,p)
    UUtran=Ustar=matrix(0,p,p)
    locx=BEGINEND(n)

    for(i in 1:length(n)){
      X_c=X[locx[i,1]:locx[i,2],,drop=FALSE]
      y_c=y[locx[i,1]:locx[i,2]]

      U_c=rep(0,p)
      Ustar_c=matrix(0,p,p)
      mu_c=exp(c(X_c%*%beta))

      C=X_c*mu_c
      A=y_c-mu_c
      INVR=diag(1/(1-alpha),n[i])-
        matrix(alpha/((1-alpha)*(1-alpha+n[i]*alpha)),n[i],n[i])
      INVB=diag(1/sqrt(mu_c),n[i]) %*%
        INVR %*% diag(1/sqrt(mu_c),n[i])

      U_c=t(C)%*%INVB%*%A
      UUtran_c=tcrossprod(U_c)
      Ustar_c=t(C)%*%INVB%*%C
      U=U+U_c
      UUtran=UUtran+UUtran_c
```

```
    Ustar=Ustar+Ustar_c
  }
  return(list(U=U,UUtran=UUtran,Ustar=Ustar))
}

# Generates initial values for beta
INITBETA=function(y,X){
  beta=as.numeric(coef(glm(y~-1+X,
                           family=poisson(link=log))))
  return(beta)
}

# Compute ALPHA
getalpha=function(y,X,beta,n){
  locx=BEGINEND(n)
  alpsum=0
  for(i in 1:length(n)){
    X_c=X[locx[i,1]:locx[i,2],,drop=FALSE]
    y_c=y[locx[i,1]:locx[i,2]]
    mu_c=exp(c(X_c%*%beta))
    r=(y_c-mu_c)/sqrt(mu_c*(1-mu_c))
    rrtran=tcrossprod(r)
    alpsum=alpsum+sum(rrtran[upper.tri(rrtran)])
  }
  alpha=alpsum/sum(n*(n-1)/2-2)
  return(alpha)
}

# compute (A - mm‘)^{-1}c without performing the inverse directly
INVBIG=function(ainvc,ainvm,m,c,start,end){
  for(i in start:end){
    b=ainvm[,i]
    bt=t(b)
    btm=bt%*%m
    btmi=btm[,i]
    gam=1-btmi
    bg=b/gam
    ainvc=ainvc+bg%*%(bt%*%c)
    if(i<end){
      ainvm=ainvm+bg%*%btm
    }
  }
  return(ainvc)
```

```
}

# Creates covariance matrix of beta
MAKEVAR=function(beta, alpha, y, X, n, p){

  SCORE_RES=SCORE(beta, alpha, y, X, n, p)
  U=SCORE_RES$U
  UUtran=SCORE_RES$UUtran
  Ustar=SCORE_RES$Ustar

  naive=ginv(Ustar)

  # new commands to compute INV(I - H1)
  eigenRES1=eigen(naive)
  evals1=eigenRES1$values
  evecs1=eigenRES1$vectors
  sqrevals1=sqrt(evals1)
  sqe1=evecs1%*%diag(sqrevals1)

  # Bias-corrected variance
  Ustar_c_array=UUtran_c_array=
    array(0,c(p,p,length(n)))
  UUtran=UUbc=UUbc2=UUbc3=Ustar=
    matrix(0,p,p)

  locx=BEGINEND(n)

  for(i in 1:length(n)){
    X_c=X[locx[i,1]:locx[i,2],,drop=FALSE]
    y_c=y[locx[i,1]:locx[i,2]]
    mu_c=exp(c(X_c%*%beta))

    U_i=U_c=rep(0,p)
    Ustar_c=matrix(0,p,p)

    # commands for beta
    C=X_c*mu_c
    A=y_c-mu_c
    INVR=diag(1/(1-alpha),n[i])-
      matrix(alpha/((1-alpha)*(1-alpha+n[i]*alpha)),n[i],n[i])
    INVB=diag(1/sqrt(mu_c),n[i]) %*%
      INVR %*% diag(1/sqrt(mu_c),n[i])
    U_i=t(C)%*%INVB%*%A
```

```
    # commands for generalized inverse - beta
    ai1=INVB
    mm1=C%*%sqe1
    ai1A=ai1%*%A
    ai1m1=ai1%*%mm1
    ai1A=INVBIG(ai1A,ai1m1,mm1,A,1,p)
    U_c=t(C)%*%ai1A

    Ustar_c=t(C)%*%INVB%*%C
    Ustar=Ustar+Ustar_c
    UUtran_c=tcrossprod(U_i)
    UUtran=UUtran+UUtran_c
    UUbc_c=tcrossprod(U_c)
    UUbc=UUbc+UUbc_c
    UUbc_ic=tcrossprod(U_c,U_i)
    UUbc2=UUbc2+UUbc_ic

    Ustar_c_array[,,i]=Ustar_c
    UUtran_c_array[,,i]=UUtran_c
  }

  # calculating adjustment factor for BC3
  for(i in 1:length(n)){
    Hi=diag(1/sqrt(1-pmin(0.75,c(diag(Ustar_c_array[,,i]%*%naive)))))
    UUbc3=UUbc3+Hi%*%UUtran_c_array[,,i]%*%Hi
  }

  # BC0 or usual Sandwich estimator of Prentice (1988);
  robust=naive%*%UUtran%*%naive

  # BC1 or Variance estimator due to Kauermann and Carroll (2001);
  varKC=naive%*%(UUbc2+t(UUbc2))%*%t(naive)/2

  # BC2 or Variance estimator due to Mancl and DeRouen (2001);
  varMD=naive%*%UUbc%*%t(naive)

  # BC3 or Variance estimator due to Fay and Graubard (2001);
  varFG=naive%*%UUbc3%*%t(naive)

  ROBFLAG=0
  if(min(diag(robust))<=0){ROBFLAG = 1}
  if(min(diag(varMD))<=0){ROBFLAG = 1}
```

```
    if(min(diag(varKC))<=0){ROBFLAG = 1}
    if(min(diag(varFG))<=0){ROBFLAG = 1}

    return(list(robust=robust,naive=naive,
                varMD=varMD,varKC=varKC,varFG=varFG,
                ROBFLAG=ROBFLAG))
  }

  # Performs modified Fisher scoring algorithm
  p=ncol(X)
  delta=rep(2*epsilon,p)
  deltaalp=2*epsilon
  converge=0
  beta=INITBETA(y,X)
  alpha=getalpha(y,X,beta,n)

  niter=1
  while((niter<=maxiter) & (max(abs(c(delta,deltaalp)))>epsilon) ){
    SCORE_RES=SCORE(beta, alpha, y, X, n, p)
    U=SCORE_RES$U
    UUtran=SCORE_RES$UUtran
    Ustar=SCORE_RES$Ustar

    psdustar=is_pos_def(Ustar)
    mineig=min(eigen(Ustar)$values)
    if(psdustar==TRUE){
      delta=solve(Ustar,U)
      beta=beta+delta
    } else{
      SINGFLAG=1
    }
    alphaold=alpha
    alpha=getalpha(y,X,beta,n)
    deltaalp=alpha-alphaold
    converge=(max(abs(c(delta,deltaalp)))<=epsilon)
    niter=niter+1
  }

  if(converge==1){
    MAKEVAR_RES=MAKEVAR(beta, alpha, y, X, n, p)
    robust=MAKEVAR_RES$robust
    naive=MAKEVAR_RES$naive
    varMD=MAKEVAR_RES$varMD
```

```
    varKC=MAKEVAR_RES$varKC
    varFG=MAKEVAR_RES$varFG
    ROBFLAG=MAKEVAR_RES$ROBFLAG
    return(list(beta=beta,alpha=alpha,naive=naive,
                robust=robust,varMD=varMD,
                varKC=varKC,varFG=varFG,niter=niter,
                converge=converge,
                ROBFLAG=ROBFLAG))
  } else{
    stop("Fisher scoring algorithm did not converge")
  }
}
```

An example usage of this function is

```
binMPreg(y=y, X=cbind(1,trt), id=id, n=nset, maxiter=50, epsilon=0.00001)
```

where `id` is the vector of sorted cluster identifiers and `nset` is the vector of cluster sizes (allowing for unequal cluster sizes).

B. Additional Web Tables Referenced in the Simulation Study

This Web Section includes 14 Web Tables summarizing additional simulation results. Specifically, the sets of Web Tables include

- Web Table 1 and 2 present the simulation results with $\bar{m} = 50$ and $P_0 = 0.30$, under the independence and exchangeable working correlation, respectively;
- Web Table 3 to 6 present the simulation results with $\bar{m} = 100$, under full combinations of the baseline prevalence $P_0 \in \{0.15, 0.30\}$ and under either the independence or exchangeable working correlation;
- Web Table 7 to 10 present the simulation results with $\bar{m} = 50$ and $P_0 \in \{0.15, 0.30\}$ and under either the independence or exchangeable working correlation, when the sample size estimate $\hat{n}$ is obtained from the "average cluster size method" ignoring cluster size variability;
- Web Table 11 to 14 present the simulation results with $\bar{m} = 100$ and $P_0 \in \{0.15, 0.30\}$ and under either the independence or exchangeable working correlation, when the sample size estimate $\hat{n}$ is obtained from the "average cluster size method" ignoring cluster size variability.

Web Table 1. Empirical power and type I error rates (in parentheses) for modified Poisson regression under independence working correlation when the mean cluster size is $\bar{m} = 50$, $P_0 = 0.30$. Empirical type I error rate between 3.6% and 6.4% and empirical power between 77.5% and 82.5% are in bold font and considered close to nominal according to the margin of error under a binomial model with 1000 replications.

ρ	CV	$\hat{n}$	Robust	MD	KC	FG	MD/KC	MD/FG	KC/FG
0.01	0	9	91.5(7.9)	76.2(3.0)	84.6(**5.3**)	**80.9**(**4.2**)	**80.6**(**4.0**)	**78.3**(**3.7**)	**82.5**(**4.6**)
	0.2	9	91.1(6.8)	76.6(3.1)	85.4(**4.6**)	**81.5**(3.3)	**80.4**(3.3)	**78.2**(3.3)	83.4(**4.2**)
	0.4	9	88.9(9.7)	71.4(3.2)	**81.0**(6.5)	77.3(**5.4**)	76.5(**5.0**)	74.2(**4.3**)	**79.0**(**6.0**)
	0.6	10	91.9(8.0)	75.4(2.4)	85.0(**4.7**)	**80.4**(3.3)	**80.2**(3.2)	**77.8**(2.7)	83.0(**4.1**)
	0.8	10	92.2(9.9)	72.0(3.2)	84.8(**5.9**)	78.2(**4.3**)	77.2(**4.2**)	74.4(**3.8**)	**80.7**(**4.9**)
0.05	0	17	86.4(6.7)	**78.3**(**5.0**)	83.0(**5.7**)	**81.4**(**5.3**)	**81.1**(**5.3**)	**80.2**(**5.0**)	**82.3**(**5.5**)
	0.2	17	85.7(**5.7**)	**77.5**(**3.7**)	**82.4**(**4.7**)	**80.6**(**4.2**)	**80.5**(**4.3**)	**79.1**(**3.8**)	**81.7**(**4.5**)
	0.4	18	84.8(6.9)	**77.5**(2.8)	**81.3**(**4.2**)	**79.3**(**3.7**)	**79.3**(**3.7**)	**78.6**(3.1)	**80.3**(**4.0**)
	0.6	20	86.3(8.2)	**78.2**(**4.3**)	**82.3**(**5.8**)	**80.7**(**4.8**)	**80.0**(**4.7**)	**79.1**(**4.5**)	**81.6**(**5.3**)
	0.8	23	88.3(8.7)	**78.5**(**4.7**)	84.6(6.9)	**82.2**(**5.8**)	**81.9**(**6.2**)	**80.6**(**5.3**)	83.5(6.5)
0.10	0	26	84.6(**5.1**)	**80.3**(3.5)	82.7(**3.9**)	**82.2**(**3.6**)	**82.0**(**3.6**)	**81.4**(3.5)	**82.4**(**3.7**)
	0.2	27	83.6(**6.1**)	**78.9**(**4.6**)	**81.2**(**5.1**)	**80.3**(**5.0**)	**80.4**(**5.0**)	**79.6**(**4.7**)	**80.9**(**5.1**)
	0.4	29	84.1(6.5)	**78.1**(**4.6**)	**81.0**(**5.7**)	**79.6**(**5.0**)	**79.7**(**5.0**)	**78.7**(**4.8**)	**80.5**(**5.2**)
	0.6	33	84.3(**6.1**)	**79.1**(**4.5**)	**81.9**(**5.1**)	**80.5**(**4.7**)	**80.8**(**4.7**)	**79.8**(**4.5**)	**81.7**(**4.8**)
	0.8	39	86.4(8.3)	**80.0**(**6.2**)	83.5(7.5)	**81.6**(6.5)	**81.8**(6.6)	**81.0**(6.5)	82.9(7.0)
0.15	0	36	**81.1**(**6.2**)	**78.8**(**4.2**)	**80.2**(**5.2**)	**79.8**(**4.5**)	**79.8**(**4.5**)	**79.4**(**4.3**)	**80.0**(**4.9**)
	0.2	37	84.3(**5.8**)	**80.4**(**4.6**)	82.6(**5.1**)	**81.7**(**5.0**)	**81.5**(**5.0**)	**80.9**(**5.0**)	**82.2**(**5.1**)
	0.4	41	86.5(**5.2**)	83.0(**4.0**)	84.8(**4.5**)	84.4(**4.4**)	84.4(**4.4**)	83.8(**4.1**)	84.5(**4.5**)
	0.6	47	84.5(**6.3**)	**80.2**(**4.2**)	**82.0**(**4.7**)	**81.0**(**4.4**)	**80.8**(**4.7**)	**80.5**(**4.3**)	**81.4**(**4.7**)
	0.8	55	84.8(**5.2**)	**81.6**(**4.1**)	83.5(**4.5**)	82.8(**4.4**)	**82.5**(**4.2**)	**82.2**(**4.1**)	83.2(**4.4**)
0.20	0	46	82.8(**4.8**)	**79.9**(**4.0**)	**81.9**(**4.4**)	**80.9**(**4.0**)	**80.9**(**4.0**)	**80.3**(**4.0**)	**81.5**(**4.1**)
	0.2	47	83.4(6.9)	**80.4**(**6.0**)	**82.0**(6.5)	**81.4**(**6.3**)	**81.3**(**6.3**)	**80.8**(**6.2**)	**81.7**(**6.3**)
	0.4	52	**82.0**(**5.2**)	**78.6**(**4.0**)	**79.8**(**4.7**)	**79.1**(**4.2**)	**79.1**(**4.2**)	**78.9**(**4.1**)	**79.3**(**4.5**)
	0.6	60	83.5(**5.0**)	**81.1**(**4.2**)	82.7(**4.5**)	**82.0**(**4.2**)	**81.8**(**4.2**)	**81.4**(**4.2**)	**82.3**(**4.4**)
	0.8	71	82.8(**5.5**)	**79.3**(**4.9**)	**80.8**(**5.2**)	**80.1**(**5.2**)	**79.8**(**5.2**)	**79.6**(**5.2**)	**80.4**(**5.2**)

Notes 1: ρ refers to ICC; CV refers to the coefficient of variation of cluster sizes; $\hat{n}$ refers to the estimated number of clusters. Notes 2: Robust refers to the t-test with the uncorrected robust sandwich variance estimator; MD refers to the t-test with the bias-corrected sandwich variance estimator due to Mancl and DeRouen; KC refers to the t-test with the bias-corrected sandwich variance estimator due to Kauermann and Carroll; FG refers to the t-test with the bias-corrected sandwich variance estimator due to Fay and Graubard; MD/KC refers to the t-test with the average MD/KC standard error estimator; MD/FG refers to the t-test with the average MD/FG standard error estimator; KC/FG refers to the t-test with the average KC/FG standard error estimator.

Web Table 2. Empirical power and type I error rates (in parentheses) for modified Poisson regression under exchangeable working correlation when the mean cluster size is $\bar{m} = 50$, $P_0 = 0.30$. Empirical type I error rate between 3.6% and 6.4% and empirical power between 77.5% and 82.5% are in bold font and considered close to nominal according to the margin of error under a binomial model with 1000 replications.

ρ	CV	$\hat{n}$	Robust	MD	KC	FG	MD/KC	MD/FG	KC/FG
0.01	0	9	91.5(7.9)	76.2(3.0)	84.6(**5.3**)	**80.9**(**4.2**)	**80.6**(**4.0**)	**78.3**(**3.7**)	**82.5**(**4.6**)
	0.2	9	90.9(6.8)	76.2(2.7)	85.2(**3.8**)	**81.3**(3.0)	**80.6**(3.2)	**78.4**(2.9)	83.2(3.5)
	0.4	9	88.5(10.3)	70.3(3.3)	**81.3**(6.7)	76.3(**5.3**)	75.8(**4.7**)	73.3(**3.9**)	**79.4**(**5.8**)
	0.6	9	86.8(9.4)	66.6(2.8)	**78.8**(**5.8**)	74.2(**4.0**)	72.6(3.5)	70.7(3.2)	76.2(**4.6**)
	0.8	10	90.9(11.7)	70.5(3.1)	84.9(7.4)	**78.3**(**5.1**)	**77.9**(**4.2**)	75.1(3.1)	**81.4**(**6.1**)
0.05	0	17	86.4(6.7)	**78.3**(**5.0**)	83.0(**5.7**)	**81.4**(**5.3**)	**81.1**(**5.3**)	**80.2**(**5.0**)	**82.3**(**5.5**)
	0.2	17	86.4(6.8)	**78.9**(**4.8**)	83.1(**5.8**)	**81.3**(**5.3**)	**80.7**(**5.2**)	**80.4**(**4.8**)	**82.2**(**5.6**)
	0.4	17	85.0(**5.9**)	**77.9**(2.9)	**81.8**(**4.4**)	**80.0**(3.2)	**80.0**(3.4)	**78.9**(3.1)	**81.0**(**3.8**)
	0.6	18	87.7(7.7)	81.6(**5.3**)	84.5(**6.2**)	82.9(**5.9**)	82.9(**5.8**)	**82.1**(**5.6**)	83.5(**6.1**)
	0.8	19	86.1(8.6)	**78.9**(**4.6**)	83.2(6.6)	**80.8**(**5.5**)	**81.1**(**5.5**)	**80.4**(**5.3**)	**81.8**(**6.1**)
0.10	0	26	84.6(**5.1**)	**80.3**(3.5)	82.7(**3.9**)	**82.2**(**3.6**)	**82.0**(**3.6**)	**81.4**(3.5)	**82.4**(**3.7**)
	0.2	26	85.0(**5.8**)	**80.0**(3.5)	82.8(**4.7**)	**81.4**(**4.1**)	**81.4**(**4.1**)	**80.9**(**4.0**)	**82.1**(**4.4**)
	0.4	27	84.4(6.9)	**81.2**(**5.3**)	83.5(**6.2**)	82.9(**5.7**)	82.9(**5.6**)	**82.2**(**5.5**)	83.1(**6.0**)
	0.6	27	83.7(7.4)	77.4(**4.6**)	**80.9**(**5.7**)	**79.0**(**5.3**)	**79.0**(**5.1**)	**78.2**(**4.9**)	**80.1**(**5.5**)
	0.8	28	83.8(**5.4**)	**79.2**(**4.1**)	**81.9**(**4.8**)	**80.7**(**4.4**)	**80.8**(**4.4**)	**80.3**(**4.2**)	**81.2**(**4.7**)
0.15	0	36	**81.1**(**6.2**)	**78.8**(**4.2**)	**80.2**(**5.2**)	**79.8**(**4.5**)	**79.8**(**4.5**)	**79.4**(**4.3**)	**80.0**(**4.9**)
	0.2	36	83.0(**6.0**)	**78.8**(**4.3**)	**80.9**(**5.5**)	**80.0**(**4.8**)	**80.0**(**4.8**)	**79.3**(**4.6**)	**80.4**(**5.3**)
	0.4	37	84.9(**5.0**)	**82.1**(**3.8**)	83.7(**4.4**)	82.9(**4.1**)	82.8(**4.0**)	**82.5**(**3.9**)	83.2(**4.1**)
	0.6	37	**79.4**(7.2)	76.8(**5.1**)	**78.2**(**6.3**)	**77.7**(**5.5**)	**77.7**(**5.5**)	77.1(**5.2**)	**78.0**(**6.0**)
	0.8	38	**81.1**(**6.0**)	77.4(**5.1**)	**80.0**(**5.5**)	**78.9**(**5.3**)	**79.0**(**5.3**)	**77.6**(**5.1**)	**79.3**(**5.3**)
0.20	0	46	82.8(**4.8**)	**79.9**(**4.0**)	**81.9**(**4.4**)	**80.9**(**4.0**)	**80.9**(**4.0**)	**80.3**(**4.0**)	**81.5**(**4.1**)
	0.2	46	**80.2**(**4.1**)	76.7(**3.9**)	**77.9**(**4.0**)	**77.5**(**4.0**)	**77.6**(**4.0**)	77.0(**4.0**)	**77.8**(**4.0**)
	0.4	46	**81.6**(**5.4**)	**78.3**(**4.9**)	**80.5**(**5.0**)	**79.6**(**5.0**)	**79.6**(**5.0**)	**79.1**(**5.0**)	**80.1**(**5.0**)
	0.6	47	**82.1**(**5.5**)	**79.4**(**4.1**)	**80.8**(**4.7**)	**80.0**(**4.3**)	**79.9**(**4.3**)	**79.7**(**4.3**)	**80.5**(**4.3**)
	0.8	48	**82.2**(**5.9**)	**79.7**(**5.1**)	**80.6**(**5.8**)	**80.2**(**5.8**)	**80.2**(**5.7**)	**80.1**(**5.7**)	**80.6**(**5.8**)

Notes 1: ρ refers to ICC; CV refers to the coefficient of variation of cluster sizes; $\hat{n}$ refers to the estimated number of clusters.
Notes 2: Robust refers to the t-test with the uncorrected robust sandwich variance estimator; MD refers to the t-test with the bias-corrected sandwich variance estimator due to Mancl and DeRouen; KC refers to the t-test with the bias-corrected sandwich variance estimator due to Kauermann and Carroll; FG refers to the t-test with the bias-corrected sandwich variance estimator due to Fay and Graubard; MD/KC refers to the t-test with the average MD/KC standard error estimator; MD/FG refers to the t-test with the average MD/FG standard error estimator; KC/FG refers to the t-test with the average KC/FG standard error estimator.

Web Table 3. Empirical power and type I error rates (in parentheses) for modified Poisson regression under independence working correlation when the mean cluster size is $\bar{m} = 100$, $P_0 = 0.15$. Empirical type I error rate between 3.6% and 6.4% and empirical power between 77.5% and 82.5% are in bold font and considered close to nominal according to the margin of error under a binomial model with 1000 replications.

ρ	CV	$\hat{n}$	Robust	MD	KC	FG	MD/KC	MD/FG	KC/FG
0.01	0	8	89.6(8.5)	73.9(3.5)	83.5(**5.8**)	**78.0**(**4.9**)	**78.7**(**5.1**)	75.8(**4.6**)	**80.9**(**5.4**)
	0.2	8	89.3(7.5)	71.9(2.7)	**82.0**(**4.9**)	77.4(**3.9**)	77.3(**3.8**)	74.4(3.0)	**80.2**(**4.3**)
	0.4	9	91.5(7.5)	75.9(3.0)	85.4(**5.0**)	**81.5**(**4.4**)	**81.0**(**3.7**)	**78.5**(**3.6**)	83.8(**4.6**)
	0.6	9	93.4(10.1)	74.2(**3.9**)	85.8(6.7)	**79.4**(**5.2**)	**79.6**(**5.3**)	76.9(**4.5**)	**82.5**(**6.0**)
	0.8	10	93.9(12.8)	75.5(**4.2**)	**85.9**(7.0)	**80.5**(**5.3**)	**81.3**(**5.1**)	**77.5**(**4.7**)	84.0(6.1)
0.05	0	18	84.6(7.3)	**77.5**(**4.2**)	**81.3**(**5.4**)	**79.8**(**4.8**)	**79.5**(**4.9**)	**78.6**(**4.5**)	**80.4**(**5.1**)
	0.2	19	84.6(**5.5**)	**78.9**(3.5)	**81.5**(**4.1**)	**80.6**(**3.7**)	**80.5**(**3.8**)	**79.8**(**3.7**)	**80.8**(**3.8**)
	0.4	20	86.0(8.2)	**79.6**(**5.6**)	84.0(7.3)	82.6(6.7)	82.6(6.5)	**81.1**(**6.1**)	83.5(6.8)
	0.6	23	88.0(7.0)	**80.8**(3.4)	84.1(**5.3**)	**81.9**(**4.3**)	**82.1**(**4.4**)	**81.4**(**3.8**)	83.2(**4.7**)
	0.8	26	88.0(9.6)	**81.3**(7.1)	85.3(7.9)	83.7(7.6)	84.1(7.4)	82.7(7.3)	85.1(7.7)
0.10	0	31	**82.4**(7.6)	**79.3**(**5.0**)	**80.5**(**5.8**)	**79.8**(**5.5**)	**79.5**(**5.3**)	**78.6**(**5.2**)	**80.4**(**5.5**)
	0.2	32	84.5(**5.8**)	**80.5**(**4.3**)	82.7(**5.1**)	**81.6**(**4.6**)	**81.6**(**4.8**)	**81.1**(**4.4**)	**82.4**(**4.9**)
	0.4	35	83.9(7.5)	**80.3**(**5.2**)	**82.3**(**6.2**)	**81.1**(**5.7**)	**80.9**(**5.8**)	**80.5**(**5.5**)	**81.8**(**5.9**)
	0.6	40	83.5(6.6)	**78.7**(**5.3**)	**80.8**(**5.8**)	**79.6**(**5.4**)	**79.6**(**5.4**)	**79.3**(**5.3**)	**80.4**(**5.5**)
	0.8	48	85.1(7.2)	**81.2**(**5.0**)	83.9(**5.9**)	**82.5**(**5.3**)	**82.2**(**5.2**)	**81.6**(**5.1**)	83.2(**5.7**)
0.15	0	44	82.7(7.2)	**80.6**(**5.5**)	**81.4**(6.6)	**81.1**(**5.8**)	**81.1**(**5.9**)	**80.9**(**5.6**)	**81.1**(6.5)
	0.2	46	**82.1**(**7.1**)	**79.0**(**5.3**)	**80.9**(**6.2**)	**79.9**(**5.7**)	**79.8**(**5.8**)	**79.4**(**5.5**)	**80.7**(**5.9**)
	0.4	50	83.5(7.7)	**80.8**(6.9)	**82.5**(7.4)	**81.5**(7.0)	**81.8**(7.1)	**81.2**(6.9)	**81.9**(7.2)
	0.6	58	**82.2**(6.9)	**79.8**(**5.7**)	**81.0**(**5.9**)	**80.4**(**5.8**)	**80.4**(**5.8**)	**80.1**(**5.8**)	**80.6**(**5.9**)
	0.8	69	85.5(**6.3**)	**81.8**(**4.9**)	83.7(**5.7**)	**82.3**(**5.3**)	82.9(**5.3**)	**82.2**(**5.2**)	83.1(**5.5**)
0.20	0	57	**82.5**(**6.0**)	**80.3**(**5.2**)	**81.8**(**5.4**)	**80.9**(**5.4**)	**80.9**(**5.3**)	**80.8**(**5.3**)	**81.2**(**5.4**)
	0.2	59	84.5(**5.8**)	**82.3**(**4.4**)	83.6(**5.1**)	82.9(**4.8**)	83.1(**4.7**)	**82.4**(**4.5**)	83.2(**4.8**)
	0.4	65	84.5(7.1)	82.9(**6.0**)	83.8(6.5)	83.1(**6.1**)	82.9(**6.1**)	82.9(**6.1**)	83.2(**6.3**)
	0.6	76	85.5(**5.8**)	83.7(**4.5**)	84.6(**5.1**)	84.2(**4.7**)	84.0(**4.8**)	84.0(**4.6**)	84.4(**5.0**)
	0.8	90	85.2(8.2)	**82.4**(7.0)	83.7(7.6)	83.0(7.1)	83.2(7.2)	82.9(7.1)	83.1(7.4)

Notes 1: ρ refers to ICC; CV refers to the coefficient of variation of cluster sizes; $\hat{n}$ refers to the estimated number of clusters.
Notes 2: Robust refers to the t-test with the uncorrected robust sandwich variance estimator; MD refers to the t-test with the bias-corrected sandwich variance estimator due to Mancl and DeRouen; KC refers to the t-test with the bias-corrected sandwich variance estimator due to Kauermann and Carroll; FG refers to the t-test with the bias-corrected sandwich variance estimator due to Fay and Graubard; MD/KC refers to the t-test with the average MD/KC standard error estimator; MD/FG refers to the t-test with the average MD/FG standard error estimator; KC/FG refers to the t-test with the average KC/FG standard error estimator.

Web Table 4. Empirical power and type I error rates (in parentheses) for modified Poisson regression under exchangeable working correlation when the mean cluster size is $\bar{m} = 100$, $P_0 = 0.15$. Empirical type I error rate between 3.6% and 6.4% and empirical power between 77.5% and 82.5% are in bold font and considered close to nominal according to the margin of error under a binomial model with 1000 replications.

ρ	CV	$\hat{n}$	Robust	MD	KC	FG	MD/KC	MD/FG	KC/FG
0.01	0	8	89.6(8.5)	73.9(3.5)	83.5(**5.8**)	**78.0**(**4.9**)	**78.7**(**5.1**)	75.8(**4.6**)	**80.9**(**5.4**)
	0.2	8	89.1(7.6)	72.8(2.1)	82.8(**4.5**)	**78.1**(3.3)	77.3(3.3)	74.3(3.0)	**80.0**(**4.0**)
	0.4	8	88.8(8.7)	70.2(2.6)	**81.9**(**4.6**)	76.0(**3.7**)	74.0(3.1)	72.0(2.8)	**79.0**(**4.0**)
	0.6	9	91.5(11.5)	76.2(3.1)	86.3(7.1)	**82.2**(**4.8**)	**81.4**(**4.4**)	**79.4**(**3.6**)	84.5(**6.0**)
	0.8	9	90.5(13.7)	71.6(3.1)	83.4(8.2)	**78.2**(**5.4**)	**78.3**(**4.2**)	75.8(**4.0**)	**81.4**(6.5)
0.05	0	18	84.6(7.3)	**77.5**(**4.2**)	**81.3**(**5.4**)	**79.8**(**4.8**)	**79.5**(**4.9**)	**78.6**(**4.5**)	**80.4**(**5.1**)
	0.2	18	86.1(**5.2**)	**80.3**(2.6)	83.6(**4.2**)	**82.2**(3.3)	**82.2**(3.3)	**81.3**(2.9)	83.0(**3.7**)
	0.4	19	86.5(**6.4**)	**80.5**(**4.2**)	84.2(**5.1**)	82.9(**4.8**)	82.7(**4.7**)	**81.4**(**4.6**)	83.8(**4.9**)
	0.6	19	85.0(**6.3**)	**78.8**(**3.6**)	**82.5**(**4.9**)	**80.9**(**4.4**)	**80.7**(**4.4**)	**79.9**(**4.2**)	**81.9**(**4.6**)
	0.8	20	84.6(8.7)	**78.4**(**4.9**)	**81.4**(**7.1**)	**80.0**(**5.7**)	**79.8**(**5.7**)	**79.3**(**5.2**)	**80.8**(6.6)
0.10	0	31	**82.4**(7.6)	**79.3**(**5.0**)	**80.5**(**5.8**)	**79.8**(**5.5**)	**79.8**(**5.2**)	**79.5**(**5.3**)	**80.2**(**5.5**)
	0.2	31	84.6(**5.8**)	**80.3**(**4.2**)	**82.4**(**4.8**)	**81.3**(**4.5**)	**81.4**(**4.5**)	**80.8**(**4.4**)	**82.0**(**4.7**)
	0.4	31	84.6(**6.1**)	**79.9**(**4.8**)	82.8(**5.4**)	**81.1**(**5.1**)	**81.5**(**5.2**)	**80.4**(**4.8**)	**81.8**(**5.4**)
	0.6	32	83.3(**6.0**)	**81.3**(**4.8**)	**82.3**(**5.2**)	**81.7**(**5.0**)	**81.7**(**5.1**)	**81.5**(**4.8**)	**82.0**(**5.2**)
	0.8	32	**81.6**(**6.4**)	**77.8**(**4.8**)	**79.9**(**5.8**)	**78.8**(**5.0**)	**78.9**(**5.1**)	**78.1**(**4.8**)	**79.3**(**5.5**)
0.15	0	44	82.7(7.2)	**80.6**(**5.5**)	**81.4**(6.6)	**81.1**(**5.8**)	**81.1**(**5.9**)	**80.9**(**5.6**)	**81.1**(6.5)
	0.2	44	82.9(**5.8**)	**80.6**(**4.5**)	**81.8**(**5.2**)	**81.0**(**5.0**)	**80.9**(**5.0**)	**80.7**(**4.8**)	**81.5**(**5.2**)
	0.4	44	82.7(**5.9**)	**80.4**(**4.8**)	**81.8**(**5.4**)	**81.1**(**5.1**)	**80.9**(**5.2**)	**80.6**(**4.8**)	**81.3**(**5.4**)
	0.6	45	**82.3**(**5.8**)	**80.4**(**4.4**)	**81.6**(**5.0**)	**80.8**(**4.6**)	**80.7**(**4.5**)	**80.4**(**4.5**)	**80.9**(**4.8**)
	0.8	45	**81.5**(**6.2**)	**79.0**(**4.6**)	**80.3**(**5.4**)	**79.5**(**5.0**)	**79.5**(**5.1**)	**79.4**(**5.0**)	**79.7**(**5.2**)
0.20	0	57	**82.5**(**6.0**)	**80.3**(**5.2**)	**81.8**(**5.4**)	**80.9**(**5.4**)	**80.9**(**5.3**)	**80.8**(**5.3**)	**81.2**(**5.4**)
	0.2	57	83.7(**5.2**)	**81.9**(**4.0**)	82.9(**4.7**)	**82.4**(**4.4**)	**82.5**(**4.5**)	**82.0**(**4.1**)	82.6(**4.5**)
	0.4	57	82.8(**5.1**)	**81.2**(**4.4**)	**82.1**(**4.6**)	**81.3**(**4.5**)	**81.4**(**4.6**)	**80.9**(**4.5**)	**81.9**(**4.6**)
	0.6	58	82.6(6.8)	**81.0**(**6.4**)	**81.8**(6.6)	**81.5**(6.5)	**81.3**(6.5)	**81.1**(**6.4**)	**81.6**(6.6)
	0.8	58	84.0(6.8)	**81.4**(**5.9**)	83.0(**6.4**)	82.6(**6.0**)	**82.4**(**6.0**)	**81.8**(**6.0**)	82.7(**6.1**)

Notes 1: ρ refers to ICC; CV refers to the coefficient of variation of cluster sizes; $\hat{n}$ refers to the estimated number of clusters. Notes 2: Robust refers to the t-test with the uncorrected robust sandwich variance estimator; MD refers to the t-test with the bias-corrected sandwich variance estimator due to Mancl and DeRouen; KC refers to the t-test with the bias-corrected sandwich variance estimator due to Kauermann and Carroll; FG refers to the t-test with the bias-corrected sandwich variance estimator due to Fay and Graubard; MD/KC refers to the t-test with the average MD/KC standard error estimator; MD/FG refers to the t-test with the average MD/FG standard error estimator; KC/FG refers to the t-test with the average KC/FG standard error estimator.

Web Table 5. Empirical power and type I error rates (in parentheses) for modified Poisson regression under independence working correlation when the mean cluster size is $\bar{m} = 100$, $P_0 = 0.30$. Empirical type I error rate between 3.6% and 6.4% and empirical power between 77.5% and 82.5% are in bold font and considered close to nominal according to the margin of error under a binomial model with 1000 replications.

ρ	CV	$\hat{n}$	Robust	MD	KC	FG	MD/KC	MD/FG	KC/FG
0.01	0	7	91.4(8.4)	73.7(3.1)	84.6(**5.1**)	**80.4**(**3.9**)	**79.5**(**4.5**)	76.6(**3.7**)	**82.5**(**4.8**)
	0.2	7	91.2(9.1)	70.5(2.9)	83.1(**5.6**)	**78.1**(**4.2**)	77.4(**4.1**)	74.3(3.3)	**80.8**(**4.9**)
	0.4	7	92.2(10.1)	69.5(2.7)	**81.4**(**5.8**)	76.3(**4.5**)	75.5(**3.9**)	73.6(3.4)	**78.2**(**5.2**)
	0.6	8	94.2(11.2)	74.9(**4.8**)	87.4(7.0)	82.7(**6.0**)	**80.7**(**5.9**)	**78.5**(**5.2**)	85.1(6.8)
	0.8	8	92.7(13.4)	69.8(**4.8**)	84.3(8.1)	**77.5**(**6.5**)	76.1(**5.7**)	72.8(**5.5**)	**80.4**(6.7)
0.05	0	15	85.9(**5.0**)	**78.3**(2.6)	**82.4**(3.3)	**80.3**(2.9)	**79.6**(2.9)	**79.2**(2.7)	**81.1**(3.3)
	0.2	15	87.0(7.4)	**78.8**(**4.2**)	82.8(**5.7**)	**81.2**(**5.1**)	**80.8**(**5.0**)	**79.9**(**4.8**)	**82.2**(**5.4**)
	0.4	16	87.5(7.1)	**79.6**(**3.7**)	84.0(**5.6**)	82.6(**4.6**)	82.6(**4.5**)	**81.1**(**4.1**)	83.5(**5.0**)
	0.6	18	88.3(9.8)	**78.4**(**4.7**)	84.0(6.9)	**81.4**(**5.7**)	**81.0**(**5.4**)	**80.0**(**5.1**)	83.3(**6.0**)
	0.8	21	89.5(8.5)	**81.8**(**4.7**)	86.4(6.5)	84.7(**5.4**)	84.2(**5.4**)	82.6(**5.0**)	85.8(**5.9**)
0.10	0	24	83.1(**5.3**)	**77.8**(3.4)	**80.9**(**4.9**)	**79.1**(**4.4**)	**79.3**(**4.0**)	**78.1**(**3.6**)	**80.1**(**4.6**)
	0.2	25	84.5(**5.3**)	**79.5**(3.5)	**81.4**(**4.2**)	**80.5**(**4.1**)	**80.6**(**3.8**)	**79.8**(**3.6**)	**81.0**(**4.1**)
	0.4	28	86.8(**5.9**)	**80.7**(**4.9**)	84.0(**5.3**)	**82.5**(**5.0**)	**82.0**(**5.0**)	**81.2**(**4.9**)	83.1(**5.1**)
	0.6	32	83.2(7.4)	**78.0**(**4.7**)	**80.9**(**5.7**)	**79.3**(**5.2**)	**79.1**(**5.1**)	**78.5**(**4.8**)	**80.1**(**5.3**)
	0.8	37	87.3(8.1)	**81.5**(**5.1**)	84.7(**6.4**)	83.1(**5.7**)	83.3(**5.8**)	82.6(**5.5**)	83.6(**6.1**)
0.15	0	34	83.3(**5.0**)	**79.3**(**4.0**)	**81.5**(**4.6**)	**80.1**(**4.3**)	**80.2**(**4.2**)	**79.8**(**4.1**)	**80.6**(**4.5**)
	0.2	36	84.4(6.5)	**81.1**(**4.9**)	83.1(**5.7**)	**82.4**(**5.2**)	**82.3**(**5.2**)	**81.6**(**5.2**)	82.9(**5.5**)
	0.4	39	**82.0**(**6.0**)	**77.7**(**4.2**)	**80.0**(**5.1**)	**78.8**(**4.6**)	**78.8**(**4.4**)	**78.5**(**4.5**)	**79.5**(**4.7**)
	0.6	45	83.0(6.9)	**78.9**(**5.1**)	**81.1**(**6.3**)	**80.3**(**5.9**)	**80.0**(**5.9**)	**79.7**(**5.6**)	**80.8**(**6.1**)
	0.8	54	84.3(7.5)	**79.9**(**4.8**)	**82.1**(**5.8**)	**81.1**(**5.2**)	**81.3**(**5.4**)	**80.9**(**4.9**)	**81.6**(**5.6**)
0.20	0	44	**81.8**(**5.4**)	**79.5**(**4.8**)	**80.9**(**4.9**)	**80.4**(**4.8**)	**80.4**(**4.8**)	**80.2**(**4.8**)	**80.6**(**4.9**)
	0.2	46	83.6(**6.1**)	**80.7**(**5.4**)	**81.9**(**5.8**)	**81.3**(**5.6**)	**81.3**(**5.6**)	**81.1**(**5.5**)	**81.5**(**5.7**)
	0.4	51	84.4(**5.2**)	**81.9**(**4.6**)	82.7(**4.9**)	**82.3**(**4.7**)	**82.2**(**4.9**)	**82.1**(**4.6**)	**82.5**(**4.9**)
	0.6	59	**81.8**(6.7)	77.2(**5.9**)	**79.1**(**6.1**)	**78.2**(**6.0**)	**78.1**(**5.9**)	**77.8**(**5.9**)	**78.5**(**6.0**)
	0.8	70	85.2(**6.4**)	**81.6**(**5.2**)	83.4(**5.9**)	82.8(**5.6**)	82.7(**5.5**)	**82.3**(**5.3**)	83.3(**5.6**)

Notes 1: ρ refers to ICC; CV refers to the coefficient of variation of cluster sizes; $\hat{n}$ refers to the estimated number of clusters. Notes 2: Robust refers to the t-test with the uncorrected robust sandwich variance estimator; MD refers to the t-test with the bias-corrected sandwich variance estimator due to Mancl and DeRouen; KC refers to the t-test with the bias-corrected sandwich variance estimator due to Kauermann and Carroll; FG refers to the t-test with the bias-corrected sandwich variance estimator due to Fay and Graubard; MD/KC refers to the t-test with the average MD/KC standard error estimator; MD/FG refers to the t-test with the average MD/FG standard error estimator; KC/FG refers to the t-test with the average KC/FG standard error estimator.

Web Table 6. Empirical power and type I error rates (in parentheses) for modified Poisson regression under exchangeable working correlation when the mean cluster size is $\bar{m} = 100$, $P_0 = 0.30$. Empirical type I error rate between 3.6% and 6.4% and empirical power between 77.5% and 82.5% are in bold font and considered close to nominal according to the margin of error under a binomial model with 1000 replications.

ρ	CV	$\hat{n}$	Robust	MD	KC	FG	MD/KC	MD/FG	KC/FG
0.01	0	7	91.4(8.4)	73.7(3.1)	84.6(**5.1**)	**80.4**(**3.9**)	**79.5**(**4.5**)	76.6(**3.7**)	**82.5**(**4.8**)
	0.2	7	90.9(8.9)	69.9(2.9)	83.4(**5.5**)	**78.2**(**4.3**)	76.6(**4.4**)	73.6(3.3)	**81.0**(**4.7**)
	0.4	7	87.6(9.8)	62.8(1.9)	**78.1**(**5.0**)	71.4(**3.6**)	71.1(3.4)	67.3(2.6)	75.9(**4.3**)
	0.6	7	90.7(10.8)	61.2(1.8)	**81.4**(**5.0**)	72.8(3.4)	72.1(2.8)	67.3(2.3)	77.0(**4.0**)
	0.8	8	94.5(10.5)	73.4(2.3)	87.3(**5.9**)	**81.6**(**4.6**)	**80.2**(3.3)	77.0(2.8)	84.8(**5.1**)
0.05	0	15	85.9(**5.0**)	**78.3**(2.6)	**82.4**(3.3)	**80.3**(2.9)	**79.6**(2.9)	**79.2**(2.7)	**81.1**(3.3)
	0.2	15	87.0(8.0)	**79.9**(**4.1**)	83.6(**5.7**)	**82.0**(**5.0**)	**81.8**(**4.9**)	**81.1**(**4.4**)	82.8(**5.6**)
	0.4	15	87.9(6.9)	**79.0**(**4.4**)	84.4(**5.3**)	**81.9**(**5.0**)	**81.4**(**4.7**)	**80.2**(**4.5**)	82.8(**5.2**)
	0.6	15	85.9(7.8)	**77.6**(**4.3**)	82.6(**6.1**)	**79.5**(**5.0**)	**79.7**(**5.0**)	**78.7**(**4.8**)	**81.0**(**5.6**)
	0.8	16	86.6(8.3)	**79.4**(**5.4**)	82.7(6.8)	**81.6**(**6.0**)	**81.1**(**6.9**)	**80.3**(**5.6**)	**82.2**(**6.2**)
0.10	0	24	83.1(**5.3**)	**77.8**(3.4)	**80.9**(**4.9**)	**79.1**(**4.4**)	**79.3**(**4.0**)	**78.1**(**3.6**)	**80.1**(**4.6**)
	0.2	25	84.4(**5.4**)	**80.3**(**4.7**)	82.7(**5.1**)	**82.0**(**4.8**)	**81.7**(**4.9**)	**80.8**(**4.8**)	**82.4**(**5.0**)
	0.4	25	83.9(**5.5**)	**79.7**(**3.4**)	**82.5**(**4.1**)	**81.5**(3.5)	**81.2**(3.5)	**80.5**(3.5)	**82.0**(**3.9**)
	0.6	25	**82.1**(6.5)	**78.4**(**4.9**)	**80.3**(**5.3**)	**79.0**(**5.2**)	**78.6**(**5.1**)	**78.5**(**5.0**)	**79.5**(**5.3**)
	0.8	26	84.4(**6.0**)	**79.6**(3.5)	82.7(**5.0**)	**81.1**(**4.3**)	**80.9**(**4.2**)	**80.3**(**4.0**)	**81.9**(**4.5**)
0.15	0	34	83.3(**5.0**)	**79.3**(**4.0**)	**81.5**(**4.6**)	**80.1**(**4.3**)	**80.2**(**4.2**)	**79.8**(**4.1**)	**80.6**(**4.5**)
	0.2	34	83.3(**4.5**)	**80.2**(3.5)	**81.8**(**3.8**)	**80.9**(**3.7**)	**80.8**(**3.7**)	**80.6**(3.5)	**81.3**(**3.7**)
	0.4	35	84.6(**6.0**)	**81.4**(**4.6**)	82.8(**5.0**)	**82.2**(**4.6**)	**82.1**(**4.6**)	**81.9**(**4.6**)	82.7(**4.7**)
	0.6	35	83.8(**5.5**)	**81.3**(**4.2**)	**82.5**(**4.8**)	**81.7**(**4.4**)	**81.7**(**4.4**)	**81.5**(**4.3**)	**82.3**(**4.6**)
	0.8	35	82.6(**5.7**)	**78.3**(**4.9**)	**80.3**(**5.5**)	**79.1**(**5.1**)	**79.3**(**5.2**)	**78.9**(**5.0**)	**79.6**(**5.4**)
0.20	0	44	**81.8**(**5.4**)	**79.5**(**4.8**)	**80.9**(**4.9**)	**80.4**(**4.8**)	**80.4**(**4.8**)	**80.2**(**4.8**)	**80.6**(**4.9**)
	0.2	44	**81.7**(**4.9**)	**78.5**(**3.9**)	**80.3**(**4.2**)	**79.5**(**4.1**)	**79.5**(**4.1**)	**78.9**(**4.0**)	**79.6**(**4.1**)
	0.4	44	82.6(**5.6**)	**81.0**(**4.2**)	**81.8**(**5.0**)	**81.5**(**4.3**)	**81.3**(**4.2**)	**81.1**(**4.2**)	**81.6**(**4.7**)
	0.6	45	85.3(**5.6**)	82.7(**4.9**)	84.2(**5.3**)	83.2(**5.1**)	83.3(**5.0**)	82.9(**4.9**)	83.6(**5.2**)
	0.8	45	83.8(**5.6**)	**80.9**(**4.3**)	**82.2**(**4.5**)	**81.5**(**4.5**)	**81.5**(**4.5**)	**81.3**(**4.4**)	**81.6**(**4.5**)

Notes 1: ρ refers to ICC; CV refers to the coefficient of variation of cluster sizes; $\hat{n}$ refers to the estimated number of clusters. Notes 2: Robust refers to the t-test with the uncorrected robust sandwich variance estimator; MD refers to the t-test with the bias-corrected sandwich variance estimator due to Mancl and DeRouen; KC refers to the t-test with the bias-corrected sandwich variance estimator due to Kauermann and Carroll; FG refers to the t-test with the bias-corrected sandwich variance estimator due to Fay and Graubard; MD/KC refers to the t-test with the average MD/KC standard error estimator; MD/FG refers to the t-test with the average MD/FG standard error estimator; KC/FG refers to the t-test with the average KC/FG standard error estimator.

Web Table 7. Empirical power and type I error rates (in parentheses) for modified Poisson regression under independence working correlation when the mean cluster size is $\bar{m} = 50$, $P_0 = 0.15$. The required sample size estimate $\hat{n}$ is obtained by "average cluster size method" and ignores variable cluster sizes. Empirical type I error rate between 3.6% and 6.4% and empirical power between 77.5% and 82.5% are in bold font and considered close to nominal according to the margin of error under a binomial model with 1000 replications.

ρ	CV	$\hat{n}$	Robust	MD	KC	FG	MD/KC	MD/FG	KC/FG
	0	11	91.0(7.1)	**79.7**(3.4)	86.5(**5.1**)	83.3(**4.0**)	83.1(**4.3**)	**82.2**(**3.7**)	84.5(**4.6**)
	0.2	11	89.1(7.6)	77.2(3.2)	83.6(**4.8**)	**80.9**(**3.6**)	**81.0**(**4.0**)	79.3(3.3)	**82.3**(**4.2**)
0.01	0.4	11	88.5(8.4)	75.6(**3.7**)	**82.0**(**5.5**)	**78.9**(**4.5**)	**78.3**(**4.3**)	**77.6**(**4.0**)	**80.1**(**5.1**)
	0.6	11	88.5(8.9)	70.0(3.4)	**81.2**(**5.6**)	75.4(**4.4**)	75.1(**4.5**)	72.3(**3.9**)	**77.9**(**5.0**)
	0.8	11	87.0(10.7)	68.5(3.5)	77.9(7.2)	73.5(**5.7**)	73.8(**5.4**)	71.4(**4.7**)	76.1(**6.1**)
	0	21	86.9(**5.0**)	**82.1**(3.3)	84.6(**3.9**)	83.5(**3.7**)	83.6(**3.8**)	82.7(3.4)	84.1(**3.8**)
	0.2	21	84.8(**5.9**)	77.3(**3.8**)	**81.5**(**4.8**)	**79.1**(**4.1**)	**79.5**(**4.2**)	**78.2**(**3.9**)	**80.6**(**4.3**)
0.05	0.4	21	83.2(6.8)	74.2(**3.9**)	**78.9**(**5.1**)	76.8(**4.4**)	77.0(**4.3**)	76.0(**4.1**)	**78.1**(**4.6**)
	0.6	21	76.9(8.8)	69.3(**5.6**)	73.9(6.9)	71.2(**6.0**)	71.5(**6.1**)	70.1(**5.8**)	72.6(**6.3**)
	0.8	21	**78.5**(10.3)	66.7(**5.8**)	72.1(8.5)	68.9(6.8)	69.6(6.6)	67.9(**6.2**)	70.5(7.3)
	0	33	84.3(**6.3**)	**81.7**(**5.4**)	82.9(**5.8**)	**82.3**(**5.6**)	**82.1**(**5.5**)	**81.8**(**5.4**)	**82.5**(**5.6**)
	0.2	33	**81.3**(6.6)	**77.5**(**4.5**)	**79.5**(**5.6**)	**78.7**(**4.8**)	**78.7**(**4.9**)	**78.3**(**4.8**)	**79.0**(**5.3**)
0.10	0.4	33	**79.7**(6.9)	73.7(**4.6**)	76.2(**5.7**)	74.5(**5.4**)	74.8(**5.5**)	73.9(**5.1**)	75.2(**5.6**)
	0.6	33	74.7(6.5)	69.2(**4.1**)	72.6(**5.0**)	70.6(**4.7**)	71.0(**4.7**)	70.0(**4.5**)	71.2(**4.8**)
	0.8	33	70.0(9.3)	59.9(**5.6**)	64.8(7.9)	61.9(**6.4**)	62.8(**6.4**)	60.7(**5.9**)	63.4(6.8)
	0	46	**82.5**(**4.8**)	**79.8**(**3.7**)	**80.8**(**4.2**)	**80.2**(**3.8**)	**80.4**(**3.9**)	**80.1**(**3.7**)	**80.5**(**3.9**)
	0.2	46	**79.1**(**5.6**)	77.2(**5.1**)	**78.4**(**5.3**)	**77.8**(**5.1**)	**77.9**(**5.1**)	**77.5**(**5.1**)	**78.2**(**5.1**)
0.15	0.4	46	77.4(8.0)	73.5(6.7)	75.6(7.1)	74.1(6.7)	74.4(6.9)	74.1(6.7)	74.9(7.0)
	0.6	46	75.9(6.6)	71.2(**5.2**)	73.9(**6.0**)	72.2(**5.3**)	72.5(**5.6**)	71.9(**5.2**)	73.0(**5.6**)
	0.8	46	68.9(8.4)	62.8(**6.1**)	66.1(7.1)	64.7(6.5)	64.6(6.6)	63.3(**6.4**)	65.2(6.9)
	0	59	83.6(**6.4**)	**81.6**(**5.9**)	82.7(**6.1**)	**82.0**(**5.9**)	**82.1**(**6.0**)	**81.7**(**5.9**)	**82.2**(**6.1**)
	0.2	59	84.0(6.5)	**82.2**(**5.4**)	82.7(**5.9**)	**82.4**(**5.6**)	82.6(**5.6**)	**82.2**(**5.6**)	82.7(**5.6**)
0.20	0.4	59	77.4(**6.2**)	74.7(**5.4**)	76.1(**5.8**)	75.2(**5.8**)	75.3(**5.6**)	74.9(**5.4**)	75.9(**5.8**)
	0.6	59	71.5(6.9)	67.5(**5.4**)	69.5(**5.9**)	68.5(**5.6**)	68.7(**5.7**)	68.4(**5.5**)	68.8(**5.7**)
	0.8	59	67.1(9.7)	62.4(7.7)	64.8(8.7)	63.4(8.2)	63.4(8.2)	62.7(8.2)	64.1(8.5)

Notes 1: ρ refers to ICC; CV refers to the coefficient of variation of cluster sizes; $\hat{n}$ refers to the estimated number of clusters.
Notes 2: Robust refers to the t-test with the uncorrected robust sandwich variance estimator; MD refers to the t-test with the bias-corrected sandwich variance estimator due to Mancl and DeRouen; KC refers to the t-test with the bias-corrected sandwich variance estimator due to Kauermann and Carroll; FG refers to the t-test with the bias-corrected sandwich variance estimator due to Fay and Graubard; MD/KC refers to the t-test with the average MD/KC standard error estimator; MD/FG refers to the t-test with the average MD/FG standard error estimator; KC/FG refers to the t-test with the average KC/FG standard error estimator.

Web Table 8. Empirical power and type I error rates (in parentheses) for modified Poisson regression under exchangeable working correlation when the mean cluster size is $\bar{m} = 50$, $P_0 = 0.15$. The required sample size estimate $\hat{n}$ is obtained by "average cluster size method" and ignores variable cluster sizes. Empirical type I error rate between 3.6% and 6.4% and empirical power between 77.5% and 82.5% are in bold font and considered close to nominal according to the margin of error under a binomial model with 1000 replications.

ρ	CV	$\hat{n}$	Robust	MD	KC	FG	MD/KC	MD/FG	KC/FG
	0	11	91.0(7.1)	**79.7**(3.4)	86.5(**5.1**)	83.3(**4.0**)	83.1(**4.3**)	**82.2**(**3.7**)	84.5(**4.6**)
	0.2	11	88.8(7.7)	**78.1**(3.5)	83.9(**5.2**)	**81.1**(**4.0**)	**80.7**(**4.1**)	79.6(3.4)	**82.2**(**4.5**)
0.01	0.4	11	88.3(8.0)	74.9(3.1)	82.8(**5.8**)	**78.8**(**4.2**)	**78.9**(**4.4**)	77.0(**3.9**)	**81.1**(**4.7**)
	0.6	11	87.0(9.1)	69.4(2.4)	**80.8**(**5.4**)	76.1(**4.1**)	75.2(**3.5**)	71.9(**3.0**)	**78.5**(**4.7**)
	0.8	11	86.3(11.1)	67.6(3.8)	**78.7**(7.9)	73.2(6.6)	74.1(**6.0**)	70.9(**5.3**)	76.5(7.3)
	0	21	86.9(**5.0**)	**82.1**(3.3)	84.6(**3.9**)	83.5(**3.7**)	83.6(**3.8**)	82.7(3.4)	84.1(**3.8**)
	0.2	21	85.6(**5.6**)	**78.7**(**3.7**)	82.8(**4.0**)	**80.4**(**3.8**)	**80.7**(**3.7**)	**79.6**(**3.7**)	**81.6**(**4.0**)
0.05	0.4	21	85.3(**5.9**)	**77.5**(**3.8**)	**81.5**(**4.4**)	**79.2**(**3.9**)	**78.9**(**4.1**)	**78.2**(**3.9**)	**80.3**(**4.1**)
	0.6	21	**81.8**(7.3)	77.2(**5.4**)	**79.0**(**6.1**)	**77.9**(**5.8**)	**78.0**(**5.9**)	**77.7**(**5.6**)	**78.7**(**6.0**)
	0.8	21	83.4(8.4)	77.1(**5.6**)	**80.7**(7.3)	**78.7**(**6.4**)	**78.4**(**6.3**)	**78.0**(**6.1**)	**79.8**(6.8)
	0	33	84.3(**6.3**)	**81.7**(**5.4**)	82.9(**5.8**)	**82.3**(**5.6**)	**82.1**(**5.5**)	**81.8**(**5.4**)	**82.5**(**5.6**)
	0.2	33	**82.4**(**5.9**)	**78.6**(**4.8**)	**80.6**(**4.9**)	**79.7**(**4.8**)	**79.8**(**4.8**)	**79.2**(**4.8**)	**80.1**(**4.9**)
0.10	0.4	33	**81.4**(6.6)	**78.1**(**5.2**)	**79.8**(**5.5**)	**79.1**(**5.3**)	**79.1**(**5.3**)	**78.8**(**5.3**)	**79.4**(**5.3**)
	0.6	33	**81.8**(**5.0**)	**78.0**(**3.9**)	**80.2**(**4.6**)	**79.0**(**4.2**)	**79.2**(**4.1**)	**78.6**(**4.1**)	**79.6**(**4.2**)
	0.8	33	**79.5**(7.5)	74.7(**5.6**)	**77.5**(**6.4**)	76.5(**6.0**)	76.1(**5.9**)	75.3(**5.8**)	77.3(**6.1**)
	0	46	**82.5**(**4.8**)	**79.8**(**3.7**)	**80.8**(**4.2**)	**80.2**(**3.8**)	**80.4**(**3.9**)	**80.1**(**3.7**)	**80.5**(**3.9**)
	0.2	46	**79.9**(**5.1**)	**78.4**(**4.5**)	**79.3**(**4.6**)	**79.0**(**4.6**)	**78.9**(**4.6**)	**78.7**(**4.6**)	**79.2**(**4.6**)
0.15	0.4	46	**81.5**(7.6)	**78.3**(**5.7**)	**80.2**(6.7)	**79.2**(**6.2**)	**79.1**(**5.8**)	**78.6**(**5.8**)	**79.6**(**6.3**)
	0.6	46	83.4(**5.2**)	**81.5**(**4.2**)	**82.4**(**4.8**)	**81.8**(**4.4**)	**81.9**(**4.7**)	**81.7**(**4.3**)	**82.1**(**4.7**)
	0.8	46	**78.6**(6.5)	76.1(**5.6**)	77.3(**6.1**)	76.8(**5.8**)	76.8(**5.8**)	76.7(**5.7**)	77.0(**6.0**)
	0	59	83.6(**6.4**)	**81.6**(**5.9**)	82.7(**6.1**)	**82.0**(**5.9**)	**82.1**(**6.0**)	**81.7**(**5.9**)	**82.2**(**6.1**)
	0.2	59	85.2(**5.9**)	83.0(**5.3**)	84.0(**5.4**)	83.3(**5.3**)	83.3(**5.3**)	83.0(**5.3**)	83.5(**5.4**)
0.20	0.4	59	**82.4**(**5.7**)	**80.8**(**5.2**)	**81.5**(**5.2**)	**81.1**(**5.2**)	**81.1**(**5.2**)	**80.9**(**5.2**)	**81.4**(**5.2**)
	0.6	59	**82.5**(**6.0**)	**80.4**(**5.6**)	**81.5**(**5.8**)	**81.1**(**5.6**)	**81.4**(**5.6**)	**80.7**(**5.6**)	**81.4**(**5.6**)
	0.8	59	83.9(6.9)	**81.6**(6.5)	82.9(6.7)	**82.1**(6.5)	**82.1**(6.5)	**81.8**(6.5)	82.8(6.5)

Notes 1: ρ refers to ICC; CV refers to the coefficient of variation of cluster sizes; $\hat{n}$ refers to the estimated number of clusters.
Notes 2: Robust refers to the t-test with the uncorrected robust sandwich variance estimator; MD refers to the t-test with the bias-corrected sandwich variance estimator due to Mancl and DeRouen; KC refers to the t-test with the bias-corrected sandwich variance estimator due to Kauermann and Carroll; FG refers to the t-test with the bias-corrected sandwich variance estimator due to Fay and Graubard; MD/KC refers to the t-test with the average MD/KC standard error estimator; MD/FG refers to the t-test with the average MD/FG standard error estimator; KC/FG refers to the t-test with the average KC/FG standard error estimator.

Web Table 9. Empirical power and type I error rates (in parentheses) for modified Poisson regression under independence working correlation when the mean cluster size is $\bar{m} = 50$, $P_0 = 0.30$. The required sample size estimate $\hat{n}$ is obtained by "average cluster size method" and ignores variable cluster sizes. Empirical type I error rate between 3.6% and 6.4% and empirical power between 77.5% and 82.5% are in bold font and considered close to nominal according to the margin of error under a binomial model with 1000 replications.

ρ	CV	$\hat{n}$	Robust	MD	KC	FG	MD/KC	MD/FG	KC/FG
	0	9	91.5(7.9)	76.2(3.0)	84.6(**5.3**)	**80.9**(**4.2**)	**80.6**(**4.0**)	**78.3**(**3.7**)	**82.5**(**4.6**)
	0.2	9	91.1(6.8)	76.6(3.1)	85.4(**4.6**)	**81.5**(3.3)	**80.4**(3.3)	**78.2**(3.1)	83.4(**4.2**)
0.01	0.4	9	88.9(9.7)	71.4(3.2)	**81.0**(6.5)	77.3(**5.4**)	76.5(**5.0**)	74.2(**4.3**)	**79.0**(**6.0**)
	0.6	9	88.2(9.9)	69.1(3.2)	**80.3**(**5.9**)	75.1(**4.3**)	73.6(**4.2**)	71.9(**3.8**)	**78.3**(**4.9**)
	0.8	9	88.3(12.0)	63.5(**4.4**)	**77.6**(**6.4**)	71.7(**4.9**)	71.3(**4.8**)	67.8(**4.5**)	74.7(**5.7**)
	0	17	86.4(6.7)	**78.3**(**5.0**)	83.0(**5.7**)	**81.4**(**5.3**)	**81.1**(**5.3**)	**80.2**(**5.0**)	**82.3**(**5.5**)
	0.2	17	86.2(7.1)	**78.1**(**4.7**)	**82.2**(**6.1**)	**80.2**(**5.3**)	**79.8**(**5.3**)	**78.7**(**4.9**)	**81.3**(**5.4**)
0.05	0.4	17	**85.0**(7.8)	76.0(**4.2**)	**81.5**(**5.8**)	**78.6**(**5.2**)	**78.7**(**5.4**)	77.3(**4.8**)	**79.8**(**5.7**)
	0.6	17	**83.3**(8.8)	71.7(**5.6**)	**78.5**(7.0)	74.5(**6.1**)	75.2(**6.2**)	73.4(**5.7**)	76.8(6.8)
	0.8	17	**77.8**(11.6)	62.9(**5.7**)	71.4(8.3)	67.4(7.0)	67.5(7.1)	65.5(6.7)	69.0(7.4)
	0	26	84.6(**5.1**)	**80.3**(3.5)	82.7(**3.9**)	**82.2**(**3.9**)	**82.0**(**3.6**)	**81.4**(3.5)	**82.4**(**3.7**)
	0.2	26	83.8(**5.6**)	**77.5**(**4.7**)	**81.5**(**5.2**)	**79.2**(**5.0**)	**79.3**(**5.0**)	**78.1**(**4.8**)	**80.6**(**5.1**)
0.10	0.4	26	**79.9**(**5.4**)	73.4(**3.8**)	76.1(**4.2**)	75.2(**4.1**)	74.7(**4.0**)	74.4(**3.9**)	75.5(**4.1**)
	0.6	26	76.3(**5.3**)	66.1(3.1)	71.8(**4.1**)	68.6(**3.8**)	68.9(**3.8**)	67.0(3.2)	70.7(**4.0**)
	0.8	26	72.1(8.6)	60.2(**4.5**)	66.9(**6.2**)	62.9(**5.2**)	62.8(**5.4**)	61.4(**4.9**)	64.9(**5.7)**
	0	36	**81.1**(**6.2**)	**78.8**(**4.2**)	**80.2**(**5.2**)	**79.8**(**4.5**)	**79.8**(**4.5**)	**79.4**(**4.3**)	**80.0**(**4.9**)
	0.2	36	**81.3**(**6.0**)	**77.5**(**4.5**)	**79.4**(**5.1**)	**78.8**(**4.7**)	**78.8**(**4.8**)	**78.1**(**4.7**)	**79.1**(**4.9**)
0.15	0.4	36	**78.0**(**6.3**)	73.0(**4.6**)	75.7(**5.3**)	74.8(**4.9**)	74.6(**4.9**)	74.0(**4.9**)	75.4(**5.0**)
	0.6	36	72.6(7.8)	66.5(**5.4**)	70.1(7.0)	68.4(**6.1**)	68.3(**6.3**)	67.6(**5.9**)	69.1(6.8)
	0.8	36	70.4(7.0)	61.5(**4.8**)	66.3(**5.7**)	63.9(**5.2**)	64.7(**5.0**)	62.9(**5.0**)	65.3(**5.3**)
	0	46	82.8(**4.8**)	**79.9**(**4.0**)	**81.9**(**4.4**)	**80.9**(**4.0**)	**80.9**(**4.0**)	**80.3**(**4.0**)	**81.5**(**4.1**)
	0.2	46	**78.8**(**4.7**)	75.6(**4.2**)	**77.6**(**4.6**)	76.5(**4.2**)	76.5(**4.2**)	76.1(**4.2**)	77.2(**4.4**)
0.20	0.4	46	75.6(7.4)	72.4(**4.8**)	73.8(**6.1**)	73.2(**5.4**)	73.4(**5.4**)	72.7(**5.2**)	73.4(**5.5**)
	0.6	46	75.8(**5.6**)	69.6(**4.0**)	73.0(**5.1**)	71.1(**4.6**)	70.8(**4.7**)	70.3(**4.4**)	72.5(**4.9**)
	0.8	46	68.1(7.0)	62.8(**5.6**)	65.6(6.5)	64.1(**6.0**)	64.3(**5.6**)	63.4(**5.6**)	64.9(**6.2**)

Notes 1: ρ refers to ICC; CV refers to the coefficient of variation of cluster sizes; $\hat{n}$ refers to the estimated number of clusters.
Notes 2: Robust refers to the t-test with the uncorrected robust sandwich variance estimator; MD refers to the t-test with the bias-corrected sandwich variance estimator due to Mancl and DeRouen; KC refers to the t-test with the bias-corrected sandwich variance estimator due to Kauermann and Carroll; FG refers to the t-test with the bias-corrected sandwich variance estimator due to Fay and Graubard; MD/KC refers to the t-test with the average MD/KC standard error estimator; MD/FG refers to the t-test with the average MD/FG standard error estimator; KC/FG refers to the t-test with the average KC/FG standard error estimator.

Web Table 10. Empirical power and type I error rates (in parentheses) for modified Poisson regression under exchangeable working correlation when the mean cluster size is $\bar{m} = 50$, $P_0 = 0.30$. The required sample size estimate $\hat{n}$ is obtained by “average cluster size method” and ignores variable cluster sizes. Empirical type I error rate between 3.6% and 6.4% and empirical power between 77.5% and 82.5% are in bold font and considered close to nominal according to the margin of error under a binomial model with 1000 replications.

ρ	CV	$\hat{n}$	Robust	MD	KC	FG	MD/KC	MD/FG	KC/FG
0.01	0	9	91.5(7.9)	76.2(3.0)	84.6(**5.3**)	**80.9**(**4.2**)	**80.6**(**4.0**)	**78.3**(**3.7**)	**82.5**(**4.6**)
	0.2	9	90.9(6.8)	76.3(2.7)	85.2(**3.8**)	**81.3**(3.0)	**80.6**(3.2)	**78.4**(2.9)	83.2(3.5)
	0.4	9	88.5(10.3)	70.3(3.3)	**81.3**(6.7)	76.3(**5.3**)	75.8(**4.7**)	73.3(**3.9**)	**79.4**(**5.8**)
	0.6	9	86.8(9.4)	66.6(2.8)	**78.8**(**5.7**)	74.2(**4.0**)	72.6(3.5)	70.7(3.2)	76.3(**4.6**)
	0.8	9	86.1(11.7)	59.9(2.7)	76.3(**5.0**)	68.1(**3.8**)	67.5(3.4)	64.1(2.8)	72.2(**4.3**)
0.05	0	17	86.4(6.7)	**78.3**(**5.0**)	83.0(**5.7**)	**81.4**(**5.3**)	**81.1**(**5.3**)	**80.2**(**5.0**)	**82.3**(**5.5**)
	0.2	17	86.4(6.8)	**78.9**(**4.8**)	83.1(**5.8**)	**81.3**(**5.3**)	**80.7**(**5.2**)	**80.4**(**4.8**)	**82.2**(**5.6**)
	0.4	17	86.5(8.1)	**79.8**(**4.4**)	82.9(**5.9**)	**81.8**(**5.2**)	**81.7**(**5.1**)	**81.2**(**4.8**)	**82.5**(**5.5**)
	0.6	17	85.5(7.9)	**78.2**(**5.7**)	**82.4**(6.9)	**80.8**(**6.0**)	**80.4**(**6.1**)	**79.8**(**5.9**)	**81.3**(6.5)
	0.8	17	**82.4**(9.0)	72.8(**6.0**)	**78.1**(7.2)	75.9(**6.3**)	75.6(6.5)	74.4(**6.3**)	76.7(6.7)
0.10	0	26	84.6(**5.1**)	**80.3**(3.5)	82.7(**3.9**)	**82.2**(**3.6**)	**82.0**(**3.6**)	**81.4**(3.5)	**82.4**(**3.7**)
	0.2	26	85.0(**5.8**)	**80.0**(3.5)	82.8(**4.7**)	**81.4**(**4.1**)	**81.4**(**4.1**)	**80.9**(**4.0**)	**82.1**(**4.4**)
	0.4	26	83.4(**5.6**)	**78.3**(**3.9**)	**81.5**(**4.6**)	**79.4**(**4.1**)	**79.4**(**4.0**)	**79.3**(**3.9**)	**80.2**(**4.2**)
	0.6	26	**81.4**(**3.8**)	76.2(2.5)	**79.2**(3.1)	**77.9**(2.9)	**77.7**(**3.0**)	76.9(2.6)	**78.6**(3.1)
	0.8	26	**80.8**(**4.8**)	74.5(3.1)	77.3(**4.0**)	76.0(3.2)	76.2(3.1)	75.1(3.1)	76.6(3.4)
0.15	0	36	**81.1**(**6.2**)	**78.8**(**4.2**)	**80.2**(**5.2**)	**79.8**(**4.5**)	**79.8**(**4.5**)	**79.4**(**4.3**)	**80.0**(**4.9**)
	0.2	36	83.0(**6.0**)	**78.8**(**4.3**)	**80.9**(**5.5**)	**80.0**(**4.8**)	**80.0**(**4.8**)	**79.3**(**4.6**)	**80.4**(**5.3**)
	0.4	36	82.9(**5.2**)	**79.1**(**4.1**)	**80.3**(**4.8**)	**79.5**(**4.2**)	**79.5**(**4.2**)	**79.3**(**4.1**)	**79.6**(**4.6**)
	0.6	36	**81.0**(**5.5**)	77.2(**4.4**)	**79.0**(**5.1**)	**78.1**(**4.8**)	**78.1**(**4.8**)	**77.7**(**4.6**)	**78.7**(**5.0**)
	0.8	36	**79.3**(**6.3**)	76.3(**5.1**)	**78.0**(**5.7**)	77.3(**5.2**)	76.8(**5.4**)	76.3(**5.2**)	**77.8**(**5.6**)
0.20	0	46	82.8(**4.8**)	**79.9**(**4.0**)	**81.9**(**4.4**)	**80.9**(**4.0**)	**80.9**(**4.0**)	**80.3**(**4.0**)	**81.5**(**4.1**)
	0.2	46	**80.2**(**4.1**)	76.7(**3.9**)	**77.9**(**4.0**)	**77.5**(**4.0**)	**77.6**(**4.0**)	**77.0**(**4.0**)	**77.8**(**4.0**)
	0.4	46	**81.6**(**5.4**)	**78.3**(**4.9**)	**80.5**(**5.0**)	**79.6**(**5.0**)	**79.6**(**5.0**)	**79.1**(**5.0**)	**80.1**(**5.0**)
	0.6	46	84.2(**4.6**)	**82.4**(**3.9**)	83.1(**4.2**)	82.6(**4.1**)	82.7(**4.1**)	82.6(**4.0**)	83.0(**4.2**)
	0.8	46	**81.5**(**5.6**)	**78.1**(**4.4**)	**80.2**(**4.9**)	**79.2**(**4.8**)	**79.4**(**4.7**)	**78.9**(**4.4**)	**79.5**(**4.8**)

Notes 1: ρ refers to ICC; CV refers to the coefficient of variation of cluster sizes; $\hat{n}$ refers to the estimated number of clusters.
Notes 2: Robust refers to the t-test with the uncorrected robust sandwich variance estimator; MD refers to the t-test with the bias-corrected sandwich variance estimator due to Mancl and DeRouen; KC refers to the t-test with the bias-corrected sandwich variance estimator due to Kauermann and Carroll; FG refers to the t-test with the bias-corrected sandwich variance estimator due to Fay and Graubard; MD/KC refers to the t-test with the average MD/KC standard error estimator; MD/FG refers to the t-test with the average MD/FG standard error estimator; KC/FG refers to the t-test with the average KC/FG standard error estimator.

Web Table 11. Empirical power and type I error rates (in parentheses) for modified Poisson regression under independence working correlation when the mean cluster size is $\bar{m} = 100$, $P_0 = 0.15$. The required sample size estimate $\hat{n}$ is obtained by "average cluster size method" and ignores variable cluster sizes. Empirical type I error rate between 3.6% and 6.4% and empirical power between 77.5% and 82.5% are in bold font and considered close to nominal according to the margin of error under a binomial model with 1000 replications.

ρ	CV	$\hat{n}$	Robust	MD	KC	FG	MD/KC	MD/FG	KC/FG
	0	8	89.6(8.5)	73.9(3.5)	83.5(**5.8**)	**78.0**(**4.9**)	**78.7**(**5.1**)	75.8(**4.6**)	**80.9**(**5.4**)
	0.2	8	89.3(7.5)	71.9(2.7)	**82.0**(**4.9**)	77.4(**3.9**)	77.3(**3.8**)	74.4(3.0)	**80.2**(**4.3**)
0.01	0.4	8	88.8(8.4)	69.3(2.7)	**81.7**(**4.8**)	74.9(**3.9**)	76.7(3.5)	72.3(3.0)	**78.3(4.0)**
	0.6	8	88.6(9.2)	62.5(2.8)	**78.4**(**5.5**)	70.8(**4.5**)	71.3(**3.9**)	66.9(3.5)	75.4(**4.9**)
	0.8	8	86.5(12.5)	59.1(**4.5**)	75.4(**8.0**)	66.2(**6.1**)	66.4(**6.0**)	62.4(**5.1**)	70.7(7.1)
	0	18	84.6(7.3)	**77.5**(**4.2**)	**81.3**(**5.4**)	**79.8**(4.8)	**79.5**(4.9)	**78.6**(**4.5**)	**80.4**(**5.1**)
	0.2	18	84.7(**5.3**)	**79.0**(**3.6**)	82.6(**4.2**)	**81.3**(**3.7**)	**81.3**(**3.9**)	**79.9**(3.5)	**81.9**(**4.0**)
0.05	0.4	18	82.7(6.6)	72.0(**3.7**)	**77.8**(**5.5**)	74.8(**4.4**)	74.5(**4.7**)	73.1(**4.2**)	76.6(**4.9**)
	0.6	18	**80.9**(9.0)	68.7(**4.9**)	75.9(**6.4**)	72.2(**5.2**)	72.3(**5.1**)	70.0(**4.9**)	74.2(**5.8**)
	0.8	18	76.0(9.4)	62.7(**4.2**)	69.5(6.5)	65.2(**5.5**)	66.5(**5.5**)	64.2(**5.1**)	67.1(**6.0**)
	0	31	**82.4**(7.6)	**79.3**(**5.0**)	**80.5**(**5.8**)	**79.8**(**5.5**)	**79.8**(**5.3**)	**79.5**(**5.2**)	**80.2**(**5.5**)
	0.2	31	83.0(**5.1**)	**79.2**(4.4)	**81.0**(4.5)	**80.0**(4.4)	**79.9**(4.4)	**79.6**(4.4)	**80.4**(4.4)
0.10	0.4	31	**81.0**(7.6)	74.9(**5.6**)	77.3(6.5)	76.3(**5.9**)	76.6(**6.0**)	75.9(**5.6**)	76.6(**6.1**)
	0.6	31	76.4(7.1)	67.9(**4.9**)	71.8(**5.5**)	68.9(**5.1**)	69.4(**5.0**)	68.4(**5.0**)	69.8(**5.2**)
	0.8	31	71.1(9.0)	62.5(**4.8**)	67.4(6.8)	64.0(**5.8**)	64.9(**6.2**)	63.8(**5.4**)	65.7(**6.2**)
	0	44	82.7(7.2)	**80.6**(**5.5**)	**81.4**(6.6)	**81.1**(**5.8**)	**81.1**(**5.9**)	**80.9**(**5.6**)	**81.1**(6.5)
	0.2	44	**81.8**(**5.9**)	**78.7**(**5.1**)	**80.2**(**5.5**)	**79.9**(**5.3**)	**79.9**(**5.3**)	**79.3**(**5.3**)	**80.0**(**5.4**)
0.15	0.4	44	**77.8**(**5.5**)	74.5(**4.3**)	76.2(**4.6**)	75.3(**4.5**)	75.2(**4.4**)	74.7(**4.3**)	75.6(**4.6**)
	0.6	44	74.9(6.7)	69.2(**5.5**)	72.2(**6.3**)	71.0(**5.8**)	70.6(**6.2**)	**79.8**(**5.5**)	71.9(**6.2**)
	0.8	44	69.2(7.7)	62.0(**5.1**)	65.6(**6.3**)	63.7(**5.5**)	63.9(**5.9**)	62.6(**5.4**)	64.7(**5.9**)
	0	57	**82.5**(**6.0**)	**80.3**(**5.2**)	**81.8**(**5.4**)	**80.9**(**5.4**)	**80.9**(**5.3**)	**80.8**(**5.3**)	**81.2**(**5.4**)
	0.2	57	**81.1**(**5.7**)	**79.2**(4.4)	**80.0**(4.6)	**79.7**(4.5)	**79.5**(4.4)	**79.3**(4.4)	**79.9**(4.5)
0.20	0.4	57	**78.0**(**5.3**)	74.4(**4.1**)	75.8(**4.8**)	74.9(**4.7**)	75.0(**4.7**)	74.7(**4.5**)	75.3(**4.7**)
	0.6	57	72.0(**6.1**)	67.5(**4.9**)	70.4(**5.4**)	68.9(**5.0**)	69.0(**5.0**)	68.3(**5.0**)	69.4(**5.3**)
	0.8	57	66.7(7.6)	62.8(**5.7**)	65.2(**6.3**)	63.6(**6.0**)	64.2(**5.9**)	63.4(**5.8**)	64.4(**6.1**)

Notes 1: ρ refers to ICC; CV refers to the coefficient of variation of cluster sizes; $\hat{n}$ refers to the estimated number of clusters.
Notes 2: Robust refers to the t-test with the uncorrected robust sandwich variance estimator; MD refers to the t-test with the bias-corrected sandwich variance estimator due to Mancl and DeRouen; KC refers to the t-test with the bias-corrected sandwich variance estimator due to Kauermann and Carroll; FG refers to the t-test with the bias-corrected sandwich variance estimator due to Fay and Graubard; MD/KC refers to the t-test with the average MD/KC standard error estimator; MD/FG refers to the t-test with the average MD/FG standard error estimator; KC/FG refers to the t-test with the average KC/FG standard error estimator.

Web Table 12. Empirical power and type I error rates (in parentheses) for modified Poisson regression under exchangeable working correlation when the mean cluster size is $\bar{m} = 100$, $P_0 = 0.15$. The required sample size estimate $\hat{n}$ is obtained by “average cluster size method” and ignores variable cluster sizes. Empirical type I error rate between 3.6% and 6.4% and empirical power between 77.5% and 82.5% are in bold font and considered close to nominal according to the margin of error under a binomial model with 1000 replications.

ρ	CV	$\hat{n}$	Robust	MD	KC	FG	MD/KC	MD/FG	KC/FG
	0	8	89.6(8.5)	73.9(3.5)	83.5(**5.8**)	**78.0**(**4.9**)	**78.7**(**5.1**)	75.8(**4.6**)	**80.9**(**5.4**)
	0.2	8	89.1(7.6)	72.8(2.1)	82.8(**4.5**)	**78.1**(3.3)	77.3(3.3)	74.4(3.0)	**80.0**(**4.0**)
0.01	0.4	8	88.8(8.7)	70.2(2.6)	**81.9**(**4.6**)	75.8(**3.7**)	76.1(3.1)	72.7(2.8)	**79.1**(**4.0**)
	0.6	8	87.8(9.6)	62.8(2.2)	**78.1**(**5.6**)	71.4(**4.5**)	71.1(**3.6**)	67.3(3.2)	75.9(**5.2**)
	0.8	8	84.9(12.0)	58.4(2.5)	73.9(7.1)	66.0(**5.6**)	66.4(**4.1**)	62.1(3.3)	70.2(6.5)
	0	18	84.6(7.3)	**77.5**(**5.0**)	**81.3**(**5.8**)	**79.8**(**5.5**)	**79.5**(**5.3**)	**78.6**(**5.2**)	**80.4**(**5.5**)
	0.2	18	86.1(**5.2**)	**80.3**(2.6)	83.6(**4.2**)	**82.2**(3.3)	**82.2**(3.3)	**81.3**(2.9)	83.0(**3.7**)
0.05	0.4	18	84.9(**5.2**)	77.1(3.5)	**81.1**(**4.6**)	**79.2**(**4.0**)	**79.2**(**4.1**)	**78.1**(**3.8**)	**80.3**(**4.3**)
	0.6	18	83.3(8.1)	76.8(**5.1**)	**80.7**(6.8)	**78.7**(**6.2**)	**78.9**(**6.0**)	**77.9**(**5.7**)	**80.4**(6.5)
	0.8	18	**80.8**(8.2)	72.4(**5.2**)	77.1(6.8)	74.4(**5.9**)	74.8(**5.9**)	73.7(**5.5**)	76.1(**6.2**)
	0	31	**82.4**(7.6)	**79.3**(**5.0**)	**80.5**(**5.8**)	**79.8**(**5.5**)	**79.8**(**5.3**)	**79.5**(**5.2**)	**80.2**(**5.5**)
	0.2	31	84.6(**5.8**)	**80.3**(**4.2**)	**82.4**(**4.8**)	**81.3**(**4.5**)	**81.4**(**4.5**)	**80.8**(**4.4**)	**82.0**(**4.7**)
0.10	0.4	31	84.6(**6.1**)	**79.9**(**5.3**)	82.8(**5.9**)	**81.1**(**5.6**)	**81.5**(**5.6**)	**80.4**(**5.4**)	**81.8**(**5.8**)
	0.6	31	83.6(**5.2**)	**79.8**(**4.5**)	**81.3**(**4.9**)	**80.4**(**4.7**)	**80.2**(**4.7**)	**80.0**(**4.6**)	**80.9**(**4.8**)
	0.8	31	**81.2**(**6.3**)	76.2(**4.7**)	**78.9**(**5.4**)	77.2(**4.9**)	77.2(**4.9**)	76.8(**4.8**)	**78.0**(**5.1**)
	0	44	82.7(7.2)	**80.6**(**5.5**)	**81.4**(6.6)	**81.1**(**5.8**)	**81.1**(**5.9**)	**80.9**(**5.6**)	**81.1**(6.5)
	0.2	44	82.9(**5.8**)	**80.6**(**4.5**)	**81.8**(**5.2**)	**81.0**(**5.0**)	**80.9**(**5.0**)	**80.7**(**4.8**)	**81.5**(**5.2**)
0.15	0.4	44	82.7(**5.9**)	**80.4**(**4.8**)	**81.8**(**5.4**)	**81.1**(**5.1**)	**80.9**(**5.2**)	**80.6**(**4.8**)	**81.1**(**5.4**)
	0.6	44	84.0(**6.3**)	**81.3**(**5.0**)	83.0(**5.5**)	**81.7**(**5.3**)	**81.8**(**5.3**)	**81.5**(**5.2**)	**82.5**(**5.4**)
	0.8	44	**80.5**(6.8)	**78.1**(**5.8**)	**79.4**(6.6)	**79.2**(**5.8**)	**79.2**(**6.0**)	**78.8**(**5.8**)	**79.2**(**6.1**)
	0	57	**82.5**(**6.0**)	**80.3**(**5.2**)	**81.8**(**5.4**)	**80.9**(**5.4**)	**80.9**(**5.3**)	**80.8**(**5.3**)	**81.2**(**5.4**)
	0.2	57	83.7(**5.2**)	**81.9**(**4.0**)	82.9(**4.7**)	**82.4**(**4.4**)	**82.5**(**4.5**)	**82.0**(**4.1**)	82.6(**4.5**)
0.20	0.4	57	82.8(**5.1**)	**81.2**(**4.4**)	**82.1**(**4.6**)	**81.5**(**4.5**)	**81.5**(**4.6**)	**81.4**(**4.5**)	**81.5**(**4.6**)
	0.6	57	82.6(**5.4**)	**80.5**(**4.9**)	**81.7**(**5.1**)	**81.0**(**5.1**)	**81.1**(**5.1**)	**80.6**(**5.0**)	**81.4**(**5.1**)
	0.8	57	**79.5**(**5.6**)	**78.2**(**4.7**)	**78.9**(**5.1**)	**78.6**(**4.9**)	**78.5**(**4.7**)	**78.3**(**4.7**)	**78.9**(**4.9**)

Notes 1: ρ refers to ICC; CV refers to the coefficient of variation of cluster sizes; $\hat{n}$ refers to the estimated number of clusters.
Notes 2: Robust refers to the t-test with the uncorrected robust sandwich variance estimator; MD refers to the t-test with the bias-corrected sandwich variance estimator due to Mancl and DeRouen; KC refers to the t-test with the bias-corrected sandwich variance estimator due to Kauermann and Carroll; FG refers to the t-test with the bias-corrected sandwich variance estimator due to Fay and Graubard; MD/KC refers to the t-test with the average MD/KC standard error estimator; MD/FG refers to the t-test with the average MD/FG standard error estimator; KC/FG refers to the t-test with the average KC/FG standard error estimator.

Web Table 13. Empirical power and type I error rates (in parentheses) for modified Poisson regression under independence working correlation when the mean cluster size is $\bar{m} = 100$, $P_0 = 0.30$. The required sample size estimate $\hat{n}$ is obtained by "average cluster size method" and ignores variable cluster sizes. Empirical type I error rate between 3.6% and 6.4% and empirical power between 77.5% and 82.5% are in bold font and considered close to nominal according to the margin of error under a binomial model with 1000 replications.

ρ	CV	$\hat{n}$	Robust	MD	KC	FG	MD/KC	MD/FG	KC/FG
0.01	0	7	91.4(8.4)	73.7(3.1)	84.6(**5.1**)	**80.4**(**3.9**)	**79.5**(**4.5**)	76.6(**3.7**)	**82.5**(**4.8**)
	0.2	7	91.2(9.1)	70.5(2.9)	83.1(**5.6**)	**78.1**(**4.2**)	77.4(**4.1**)	74.3(3.3)	**80.8**(**4.9**)
	0.4	7	92.2(10.1)	69.5(2.7)	**81.4**(**5.8**)	76.3(**4.5**)	75.5(**3.9**)	73.6(3.4)	**78.2**(**5.2**)
	0.6	7	91.8(10.6)	64.0(2.7)	**82.0**(**6.1**)	73.7(**4.5**)	73.0(**4.5**)	68.6(**3.8**)	**78.1**(**4.9**)
	0.8	7	90.5(14.7)	58.7(**4.1**)	**79.1**(8.8)	70.7(6.7)	67.8(**5.7**)	63.6(**5.1**)	75.3(7.6)
0.05	0	15	85.9(**5.0**)	**78.3**(2.6)	**82.4**(3.3)	**80.3**(2.9)	**79.6**(2.9)	**79.2**(2.7)	**81.1**(3.3)
	0.2	15	83.8(6.7)	76.1(**4.1**)	**80.0**(**5.9**)	**78.0**(**4.8**)	**77.8**(**4.9**)	76.8(**4.5**)	**79.1**(**5.3**)
	0.4	15	83.8(7.0)	71.8(**3.8**)	**78.3**(**5.0**)	75.1(**4.4**)	74.8(**4.5**)	73.4(**4.1**)	76.3(**4.8**)
	0.6	15	**81.8**(9.0)	67.6(**4.9**)	75.3(7.0)	71.1(**6.1**)	71.9(**6.0**)	69.7(**5.4**)	73.5(**6.4**)
	0.8	15	**79.6**(10.8)	61.6(**5.7**)	71.3(8.4)	66.5(7.0)	66.5(**6.4**)	63.6(**5.9**)	68.3(7.6)
0.10	0	24	83.1(**5.3**)	**77.8**(3.4)	**80.9**(**4.9**)	**79.1**(**4.4**)	**79.3**(**4.0**)	**78.1**(**3.6**)	**80.1**(**4.6**)
	0.2	24	83.6(7.4)	**77.9**(**4.6**)	**81.2**(**5.7**)	**80.0**(**5.3**)	**79.7**(**5.4**)	**79.4**(**4.9**)	**80.4**(**5.4**)
	0.4	24	**81.4**(7.8)	75.0(**5.4**)	**78.5**(6.8)	76.7(**6.2**)	76.2(**6.3**)	75.5(**5.8**)	**77.8**(6.6)
	0.6	24	**77.7**(7.2)	68.0(**4.0**)	73.8(**5.8**)	70.7(**4.6**)	70.8(**4.7**)	68.8(**4.3**)	72.4(**5.2**)
	0.8	24	72.7(9.2)	63.2(**5.5**)	68.6(6.9)	65.6(**6.3**)	66.0(**6.1**)	64.6(**6.0**)	67.5(6.5)
0.15	0	34	83.3(**5.0**)	**79.3**(**4.0**)	**81.5**(**4.6**)	**80.1**(**4.3**)	**80.2**(**4.2**)	**79.8**(**4.1**)	**80.6**(**4.5**)
	0.2	34	**82.1**(**4.5**)	**78.1**(**3.8**)	**80.1**(**4.1**)	**79.1**(**4.0**)	**79.1**(**4.0**)	**78.4**(**3.9**)	**79.6**(**4.0**)
	0.4	34	**78.6**(7.5)	73.2(**6.1**)	75.4(6.6)	74.5(**6.3**)	74.3(**6.3**)	74.7(**6.2**)	75.2(6.5)
	0.6	34	70.4(8.0)	63.8(**5.5**)	66.6(6.8)	64.9(**6.1**)	64.9(**6.1**)	64.6(**5.8**)	65.4(**6.4**)
	0.8	34	70.3(8.9)	62.2(**6.0**)	65.9(7.5)	63.5(6.8)	63.6(6.7)	62.6(**6.3**)	64.7(7.1)
0.20	0	44	**81.8**(**5.4**)	**79.5**(**4.8**)	**80.9**(**4.9**)	**80.4**(**4.8**)	**80.4**(**4.8**)	**80.2**(**4.8**)	**80.6**(**4.9**)
	0.2	44	**79.4**(**5.9**)	76.6(**4.7**)	77.4(**5.5**)	76.8(**5.0**)	76.8(**4.9**)	76.8(**4.8**)	76.9(**5.1**)
	0.4	44	**78.5**(**6.3**)	74.0(**4.9**)	76.3(**5.5**)	75.2(**5.2**)	75.4(**5.0**)	74.6(**5.0**)	76.0(**5.2**)
	0.6	44	73.3(**6.4**)	68.8(**4.5**)	71.2(**5.5**)	70.0(**4.9**)	70.3(**4.9**)	69.2(**4.7**)	70.9(**5.1**)
	0.8	44	66.9(7.9)	59.5(**5.7**)	63.1(6.7)	61.1(**5.9**)	61.1(**6.3**)	60.6(**5.8**)	61.9(6.6)

Notes 1: ρ refers to ICC; CV refers to the coefficient of variation of cluster sizes; $\hat{n}$ refers to the estimated number of clusters.
Notes 2: Robust refers to the t-test with the uncorrected robust sandwich variance estimator; MD refers to the t-test with the bias-corrected sandwich variance estimator due to Mancl and DeRouen; KC refers to the t-test with the bias-corrected sandwich variance estimator due to Kauermann and Carroll; FG refers to the t-test with the bias-corrected sandwich variance estimator due to Fay and Graubard; MD/KC refers to the t-test with the average MD/KC standard error estimator; MD/FG refers to the t-test with the average MD/FG standard error estimator; KC/FG refers to the t-test with the average KC/FG standard error estimator.

Web Table 14. Empirical power and type I error rates (in parentheses) for modified Poisson regression under exchangeable working correlation when the mean cluster size is $\bar{m} = 100$, $P_0 = 0.30$. The required sample size estimate $\hat{n}$ is obtained by "average cluster size method" and ignores variable cluster sizes. Empirical type I error rate between 3.6% and 6.4% and empirical power between 77.5% and 82.5% are in bold font and considered close to nominal according to the margin of error under a binomial model with 1000 replications.

ρ	CV	$\hat{n}$	Robust	MD	KC	FG	MD/KC	MD/FG	KC/FG
0.01	0	7	91.4(8.4)	73.7(3.1)	84.6(**5.1**)	**80.4(3.9)**	**79.5(4.5)**	76.6(**3.7**)	**82.5(4.8)**
	0.2	7	90.9(8.9)	69.9(2.9)	83.4(**5.5**)	**78.2(4.3)**	76.6(**4.4**)	73.6(3.3)	**81.0(4.7)**
	0.4	7	91.0(9.8)	68.9(1.9)	**81.9(5.0)**	76.0(**3.6**)	74.1(3.4)	71.6(2.6)	**79.0(4.3)**
	0.6	7	90.7(10.8)	61.2(1.8)	**81.4(5.0)**	72.8(3.4)	72.1(2.8)	67.3(2.3)	77.0(**4.0**)
	0.8	7	88.1(13.5)	54.1(2.5)	76.8(7.1)	66.0(**5.3**)	65.4(**4.0**)	60.1(3.2)	71.8(**6.4**)
0.05	0	15	85.9(**5.0**)	**78.3**(2.6)	**82.4**(3.3)	**80.3**(2.9)	**79.6**(2.9)	**79.2**(2.7)	**81.1**(3.3)
	0.2	15	85.1(6.7)	76.6(**3.8**)	**80.8(5.5)**	**78.8(4.7)**	**78.5(4.7)**	**78.0(4.4)**	**79.8(5.0)**
	0.4	15	85.2(6.8)	76.0(3.4)	**81.5(4.5)**	**78.9(3.6)**	**78.7(3.6)**	**77.6**(3.4)	**80.1(3.9)**
	0.6	15	85.3(**8.2**)	76.0(**4.5**)	**80.9**(6.7)	**79.3(5.7)**	**79.5(5.6)**	**77.8(5.1)**	**80.4(6.2)**
	0.8	15	82.8(8.4)	72.6(**4.2**)	**78.2(6.0)**	75.5(**4.9**)	75.4(**4.9**)	73.9(**4.7**)	77.0(**5.4**)
0.10	0	24	83.1(**5.3**)	**77.8**(3.4)	**80.9(4.9)**	**79.1(4.4)**	**79.3(4.0)**	**78.1(3.6)**	**80.1(4.6)**
	0.2	24	83.9(7.2)	**79.7(5.3)**	**82.5(6.0)**	**81.5(5.6)**	**81.2(5.5)**	**80.5(5.3)**	**82.0(5.7)**
	0.4	24	83.6(6.6)	**79.3(5.0)**	**80.9(6.0)**	**80.2(5.5)**	**80.5(5.5)**	**79.5(5.3)**	**80.8(5.8)**
	0.6	24	83.0(**5.1**)	**78.1(3.8)**	**79.9(4.4)**	**78.9(3.9)**	**78.7(4.1)**	**78.5(3.9)**	**79.2(4.2)**
	0.8	24	83.8(6.5)	76.9(**4.9**)	**81.0(5.9)**	**78.8(5.4)**	**78.8(5.5)**	**78.0(5.2)**	**80.1(5.9)**
0.15	0	34	83.3(**5.0**)	**79.3(4.0)**	**81.5(4.6)**	**80.1(4.3)**	**80.2(4.2)**	**79.8(4.1)**	**80.6(4.5)**
	0.2	34	83.3(**4.5**)	**80.2**(3.5)	**81.8(3.8)**	**80.9(3.7)**	**80.8(3.7)**	**80.6**(3.5)	**81.3(3.7)**
	0.4	34	83.2(6.7)	**79.8(5.1)**	**81.6(5.7)**	**80.5(5.2)**	**80.5(5.2)**	**80.2(5.2)**	**81.0(5.4)**
	0.6	34	**79.1(5.6)**	75.6(**4.5**)	**77.5(5.0)**	76.7(**4.6**)	76.7(**4.6**)	76.3(**4.5**)	77.1(**4.8**)
	0.8	34	85.1(**5.6**)	**81.4(3.9)**	83.5(**5.1**)	83.0(**4.3**)	82.8(**4.4**)	**82.3(4.1)**	83.1(**4.6**)
0.20	0	44	**81.8(5.4)**	**79.5(4.8)**	**80.9(4.9)**	**80.4(4.8)**	**80.4(4.8)**	**80.2(4.8)**	**80.6(4.9)**
	0.2	44	**81.7(4.9)**	**78.5(3.9)**	**80.3(4.2)**	**79.5(4.1)**	**79.5(4.1)**	**78.9(4.0)**	**79.6(4.1)**
	0.4	44	83.4(**5.6**)	**80.7(4.2)**	**82.1(5.0)**	**81.3(4.3)**	**81.4(4.2)**	**80.9(4.2)**	**81.9(4.7)**
	0.6	44	82.8(**5.5**)	**80.7(4.9)**	**81.9(5.3)**	**81.2(5.1)**	**81.2(5.2)**	**81.0(5.0)**	**81.4(5.3)**
	0.8	44	**82.3(6.3)**	**80.4(5.1)**	**81.3(5.8)**	**80.7(5.2)**	**80.7(5.4)**	**80.5(5.2)**	**80.9(5.4)**

Notes 1: ρ refers to ICC; CV refers to the coefficient of variation of cluster sizes; $\hat{n}$ refers to the estimated number of clusters.
Notes 2: Robust refers to the t-test with the uncorrected robust sandwich variance estimator; MD refers to the t-test with the bias-corrected sandwich variance estimator due to Mancl and DeRouen; KC refers to the t-test with the bias-corrected sandwich variance estimator due to Kauermann and Carroll; FG refers to the t-test with the bias-corrected sandwich variance estimator due to Fay and Graubard; MD/KC refers to the t-test with the average MD/KC standard error estimator; MD/FG refers to the t-test with the average MD/FG standard error estimator; KC/FG refers to the t-test with the average KC/FG standard error estimator.

C. Comparisons with Sample Size Formulas based on Risk Differences

To clarify differences between the proposed sample size formulas based on the relative risk and those based on the risk difference, we provide a detailed comparison. As suggested by a reviewer, we focus on the Cornfield's approach[1] and the Donner et al's approach.[2] In fact, as will be seen in due course, these two approaches are equivalent. Recall the general set up of sample size requirements in Section 2.3 of the main manuscript. Given the pre-specified type I error rate ϵ_1, and type II error rate ϵ_2, the general sample size requirement in a CRT based on a t-test is given by

$$n \geq \frac{(t_{n-2,\epsilon_1/2} + t_{n-2,\epsilon_2})^2 \sigma^2}{\Delta^2},$$

where $t_{n-2,q}$ is the qth quantile of the t distribution with $n-2$ degree of freedom, Δ is the effect size and σ^2 is the corresponding large-sample variance of the *test statistic*. While the Cornfield's approach[1] and the Donner et al's approach[2] all used a z-test, we provide versions of their approach based on the t-test since (i) it is more comparable with our approach in the main text (we would like to exclude difference arising from the choice of the z- and t-quantiles); and (2) the t-test frequently provides better control of empirical type I error rates in CRTs when the number of clusters n is not large, as evidenced by our own research and others.[3–8] Before we compare their approaches with ours, we first note that both the Cornfield's approach and the Donner et al's approach make two (strong) assumptions:

(1) Both the Cornfield's approach and the Donner et al's approach assume equal randomization, and therefore only applies when $\pi = 1/2$, while our approach allows for general values of randomization proportion $\pi \in (0,1)$;

(2) Both the Cornfield's approach and the Donner et al's approach assume equal cluster sizes, while our proposal accounts for variable cluster sizes through explicit variance inflation factors, $\text{VIF}_{\text{indep}}$ and VIF_{exch}.

We next show that the Cornfield's approach and the Donner et al's approach are identical. Let P_1, P_0 be the expected prevalence of outcome in the intervention and control groups, both Cornfield's approach and the Donner et al's approach defines the effect size on the risk difference scale, namely $\Delta = P_1 - P_0$. Using our notation, the Donner et al's approach specifies the variance parameter as

$$\sigma^2_{\text{Donner}} = \frac{2\{1+(m-1)\rho\}}{m}\{(P_1(1-P_1) + P_0(1-P_0)\}.$$

This is adapted from equation (7) of Donner et al.[2] and equation (5) of Rutterford et al.[9] Under variable cluster sizes, one may simply replace m with $\bar{m}$ and then use the so-called "average cluster size method". The variance parameter used in the Cornfield's approach is not immediate from their original article,[1] and requires the following steps to obtain its explicit form. We first define $\hat{p}_i = \sum_{j=1}^{m} Y_{ij}/m$ (recall that their approach only assumes equal cluster sizes, and hence we use m). Then among the intervention clusters,

$$\text{var}(\hat{p}_i | X_i = 1) = \frac{1+(m-1)\rho}{m}\{P_1(1-P_1)\}.$$

Likewise, among the control clusters, we have

$$\text{var}(\hat{p}_i | X_i = 0) = \frac{1+(m-1)\rho}{m}\{P_0(1-P_0)\}.$$

Then using equations (3) and (4) in Cornfield,[1] we obtain the variance parameter

$$\begin{aligned}\sigma^2_{\text{Cornfield}} =& m \times \text{var}(\hat{p}_i) \times \left(\frac{2}{m} + \frac{2}{m}\right) = 4\left\{\frac{1}{2}\text{var}(\hat{p}_i|X_i=1) + \frac{1}{2}\text{var}(\hat{p}_i|X_i=0)\right\} \\ =& \frac{2\{1+(m-1)\rho\}}{m}\{(P_1(1-P_1) + P_0(1-P_0)\} = \sigma^2_{\text{Donner}} \qquad (1)\end{aligned}$$

Equation (1) clarifies that the Cornfield's approach and the Donner et al's approach are identical, whose variance parameter clearly differs from the σ^2 derived in our paper, even under equal cluster sizes.

We next provide some numerical evidence for comparing our approach with the Cornfield/Donner formula. Web Table 15 presents the estimated number of clusters using our formula with that using the Cornfield/Donner formula when the average cluster size is $\bar{m} \in \{50, 100\}$, and $P_0 = 0.15$, $P_1 = 0.3$, across a range of ICC and CV of cluster sizes used in our simulation study. It is evident that the estimated number of clusters tends to be smaller using the Cornfield/Donner formula compared those using our proposed formula. Only when the required number of clusters is fairly small (say fewer than 10), the results from all approaches agree with one another. In other cases, because the Cornfield/Donner formula returns a smaller number of clusters than our formula, CRTs designed with the Cornfield/Donner formula but analyzed by the modified Poisson regression (based on relative risks) could be underpowered (this is because we have shown in the simulations that our formula guarantees adequate empirical power for the GEE analyses). Similar conclusions can be made from Web Table 16, where we increase the prevalence of outcomes to be $P_0 = 0.3$ and $P_1 = 0.5$. Finally, because the Cornfield/Donner formula does not account for variable cluster sizes, when the actual CV of cluster sizes becomes large, the Cornfield/Donner formula mildly underestimates the required sample size for CRTs analyzed by the modified Poisson regression with an exchangeable working correlation structure, but severely underestimates the required sample size for CRTs analyzed by the modified Poisson regression with an independence working correlation structure.

Web Table 15. Estimated number of clusters using the proposed sample size formulas (under either independence or exchangeable correlation) and those using the Cornfield/Donner formula. We assume $P_0 = 0.15$, $P_1 = 0.3$. The "average cluster size method" was used in applying the Cornfield/Donner formula under variable cluster sizes (CV $\neq 0$).

		$\bar{m} = 50$			$\bar{m} = 100$		
ρ	CV	$\hat{n}_{\text{ind}}$	$\hat{n}_{\text{exch}}$	$\hat{n}_{\text{CD}}$	$\hat{n}_{\text{ind}}$	$\hat{n}_{\text{exch}}$	$\hat{n}_{\text{CD}}$
0.01	0.0	11	11	10	8	8	8
	0.2	11	11	10	8	8	8
	0.4	11	11	10	9	8	8
	0.6	12	11	10	9	9	8
	0.8	12	12	10	10	9	8
0.05	0.0	21	21	19	18	18	17
	0.2	21	21	19	19	18	17
	0.4	23	21	19	20	19	17
	0.6	25	22	19	23	19	17
	0.8	29	23	19	26	20	17
0.10	0.0	33	33	30	31	31	28
	0.2	34	34	30	32	31	28
	0.4	38	34	30	35	31	28
	0.6	43	35	30	40	32	28
	0.8	50	36	30	48	32	28
0.15	0.0	46	46	42	44	44	40
	0.2	48	46	42	46	44	40
	0.4	52	47	42	50	44	40
	0.6	60	48	42	58	45	40
	0.8	71	49	42	69	45	40
0.20	0.0	59	59	53	57	57	51
	0.2	61	59	53	59	57	51
	0.4	67	60	53	65	57	51
	0.6	78	60	53	76	58	51
	0.8	92	62	53	90	58	51

Notes 1: ρ refers to ICC; CV refers to the coefficient of variation of cluster sizes.

Notes 2: $\hat{n}_{\text{ind}}$ refers to the estimated number of clusters based on the modified Poisson regression with independence working correlation; $\hat{n}_{\text{exch}}$ refers to the estimated number of clusters based on the modified Poisson regression with exchangeable working correlation; $\hat{n}_{\text{CD}}$ refers to the estimated number of clusters based on the Cornfield/Donner formula (where the effect size is measured by risk difference).

Web Table 16. Estimated number of clusters using the proposed sample size formulas (under either independence or exchangeable correlation) and those using the Cornfield/Donner formula. We assume $P_0 = 0.3$, $P_1 = 0.5$. The "average cluster size method" was used in applying the Cornfield/Donner formula under variable cluster sizes (CV $\neq 0$).

		$\bar{m} = 50$			$\bar{m} = 100$		
ρ	CV	$\hat{n}_{\text{ind}}$	$\hat{n}_{\text{exch}}$	$\hat{n}_{\text{CD}}$	$\hat{n}_{\text{ind}}$	$\hat{n}_{\text{exch}}$	$\hat{n}_{\text{CD}}$
0.01	0.0	9	9	8	7	7	7
	0.2	9	9	8	7	7	7
	0.4	9	9	8	7	7	7
	0.6	10	9	8	8	7	7
	0.8	10	10	8	8	8	7
0.05	0.0	17	17	15	15	15	13
	0.2	17	17	15	15	15	13
	0.4	18	17	15	16	15	13
	0.6	20	18	15	18	15	13
	0.8	23	19	15	21	16	13
0.10	0.0	26	26	24	24	24	22
	0.2	27	26	24	25	25	22
	0.4	29	27	24	28	25	22
	0.6	33	27	24	32	25	22
	0.8	39	28	24	37	26	22
0.15	0.0	36	36	33	34	34	31
	0.2	37	36	33	36	34	31
	0.4	41	37	33	39	35	31
	0.6	47	37	33	45	35	31
	0.8	55	38	33	54	35	31
0.20	0.0	46	46	42	44	44	40
	0.2	47	46	42	46	44	40
	0.4	52	46	42	51	44	40
	0.6	60	47	42	59	45	40
	0.8	71	48	42	70	45	40

Notes 1: ρ refers to ICC; CV refers to the coefficient of variation of cluster sizes.

Notes 2: $\hat{n}_{\text{ind}}$ refers to the estimated number of clusters based on the modified Poisson regression with independence working correlation; $\hat{n}_{\text{exch}}$ refers to the estimated number of clusters based on the modified Poisson regression with exchangeable working correlation; $\hat{n}_{\text{CD}}$ refers to the estimated number of clusters based on the Cornfield/Donner formula (where the effect size is measured by risk difference).

References

1. Cornfield J. Symposium on chd prevention trials: design issues in testing life style intervention: randomization by group: a formal analysis. *American Journal of Epidemiology* 1978; 108(2): 100–102.
2. Donner A, Birkett N and Buck C. Randomization by cluster: sample size requirements and analysis. *American Journal of Epidemiology* 1981; 114(6): 906–914.
3. Li P and Redden DT. Small sample performance of bias-corrected sandwich estimators for cluster-randomized trials with binary outcomes. *Statistics in Medicine* 2015; 34(2): 281–296.
4. Ford WP and Westgate PM. Improved standard error estimator for maintaining the validity of inference in cluster randomized trials with a small number of clusters. *Biometrical Journal* 2017; 59(3): 478–495.
5. Li F, Turner EL and Preisser JS. Sample size determination for GEE analyses of stepped wedge cluster randomized trials. *Biometrics* 2018; 74(4): 1450–1458.
6. Li F, Forbes AB, Turner EL et al. Power and sample size requirements for GEE analyses of cluster randomized crossover trials. *Statistics in Medicine* 2019; 38(4): 636–649.
7. Li F. Design and analysis considerations for cohort stepped wedge cluster randomized trials with a decay correlation structure. *Statistics in Medicine* 2020; 39(4): 438–455.
8. Li F and Harhay MO. Commentary: Right truncation in cluster randomized trials can attenuate the power of a marginal analysis. *International Journal of Epidemiology* 2020; 49(3): 964–967.
9. Rutterford C, Copas A and Eldridge S. Methods for sample size determination in cluster randomized trials. *International Journal of Epidemiology* 2015; 44(3): 1051–1067.