

# PREVALENCE AND TREND ESTIMATION FROM OBSERVATIONAL DATA WITH HIGHLY VARIABLE POST-STRATIFICATION WEIGHTS

## SUPPLEMENTARY MATERIALS

### ADDITIONAL DETAILS AND RESULTS

BY YANNICK VANDENDIJK\*, CHRISTEL FAES\* AND NIEL HENS\*,†

*Interuniversity Institute for Biostatistics and statistical Bioinformatics,  
Hasselt University, Diepenbeek, Belgium \**

*Centre for Health Economic Research and Modelling Infectious Diseases,  
Vaccine and Infectious Disease Institute, University of Antwerp, Wilrijk,  
Belgium †*

### CONTENTS

Appendix A . . . . .	2
Appendix B . . . . .	5
Appendix C . . . . .	7
Appendix D . . . . .	13
Appendix E . . . . .	49
References . . . . .	75

Appendix A: Details on casting models (1) and (7) in the generalized linear mixed model (GLMM) framework.

Appendix B: Details on the model fitting of models (1) and (7).

Appendix C: Annotated SAS and R code to implement the proposed methods in statistical software.

Appendix D: Additional simulation results of the simulation study presented in Section 5 are given.

Appendix E: Additional results with respect to varying values of  $w_0$ , additional results for a small sample size and results on model fits and other spline basis functions.

**Appendix A.** In this Appendix we show how models (1) and (7) can be cast in the general linear mixed model (GLMM) framework.

The general form of a generalized linear mixed model is ([Molenberghs and Verbeke, 2005](#))

$$(A.1) \quad g\{E(\mathbf{y}|\mathbf{b})\} = \boldsymbol{\eta} \equiv \mathbf{N}\mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{Z}\mathbf{b},$$

where  $\mathbf{N}$  is an  $n \times H$  matrix indicating to which stratum an observation belongs ( $(\mathbf{N})_{ih} = 1$  if  $y_i$  belongs to stratum  $h$  and 0 otherwise),  $\mathbf{X}$  is a  $H \times p$  fixed-effect design matrix,  $\boldsymbol{\beta}$  is a  $p \times 1$  vector of unknown fixed-effect parameters,  $\mathbf{Z}$  is a  $H \times q$  random-effect design matrix and  $\mathbf{b}$  is a  $q \times 1$  vector of unknown random effect for which it is assumed that  $\mathbf{b} \sim \mathcal{N}_q(\mathbf{0}, \mathbf{G})$ . Formulating model (1) using model assumptions (a)-(c) in the GLMM framework yields the following replacements:

**(a) Exchangeable random-effects model (XRE):** Because it is assumed that  $\delta_h = \beta$  for all  $h$  and  $\mathbf{D} = \sigma^2 \mathbf{I}_H$ , we have

$$(A.2) \quad \mathbf{X} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}, \quad \mathbf{Z} = \mathbf{I}_H \text{ and } \mathbf{G} = \sigma^2 \mathbf{I}_H.$$

**(b) Linear model (LIN):** Because it is assumed that  $\delta_h = \beta_0 + \beta_1 x_h$  for all  $h$  and  $\mathbf{D} = \sigma^2 \mathbf{I}_H$ , we have

$$(A.3) \quad \mathbf{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_H \end{pmatrix}, \quad \mathbf{Z} = \mathbf{I}_H \text{ and } \mathbf{G} = \sigma^2 \mathbf{I}_H.$$

**(c) Non-parametric model (NPAR):** We have that  $\delta_h = f(x_h)$  for all  $h$  and  $\mathbf{D} = \sigma^2 \mathbf{I}_H$ , where  $f(x_h)$  is a twice differentiable smooth function of  $h$ ,

$$\left\{ f : f \text{ and } f' \text{ are absolutely continuous, } \int f''(u)^2 du < \infty \right\}$$

([Hastie and Tibshirani, 1990](#); [Ruppert, Wand and Carroll, 2003](#)). We use the approximating thin plate spline family to define  $f(x_h)$ . Let  $\kappa_1 = x_1, \dots, \kappa_H = x_H$  denote  $H$  fixed knots. Let  $d_{hh'} = \|x_h - \kappa_{h'}\| = |x_h - \kappa_{h'}|$  denote the Euclidean distance between  $x_h$  and knot  $\kappa_{h'}$ . The approximating thin plate spline function of degree 2 is

$$f(x_h) = \beta_0 + \beta_1 x_h + \sum_{h'=1}^H b_{uh'} d_{hh'}^3,$$

where  $(b_{u_1}, \dots, b_{u_H})$  are coefficients to be estimated which allow the non-parametric mean function  $f$  to vary smoothly (Ruppert, Wand and Carroll, 2003). These coefficients  $(b_{u_1}, \dots, b_{u_H})$  are penalized by a roughness penalty parameter to overcome overfitting of the data. Formulating this model in the GLMM framework yields

$$(A.4) \quad \mathbf{X} = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_H \end{pmatrix}, \mathbf{Z}_1 = \tilde{\mathbf{Z}}_1 \boldsymbol{\Omega}^{-1/2}, \mathbf{Z}_2 = \mathbf{I}_H, \mathbf{Z} = [\mathbf{Z}_1 | \mathbf{Z}_2]$$

and

$$(A.5) \quad \mathbf{G} = \begin{pmatrix} \sigma_u^2 \mathbf{I}_H & \mathbf{0} \\ \mathbf{0} & \sigma^2 \mathbf{I}_H \end{pmatrix},$$

where  $\tilde{\mathbf{Z}}_1$  is a  $H \times H$  matrix with  $(h, h')$ th element equal to  $d_{hh'}^3$  and  $\boldsymbol{\Omega}$  is a  $H \times H$  matrix with  $(h, h')$ th element equal to  $v_{hh'}^3$ , where  $v_{hh'} = \|\kappa_h - \kappa_{h'}\| = |\kappa_h - \kappa_{h'}|$  denotes the Euclidean distance between knots  $\kappa_h$  and  $\kappa_{h'}$ . The matrix  $\boldsymbol{\Omega}^{-1/2}$  can be calculated by singular value decomposition (Searle, 1982). For this NPAR model the random effects vector  $\mathbf{b}$  can be written as  $\mathbf{b} = [\mathbf{b}_u^T | \mathbf{b}_h^T]^T$ , where  $\mathbf{b}_u$  correspond to the random effects of the smooth non-parametric function, and  $\mathbf{b}_h$  correspond to the post-stratum random effects. Casting the NPAR model in the GLMM framework has the advantage that the roughness penalty parameter is selected automatically.

To cast model (7) with a non-parametric (NPAR) model assumption in the GLMM framework, we need to define the matrices

$$(A.6) \quad \mathbf{X} = \begin{pmatrix} 1 & x_1 & t_1 \\ \vdots & \vdots & \vdots \\ 1 & x_H & t_1 \\ 1 & x_1 & t_2 \\ \vdots & \vdots & \vdots \\ 1 & x_H & t_2 \\ \vdots & \vdots & \vdots \\ 1 & x_1 & t_T \\ \vdots & \vdots & \vdots \\ 1 & x_H & t_T \end{pmatrix},$$

$$(A.7) \quad \mathbf{Z}_1 = \tilde{\mathbf{Z}}_1 \boldsymbol{\Omega}^{-1/2}, \mathbf{Z}_2 = \tilde{\mathbf{Z}}_t \boldsymbol{\Omega}_t^{-1/2}, \mathbf{Z}_3 = [\mathbf{I}_H, \dots, \mathbf{I}_H]^T, \mathbf{Z} = [\mathbf{Z}_1 | \mathbf{Z}_2 | \mathbf{Z}_3],$$

and

$$(A.8) \quad \mathbf{G} = \begin{pmatrix} \sigma_u^2 \mathbf{I}_H & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \sigma_t^2 \mathbf{I}_K & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \sigma^2 \mathbf{I}_H \end{pmatrix}.$$

The vector  $(t_1, \dots, t_T)$  contains the different time points on which the survey data were collected.  $\tilde{\mathbf{Z}}_1$  is a  $HT \times H$  matrix with  $(ht, h')$ th element equal to  $d_{ht,h'}^3$  where  $d_{ht,h'} = |x_h - \kappa_{h'}|$ . The  $H \times H$  matrix  $\Omega$  is similar as defined in (A.4). The matrix  $\tilde{\mathbf{Z}}_t$  is a  $HT \times K$  matrix with  $(hj, j')$ th element equal to  $d_{hj,j'}^3$  where  $d_{hj,j'} = |t_j - \zeta_{j'}|$  and  $\zeta_1, \dots, \zeta_K$  denote  $K$  fixed knots of the time variable. The matrix  $\Omega_t$  is a  $K \times K$  matrix with  $(k, k')$ th element equal to  $v_{kk'}^3$ , where  $v_{kk'} = |\zeta_k - \zeta_{k'}|$  denotes the Euclidean distance between knots  $\zeta_k$  and  $\zeta_{k'}$ .

**Appendix B.** In this Appendix we give more details on model estimation of models (1) and (7) once they have been cast into the GLMM (A.1).

Let  $\boldsymbol{\theta}$  denote the vector of unknown covariance parameters in model (A.1). The corresponding likelihood is

$$\begin{aligned}\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\theta}) &= p(\mathbf{y}; \boldsymbol{\beta}, \boldsymbol{\theta}) \\ &= \int_{\mathbb{R}^q} p(\mathbf{y}|\mathbf{b})p(\mathbf{b})d\mathbf{b} \\ &= (2\pi)^{-q/2} |\mathbf{G}_{\boldsymbol{\theta}}|^{-1/2} \int_{\mathbb{R}^q} \exp\{\mathbf{y}^T \boldsymbol{\eta} - \mathbf{1}^T \log(1 + \exp(\boldsymbol{\eta})) - \frac{1}{2} \mathbf{b}^T \mathbf{G}_{\boldsymbol{\theta}}^{-1} \mathbf{b}\} d\mathbf{b}.\end{aligned}$$

Maximum or restricted maximum likelihood estimation is hindered by the presence of this  $q$ -dimensional integral. An exact analytical function of this integral is not available for binary data. We use pseudo-likelihood estimation based on linearisation to estimate model parameters ([Wolfinger and O'Connell, 1993](#)). This method employs Taylor series expansions to approximate the model by a linear mixed model based on pseudo-data. This fitting method is doubly iterative. First, model (A.1) is approximated by a linear mixed model based on current values of the parameter estimates. Second, the resulting linear mixed model is fit by restricted maximum likelihood, which is itself an iterative process. On convergence, the new parameter estimates are used to construct a new linear mixed model approximation of model (A.1). This process continues until a prespecified convergence criterion is met.

From (A.1) it follows that

$$(B.1) \quad E(\mathbf{y}|\mathbf{b}) = g^{-1}(\mathbf{N}\mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{Z}\mathbf{b}) = g^{-1}(\boldsymbol{\eta}) = \boldsymbol{\mu}.$$

Define  $\text{Var}(\mathbf{y}|\mathbf{b}) = \mathbf{A}$ , where  $\mathbf{A}$  is a  $n \times n$  diagonal matrix containing the variance components of the data. Let  $\tilde{\boldsymbol{\beta}}$  and  $\tilde{\mathbf{b}}$  denote the current estimates of  $\boldsymbol{\beta}$  and  $\mathbf{b}$ . A first order Taylor series expansion of  $\boldsymbol{\mu}$  around  $\tilde{\boldsymbol{\beta}}$  and  $\tilde{\mathbf{b}}$  yields

$$(B.2) \quad \boldsymbol{\mu} \approx g^{-1}(\tilde{\boldsymbol{\eta}}) + \tilde{\boldsymbol{\Delta}}\mathbf{N}\mathbf{X}(\boldsymbol{\beta} - \tilde{\boldsymbol{\beta}}) + \tilde{\boldsymbol{\Delta}}\mathbf{N}\mathbf{Z}(\mathbf{b} - \tilde{\mathbf{b}}),$$

where  $\tilde{\boldsymbol{\Delta}} = \left(\frac{\partial g^{-1}(\boldsymbol{\eta})}{\partial \boldsymbol{\eta}}\right)_{\tilde{\boldsymbol{\beta}}, \tilde{\mathbf{b}}}$  is a  $n \times n$  diagonal matrix of derivatives of the conditional mean evaluated at the expansion locus. Rearranging terms in (B.2) yields

$$(B.3) \quad \tilde{\boldsymbol{\Delta}}^{-1} (\boldsymbol{\mu} - g^{-1}(\tilde{\boldsymbol{\eta}})) + \mathbf{N}\mathbf{X}\tilde{\boldsymbol{\beta}} + \mathbf{N}\mathbf{Z}\tilde{\mathbf{b}} = \mathbf{N}\mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{Z}\mathbf{b}.$$

The left hand-side of (B.3) is, conditional on  $\mathbf{b}$ , the expected value of

$$(B.4) \quad \tilde{\boldsymbol{\Delta}}^{-1} (\mathbf{y} - g^{-1}(\tilde{\boldsymbol{\eta}})) + \mathbf{N}\mathbf{X}\tilde{\boldsymbol{\beta}} + \mathbf{N}\mathbf{Z}\tilde{\mathbf{b}} \equiv \mathbf{P},$$

where  $\text{Var}(\mathbf{P}|\mathbf{b}) = \tilde{\Delta}^{-1} \mathbf{A} \tilde{\Delta}^{-1}$ . One can thus consider the model

$$(B.5) \quad \mathbf{P} = \mathbf{N}\mathbf{X}\boldsymbol{\beta} + \mathbf{N}\mathbf{Z}\mathbf{b} + \boldsymbol{\epsilon},$$

which is a linear mixed model with pseudo-responses  $\mathbf{P}$  and  $\text{Var}(\boldsymbol{\epsilon}) = \text{Var}(\mathbf{P}|\mathbf{b})$ . Assuming that  $\boldsymbol{\epsilon}$  has a normal distribution, model (B.5) is fit by standard restricted maximum likelihood techniques for linear mixed models. This process of linearising model (B.1) and fitting the accompanying linear mixed model is iterated until a convergence criterion is met. Let  $\hat{\boldsymbol{\theta}}$  denote the final estimates of the covariance parameters. The fixed effects parameters are estimated and the random effects are predicted as

$$(B.6) \quad \hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{V}(\hat{\boldsymbol{\theta}})^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}(\hat{\boldsymbol{\theta}})^{-1} \bar{\mathbf{P}},$$

and

$$(B.7) \quad \hat{\mathbf{b}} = \hat{\mathbf{G}} \mathbf{Z}^T \mathbf{V}(\hat{\boldsymbol{\theta}})^{-1} (\bar{\mathbf{P}} - \mathbf{X} \hat{\boldsymbol{\beta}}),$$

where  $\mathbf{V}(\hat{\boldsymbol{\theta}}) = \mathbf{Z} \hat{\mathbf{G}} \mathbf{Z}^T + \Delta_*^{-1} \mathbf{A}_* \Delta_*^{-1} \Gamma$  is  $H \times H$  matrix with  $\mathbf{A}_*$  and  $\Delta_*$  both  $H \times H$  matrices similar as  $\mathbf{A}$  and  $\tilde{\Delta}$  but only for the  $H$  post-strata means, namely  $g^{-1}(\mathbf{X} \tilde{\boldsymbol{\beta}} + \mathbf{Z} \tilde{\mathbf{b}})$ .  $\Gamma$  is a  $H \times H$  diagonal matrix where  $(\Gamma)_{hh} = n_h^{-1}$  and  $\bar{\mathbf{P}} = \Delta_*^{-1}(\bar{\mathbf{y}} - g^{-1}(\mathbf{X} \tilde{\boldsymbol{\beta}} + \mathbf{Z} \tilde{\mathbf{b}})) + \mathbf{X} \tilde{\boldsymbol{\beta}} + \mathbf{Z} \tilde{\mathbf{b}}$  where  $\bar{\mathbf{y}} = (\bar{y}_1, \dots, \bar{y}_H)^T$ .

Equations (B.6) and (B.7) can be written as

$$(B.8) \quad \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{b}} \end{pmatrix} = (\mathbf{C}^T \hat{\Sigma}_p^{-1} \mathbf{C} + \hat{\mathbf{B}})^{-1} \mathbf{C}^T \hat{\Sigma}_p^{-1} \bar{\mathbf{P}},$$

where  $\mathbf{C} = [\mathbf{X} | \mathbf{Z}]$ ,  $\mathbf{B} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^{-1} \end{pmatrix}$ , and  $\Sigma_p = \Delta_*^{-1} \mathbf{A}_* \Delta_*^{-1} \Gamma$  ([Ruppert, Wand and Carroll, 2003](#)). The fitted post-stratum means are

$$(B.9) \quad (\hat{\mu}_1, \dots, \hat{\mu}_H)^T = \text{expit}(\mathbf{X} \hat{\boldsymbol{\beta}} + \mathbf{Z} \hat{\mathbf{b}}).$$

**Appendix C.** In this Appendix we give annotated SAS and R code for the fitting of the models discussed. We provide code for the point and variance estimation of  $\bar{y}_{ws}$  and  $\bar{y}_{ws,GREG}$ . We focus on the code for fitting model (1) using the LIN assumption. First, a dataset is simulated in R. This dataset is exported to SAS where the parameters of model (1) are estimated using the *GLIMMIX* procedure. These results are again exported to R where the point and variance estimation is done.

First, a dataset is simulated using the following R code.

```
### Input parameters
sigma.sq = 0.02; beta0 = -2; beta1 = 0.2
w0 = 3
seed = 4321

### Define population and sample size
Nh = c(7500, 7500, 8000, 8000, 8500, 8500, 9000, 9000, 9000,
      9000, 9000, 8500, 8500, 8000, 8000, 7500, 7500)
nh = c(5, 15, 30, 75, 125, 150, 200, 275, 375,
      375, 275, 200, 150, 100, 80, 40, 20, 10)
N = sum(Nh); n = sum(nh)
Ph = Nh/N; ph = nh/n; wh = Ph/ph
Xh = 1:18

### Simulate data according to model (1) using LIN assumption
set.seed(seed)
D = sigma.sq*diag(length(Xh))
delta.h = beta0 + beta1*Xh
delta.star.h = rmvnorm(1, mean = delta.h, sigma=D)
true.prop.h = expit(delta.star.h)
Y.ih = rbinom(length(Xh), Nh, true.prop.h)
population.Yh = Y.ih/Nh
population.prevalence = sum(population.Yh*Nh)/N
y.ih = rbinom(length(Xh), nh, population.Yh)
simulated.data = data.frame(cbind(y.ih, nh, Xh))
names(simulated.data) = c("resp", "nh", "xh")

### Export to SAS to estimate model
write.table(simulated.data, '...', row.names=F)
```

Note that the final dataset does not contain the individual responses from the subjects, but contains the data grouped per post-stratum. The following SAS code is used to estimate the model parameters of model (1) using the LIN assumption for this simulated dataset. The dataset is first imported in SAS:

```
proc import datafile='...' out=dataInput dbms=dlm
    replace; datarow=2;
run;
data dataInput; set dataInput;
    hcopy = xh;
run;
```

Here, the code for fitting model (1) using the LIN assumption is provided.

```
proc glimmix data=dataInput method=RSPL;
    parms (0.02);
    class hcopy;
    model resp/nh = xh /solution dist=binomial;
    random intercept /subject=hcopy solution;
    output out=DataOutput1 predicted=eta predicted(ILINK)=prob variance=var;
    ods output CovParms=DataOutput2;
run;
```

The 'method=RSPL' option requests estimation to be based on pseudo-restricted maximum likelihood. The PARMS statement defines the starting value of the  $\sigma^2$  parameter. In the MODEL statement no intercept should be specified as this is done automatically in SAS. The RANDOM statement is used to define the post-stratum specific parameters  $b_h$ . The OUTPUT and ODS OUTPUT statements are used to obtain all relevant output for constructing point and variance estimates in the next stage. To fit a model using the XRE assumption, the MODEL statement needs to be changed to:

```
model resp/nh = /solution dist=binomial;
```

The SAS code for the NPAR assumption is

```
proc glimmix data=dataInput method=RSPL;
    parms (0.001) (0.02);
    class hcopy;
    model resp/nh = xh /solution dist=binomial;
    random xh /type=rsmooth solution knotmethod=equal(18) knotinfo;
    random intercept /subject=hcopy solution;
    output out=DataOutput1 predicted=eta predicted(ILINK)=prob variance=var;
```

```
ods output CovParms=DataOutput2;
run;
```

The first RANDOM statement fits the non-parametric part of the model. The 'type=rsmooth' option specifies the usage of the approximating thin plate spline family for constructing the non-parametric function.

After model fitting, the two output datasets, *DataOutput1* and *DataOutput2*, need to be exported by a PROC EXPORT statement.

The next step it to calculate the point estimates. For this, we first import the two output datasets, *DataOutput1* and *DataOutput2*, into R.

```
import1 = read.table('.../DataOutput1', header=T)
import2 = read.table('.../DataOutput2', header=T)
```

The point estimates,  $\bar{y}_{ws}$  and  $\bar{y}_{ws,GREG}$ , are calculated with the following code.

```
nh.hat = rep(0,length(Xh))
nh.hat[wh > w0] = (Nh[wh > w0]/N)*(n/w0)
gamma = (n-sum(nh.hat)) / sum(nh[wh <= w0])
nh.hat[wh <= w0] = gamma*nh[wh <= w0]
pi.hat = nh.hat/Nh
sample.prop = y.ih / nh
pred.prop = import1$prob
y.ws = 1/N * sum(nh*sample.prop + (Nh-nh)*pred.prop)
y.ws.GREG = 1/N * sum((nh/pi.hat)*sample.prop + (Nh-(nh/pi.hat))*pred.prop)
```

The variance of  $\bar{y}_{ws}$  can be calculated analytically using the following code.

```
X = cbind(rep(1,length(Xh)), Xh)
Z = diag(length(Xh))
C = cbind(X, Z)
G = import2$Estimate*diag(18)
B = matrix(0, length(Xh)+2, length(Xh)+2)
B[3:(length(Xh)+2), 3:(length(Xh)+2)] = solve(G)
eta = import1$eta
Delta = diag(as.vector(exp(eta)/((1+exp(eta))^2)))
A = diag(as.vector(expit(eta)*(1-expit(eta))))
Gamma = diag(1/nh)
Sigma.p = solve(Delta) %*% A %*% solve(Delta) %*% Gamma
Theta = Delta %*% C %*% solve(t(C) %*% solve(Sigma.p, C) + B) %*% t(C) %*% Delta
V1 = Nh-nh
```

```
var.analytical.y.ws = t(V1)/*%Theta/*%V1 / (N^2)
```

To calculate the bootstrap variance of  $\bar{y}_{ws}$ , first, bootstrap samples are generated in R. These bootstrap samples are analysed in SAS using the SAS code given above with an extra *by index* statement. These results are again analysed in R to calculate the bootstrap variance.

```
# generate bootstrap samples
D = import2$Estimate*diag(length(Xh))
V = Z/*%G/*%t(Z) + Sigma.p
pseudo.data = solve(Delta)/*%(sample.prop - import1$prob) + eta
beta.hat = solve( t(X)/*%solve(V,X) , t(X)/*%solve(V,pseudo.data) )
y.boot = vector(); y.pop.mean = vector()
B = 250
set.seed(seed)
for (b in 1:B){
  tilde.u = rmvnorm(1, mean=rep(0,length(Xh)), sigma=D)
  tilde.delta.star = X/*%beta.hat + Z/*%as.vector(tilde.u)
  population.b = rbinom(18,round(Nh),expit(tilde.delta.star))
  y.pop.mean[b] = sum(population.b)/sum(Nh)
  sample.b = rbinom(length(Xh), nh, (population.b / Nh))
  y.boot = c(y.boot, sample.b)
}
n.boot = rep(nh,times=B)
xh.boot = rep(Xh,times=B)
b.boot = rep(1:B,each=length(Xh))
bootstrap.data <- as.data.frame(cbind(y.boot,n.boot,xh.boot,b.boot))
names(bootstrap.data) = c("y","nh","xh","b")
write.table(bootstrap.data,'...',row.names=F)

### Analyse results bootstrap samples
import3 = read.table('...',header=T)
bootstrap.y.ws <- vector()
for (b in 1:B){
  set.b <- import3[import3$b==b,]
  ph.boot.fit <- set.b$prob
  ph.boot <- set.b$y/set.b$nh
  if (set.b$prob[1]==999){bootstrap.y.ws[b]=NA}
  else{bootstrap.y.ws[b]=(1/N)*(sum( nh*ph.boot ) +
    sum( (Nh-nh)*ph.boot.fit ))}
}
```

```
v.k = (bootstrap.y.ws - y.pop.mean)^2
var.y.ws.boot = mean(v.k, na.rm=T)
```

In the last part of the code there is a *if...else* statement to eliminate those bootstrap samples that did not converge (1 out of the 250 bootstrap samples did not converge in this example).

The jackknife variance of  $\bar{y}_{ws,GREG}$  is obtained using the following code.

```
# construct jackknife samples
G=250
s=n/G
y.jack = NULL
set.seed(seed)
for (h in 1:length(Xh)){
  yy=c(rep(1,round(y.ih[h])), rep(0,round(nh[h]-y.ih[h])))
  y.jack = c(y.jack, sample(yy))
}
group.ind = rep(1:G, times=ceiling(s))[1:n]
jack.data.full = cbind(y.jack, group.ind, rep(1:length(Xh), times=nh))
y.boot=n.boot=h.boot=b.boot=NULL
for (g in 1:G){
  set = jack.data.full[group.ind!=g,]
  y.boot = c(y.boot, as.vector(tapply(set[,1],set[,3],sum)))
  n.boot = c(n.boot, as.vector(tapply(set[,1],set[,3],length)))
  h.boot = c(h.boot, Xh)
  b.boot = c(b.boot, rep(g,length(Xh)))
}
jackknife.data <- as.data.frame(cbind(y.boot,n.boot,h.boot,b.boot))
names(jackknife.data) <- c("y","nh","xh","b")
write.table(jackknife.data,'...',row.names=F)

# Analyse results jackknife samples
import4 = read.table('...', header=T)
jackknife.y.ws.greg <- vector()
for (g in 1:G){
  set.g <- import4[import4$b==g,]
  ph.boot.fit <- set.g$prob
  ph.boot <- set.g$y/set.g$nh
  jackknife.y.ws.greg[g] <- (1/N)*(sum( (set.g$nh/pi.hat)*ph.boot ) +
    sum( (Nh-set.g$nh/pi.hat)*ph.boot.fit ))
```

```
}

v.k = (jackknife.y.ws.greg - mean(jackknife.y.ws.greg))^2
var.y.ws.greg = (G-1)/G*sum(v.k)
```

**Appendix D.** In this Appendix extra simulation results are presented. Setup of the simulation study is described in Section 5.1 and Section 5.3.

*Results for Overall Prevalence Estimation.* Figures 1–4 present the bias results of the nine estimators for the eight population models. When the underlying population has a constant mean (NULL and XRE) all estimators yield unbiased estimates. The *psm* estimator remains unbiased in all simulation settings. For the populations with more structure (LIN, QUAD and EXP) the *unw* and *trim* estimators are biased. For the LIN<sub>0</sub> and LIN<sub>1</sub>, we observe that the *xre* and *xre-greg* also show some bias. For the QUAD and EXP populations we observe that the GREG adjusted estimators have less bias than their non-adjusted counterparts. As expected, we observe that the width of the boxplots decreases with increasing sample size and vice versa.

The variance results are presented in Figures 5–8. In all simulation scenarios, the *psm* estimator has a high variance associated with its estimates. This is more pronounced for smaller sample sizes. For simple underlying population models (NULL and XRE) we observe that the variances of *npar* and *npar-greg* are higher than the other estimators (excluding *psm*). We also observe that the variances associated with the GREG adjusted estimators are higher than the variances associated with the non-adjusted estimators. For the largest sample size (Figure 5), the variances of the *psm* estimator and the NPAR based estimators do not differ much.

Tables 1–4 summarize the MSE values of all simulation settings. In general, *psm* does not perform well in terms of MSE, due to the increase in variability over the other estimators. For a population with more structure (LIN, QUAD and EXP) the *unw* and *trim* estimators show a high MSE, due to the bias of the estimates. For the LIN populations, as expected, the estimators *lin* and *lin-greg* perform the best in terms of MSE. The *npar* estimators perform the best for the QUAD populations. For the EXP populations, the *lin-greg* and both *npar* estimators perform well. For the smallest sample size (Table 4), we observe that the MSE values not differ much between the *trim* estimator and the weight smoothing model based estimators.

The nominal coverage results of the different estimators are presented in Tables 5–8. The *psm* estimator attains a good coverage over all simulation conditions. Except for the smaller sample size (Table 8) coverages are somewhat lower than the 95% level. Due to the bias, the nominal coverage of the *unw* and *trim* estimators deteriorates when the underlying model has more structure. For the underlying models with non-linear mean (QUAD and EXP), the *xre* and *lin* estimators do not obtain good coverage results. The coverage results of *npar* and *npar-greg* estimators are quite consistent,

but still yield some underestimates of the actual coverages, especially for the QUAD and EXP populations. Overall, the *npar-greg* estimator seems to attain better coverage results when compared to the *npar* estimator.

Tables 9–12 show the average lengths of the 95% confidence intervals. The average lengths of the GREG adjusted estimators are greater than the average lengths of the non-adjusted counterparts. We observe that the average lengths of the *npar* and *npar-greg* estimators are smaller when compared to *psm*. Only for the small sample size (Table 12) the average lengths of *npar-greg* estimator are equal or slightly higher than for *psm*. We also noted in Table 8 that the *npar-greg* yields better coverage results for the smaller sample size.

Overall the *npar* and *npar-greg* estimators perform consistent over the different simulation settings in terms of bias, variance and MSE. The *npar* and *npar-greg* estimators underestimate the actual coverage slightly, but the average length of the confidence intervals is in general smaller when compared to *psm*.

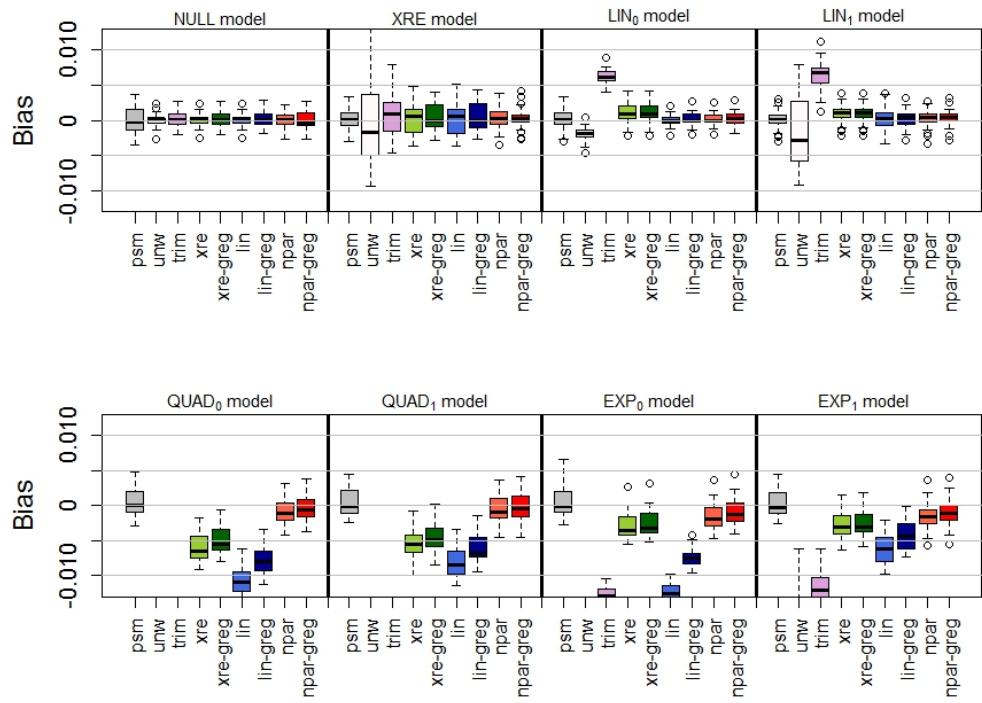


FIG 1. *Bias results of nine estimators and eight population models for the large population  $N^{(1)} = 6,000,000$  with a sample size of 25,000. Boxplots of the average bias associated with the 25 simulated populations are shown. The bias of some estimators exceeds the range of the y-axis used and are, therefore, not depicted on the figure.*

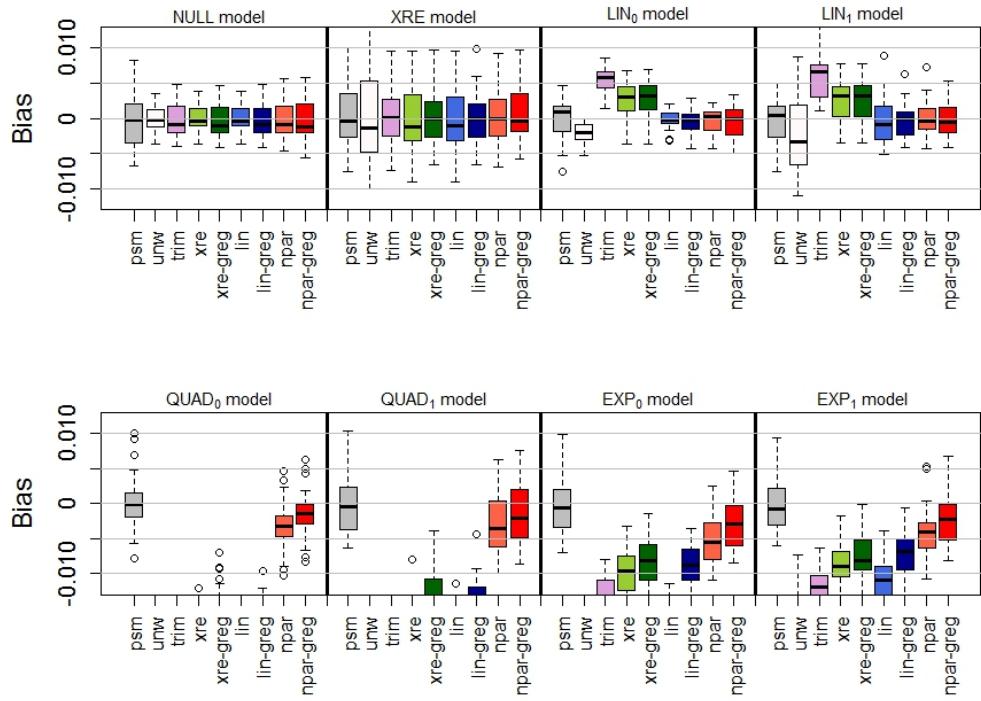


FIG 2. *Bias results of nine estimators and eight population models for the large population  $N^{(1)} = 6,000,000$  with a sample size of 5,000. Boxplots of the average bias associated with the 25 simulated populations are shown. The bias of some estimators exceeds the range of the y-axis used and are, therefore, not depicted on the figure.*

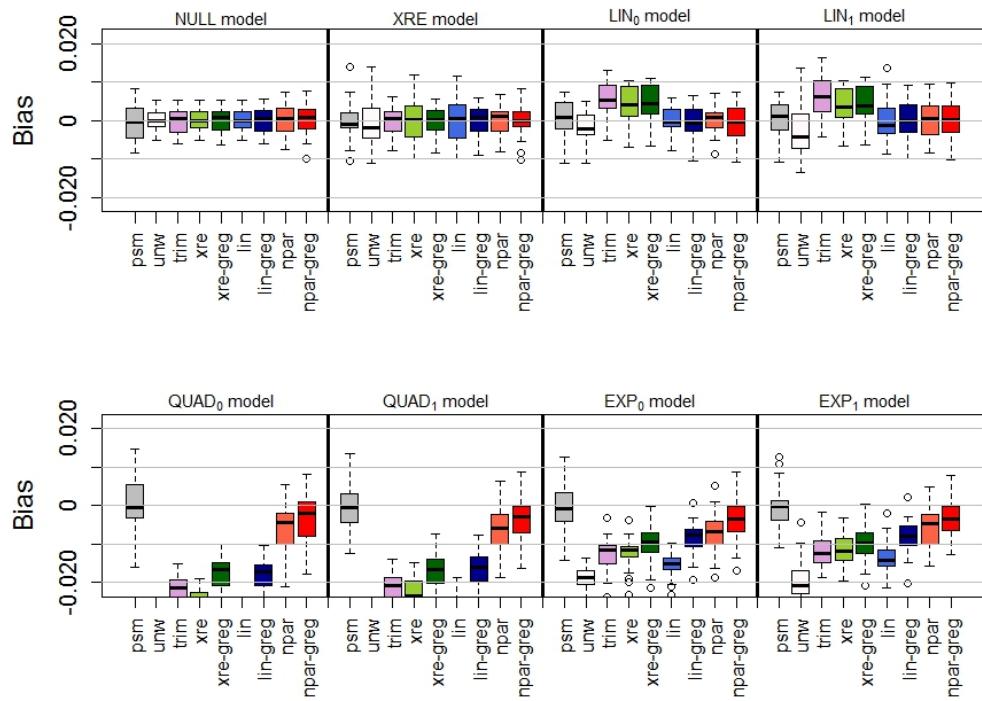


FIG 3. *Bias results of nine estimators and eight population models for the small population  $N^{(1)} = 150,000$  with a sample size of 2,500. Boxplots of the average bias associated with the 25 simulated populations are shown. The bias of some estimators exceeds the range of the y-axis used and are, therefore, not depicted on the figure.*

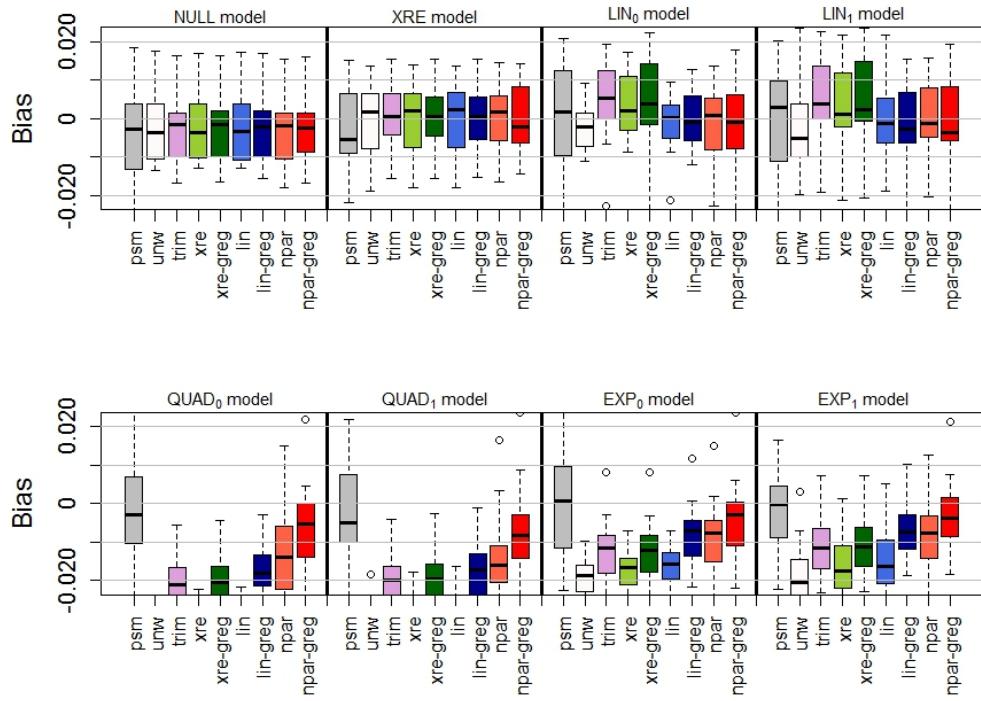


FIG 4. *Bias results of nine estimators and eight population models for the small population  $N^{(1)} = 150,000$  with a sample size of 500. Boxplots of the average bias associated with the 25 simulated populations are shown. The bias of some estimators exceeds the range of the y-axis used and are, therefore, not depicted on the figure.*

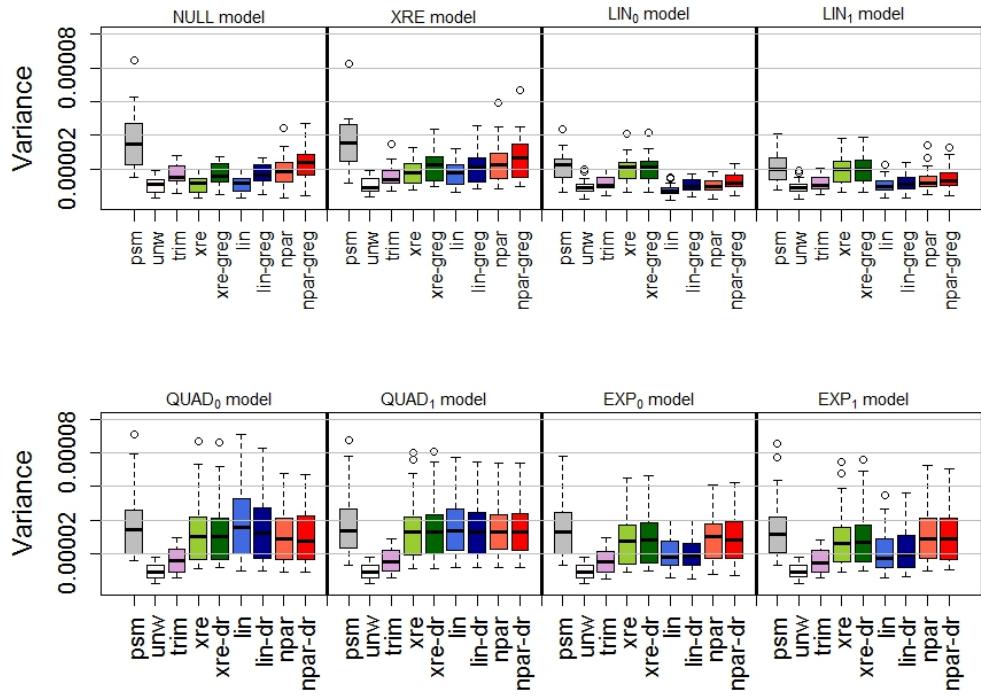


FIG 5. Variance results of nine estimators and eight population models for the large population  $N^{(1)} = 6,000,000$  with a sample size of 25,000. Boxplots of the average bias associated with the 25 simulated populations are shown. The variance of some estimators exceeds the range of the y-axis used and are, therefore, not depicted on the figure.

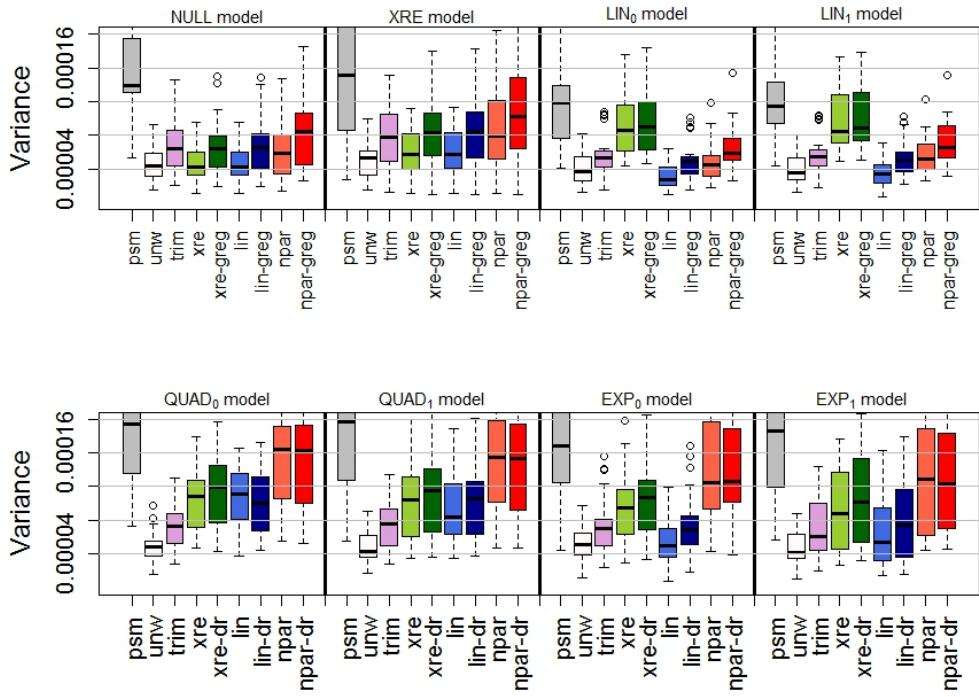


FIG 6. Variance results of nine estimators and eight population models for the large population  $N^{(1)} = 6,000,000$  with a sample size of 5,000. Boxplots of the average bias associated with the 25 simulated populations are shown. The variance of some estimators exceeds the range of the y-axis used and are, therefore, not depicted on the figure.

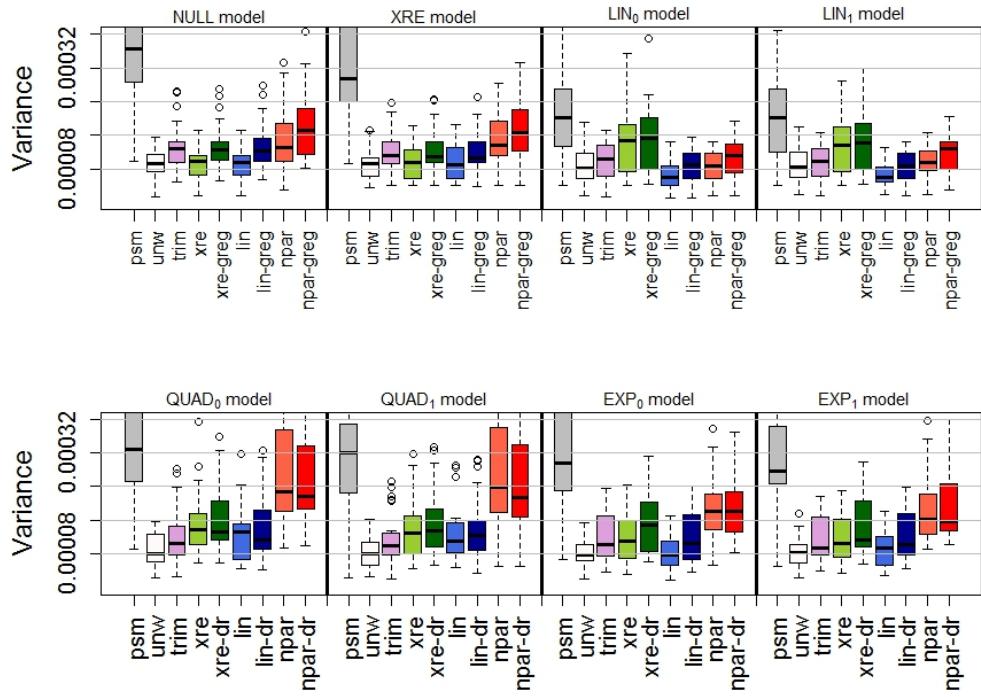


FIG 7. Variance results of nine estimators and eight population models for the small population  $N^{(1)} = 150,000$  with a sample size of 2,500. Boxplots of the average bias associated with the 25 simulated populations are shown. The variance of some estimators exceeds the range of the y-axis used and are, therefore, not depicted on the figure.

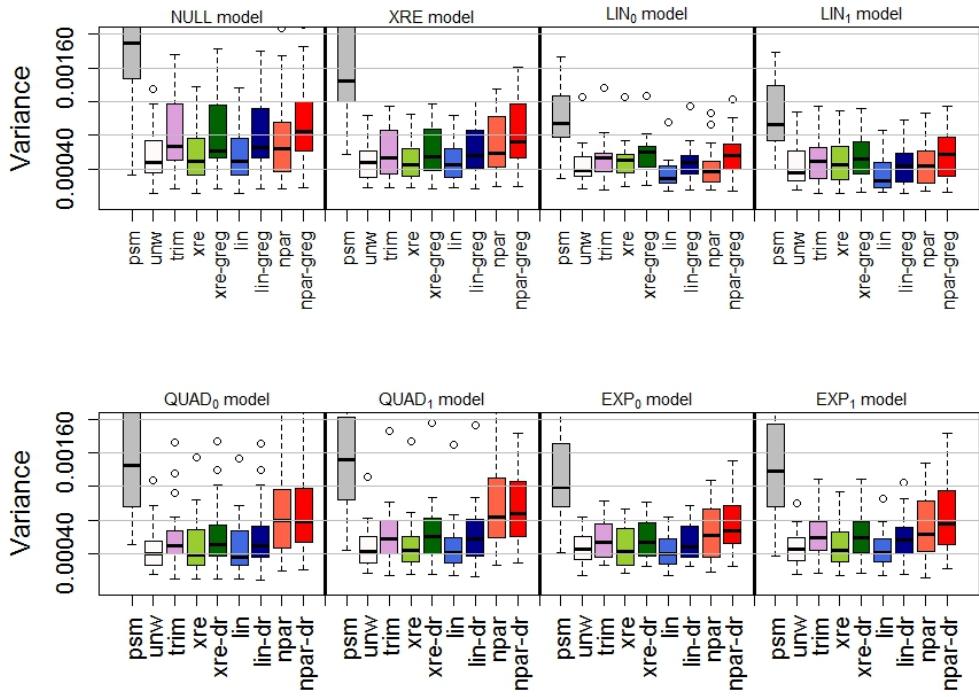


FIG 8. Variance results of nine estimators and eight population models for the small population  $N^{(1)} = 150,000$  with a sample size of 500. Boxplots of the average bias associated with the 25 simulated populations are shown. The variance of some estimators exceeds the range of the y-axis used and are, therefore, not depicted on the figure.

TABLE 1

*Mean squared error ( $\times 10^4$ ) of nine estimators and eight population models for the large population  $N^{(1)} = 6,000,000$  with a sample size of 25,000.*

Estimator	NULL	XRE	LIN <sub>0</sub>	LIN <sub>1</sub>	QUAD <sub>0</sub>	QUAD <sub>1</sub>	EXP <sub>0</sub>	EXP <sub>1</sub>
<i>psm</i>	0.38	1.32	0.21	0.75	0.38	1.31	0.36	1.26
<i>unw</i>	0.11	1.22	0.13	1.12	15.82	17.05	3.70	4.90
<i>trim</i>	0.17	1.12	0.50	1.26	5.16	5.84	1.80	2.59
<i>xre</i>	0.11	0.18	0.20	0.20	0.71	0.66	0.38	0.39
<i>xre-greg</i>	0.17	0.21	0.20	0.20	0.59	0.56	0.37	0.38
<i>lin</i>	0.11	0.18	0.07	0.10	1.55	1.02	1.72	0.60
<i>lin-greg</i>	0.17	0.21	0.10	0.12	0.97	0.71	0.76	0.42
<i>npar</i>	0.19	0.25	0.10	0.14	0.32	0.35	0.32	0.34
<i>npar-greg</i>	0.23	0.27	0.13	0.15	0.32	0.35	0.30	0.32

TABLE 2

*Mean squared error ( $\times 10^4$ ) of nine estimators and eight population models for the large population  $N^{(1)} = 6,000,000$  with a sample size of 5,000.*

Estimator	NULL	XRE	LIN <sub>0</sub>	LIN <sub>1</sub>	QUAD <sub>0</sub>	QUAD <sub>1</sub>	EXP <sub>0</sub>	EXP <sub>1</sub>
<i>psm</i>	1.70	2.85	1.16	1.73	1.88	2.92	1.82	2.90
<i>unw</i>	0.46	1.83	0.43	1.54	16.09	17.52	4.16	5.42
<i>trim</i>	0.68	1.92	0.86	1.68	5.61	6.50	2.41	3.26
<i>xre</i>	0.47	0.61	1.01	1.04	4.64	4.11	1.83	1.75
<i>xre-greg</i>	0.67	0.84	1.06	1.09	3.14	2.90	1.73	1.68
<i>lin</i>	0.48	0.61	0.30	0.35	8.35	5.59	2.97	2.04
<i>lin-greg</i>	0.67	0.83	0.50	0.53	3.66	2.96	1.49	1.32
<i>npar</i>	0.65	0.90	0.48	0.57	1.63	1.71	1.74	1.49
<i>npar-greg</i>	0.85	1.08	0.66	0.71	1.52	1.57	1.49	1.40

TABLE 3

*Mean squared error ( $\times 10^4$ ) of nine estimators and eight population models for the small population  $N^{(1)} = 150,000$  with a sample size of 2,500.*

Estimator	NULL	XRE	LIN <sub>0</sub>	LIN <sub>1</sub>	QUAD <sub>0</sub>	QUAD <sub>1</sub>	EXP <sub>0</sub>	EXP <sub>1</sub>
<i>psm</i>	3.81	4.50	2.21	2.57	3.64	4.58	3.41	4.28
<i>unw</i>	0.91	2.18	1.06	2.18	16.23	17.79	4.49	5.79
<i>trim</i>	1.33	2.25	1.46	2.20	6.10	7.07	2.97	3.76
<i>xre</i>	0.93	0.96	1.59	1.55	8.02	7.13	2.78	2.66
<i>xre-greg</i>	1.34	1.25	1.74	1.68	4.80	4.50	2.62	2.50
<i>lin</i>	0.93	0.96	0.69	0.71	11.24	8.79	3.26	2.82
<i>lin-greg</i>	1.33	1.26	0.87	0.88	4.66	4.23	1.98	1.88
<i>npar</i>	1.43	1.48	0.87	1.01	3.02	3.01	2.53	2.33
<i>npar-greg</i>	1.83	1.75	1.09	1.17	2.69	2.66	2.27	2.17

TABLE 4

*Mean squared error ( $\times 10^4$ ) of nine estimators and eight population models for the small population  $N^{(1)} = 150,000$  with a sample size of 500.*

Estimator	NULL	XRE	LIN <sub>0</sub>	LIN <sub>1</sub>	QUAD <sub>0</sub>	QUAD <sub>1</sub>	EXP <sub>0</sub>	EXP <sub>1</sub>
<i>psm</i>	19.45	17.18	11.28	11.15	16.85	18.86	14.72	16.84
<i>unw</i>	5.56	6.53	4.74	5.70	20.02	21.83	8.55	9.83
<i>trim</i>	7.78	7.62	5.96	6.48	11.04	12.50	7.86	8.82
<i>xre</i>	5.59	5.10	5.08	5.13	16.91	17.10	7.88	7.84
<i>xre-greg</i>	8.09	6.41	6.13	6.15	10.86	11.26	7.56	7.81
<i>lin</i>	5.57	5.05	3.34	3.41	16.33	16.42	7.11	7.01
<i>lin-greg</i>	8.02	6.43	4.60	4.57	9.53	9.96	6.40	6.72
<i>npar</i>	8.11	6.61	4.31	4.53	11.33	12.51	7.35	7.75
<i>npar-greg</i>	10.55	8.20	5.50	5.57	10.36	11.19	7.87	8.49

TABLE 5

*Nominal coverage of the 95% CI of nine estimators for the large population ( $N^{(1)} = 6,000,000$ ) with a sample size of 25,000.*

Estimator	NULL	XRE	LIN <sub>0</sub>	LIN <sub>1</sub>	QUAD <sub>0</sub>	QUAD <sub>1</sub>	EXP <sub>0</sub>	EXP <sub>1</sub>
<i>psm</i>	96.4	96.4	97.1	96.4	92.0	93.2	93.5	94.4
<i>unw</i>	94.3	64.0	90.2	58.8	0.0	0.0	0.0	1.6
<i>trim</i>	93.1	87.2	52.2	49.6	0.0	0.0	6.9	10.0
<i>xre</i> (analytical)	94.7	93.2	98.0	97.6	77.2	76.8	90.7	92.0
<i>xre</i> (bootstrap)	95.1	92.8	99.2	98.8	76.0	78.4	91.9	92.4
<i>xre-greg</i> (jackknife)	99.2	92.8	98.0	97.6	78.4	78.0	91.1	90.8
<i>lin</i> (analytical)	94.7	92.4	95.1	95.2	44.8	61.6	14.9	72.4
<i>lin</i> (bootstrap)	94.7	92.4	95.1	94.8	43.6	61.2	16.1	73.2
<i>lin-greg</i> (jackknife)	98.4	92.0	98.4	95.6	53.2	64.4	21.0	73.6
<i>npar</i> (analytical)	92.7	92.4	93.9	96.4	93.6	92.4	89.5	90.0
<i>npar</i> (bootstrap)	96.4	91.2	98.0	94.8	95.2	92.4	93.5	89.6
<i>npar-greg</i> (jackknife)	97.6	93.6	98.4	96.0	93.2	92.0	91.1	91.2

TABLE 6

*Nominal coverage of the 95% CI of nine estimators for the large population ( $N^{(1)} = 6,000,000$ ) with a sample size of 5,000.*

Estimator	NULL	XRE	LIN <sub>0</sub>	LIN <sub>1</sub>	QUAD <sub>0</sub>	QUAD <sub>1</sub>	EXP <sub>0</sub>	EXP <sub>1</sub>
<i>psm</i>	95.2	95.2	93.6	92.8	93.2	94.8	93.6	92.4
<i>unw</i>	96.8	87.2	95.6	88.8	0.0	0.0	22.0	26.4
<i>trim</i>	95.6	92.8	88.4	84.4	24.0	28.0	64.4	65.2
<i>xre</i> (analytical)	96.8	94.0	96.8	96.4	51.2	58.0	87.2	86.8
<i>xre</i> (bootstrap)	97.2	94.0	98.8	98.8	50.4	58.0	86.4	88.8
<i>xre-greg</i> (jackknife)	98.0	96.0	96.4	95.6	58.8	64.0	85.6	84.4
<i>lin</i> (analytical)	96.8	94.8	96.8	95.2	19.6	36.0	40.4	69.2
<i>lin</i> (bootstrap)	97.2	94.4	96.8	95.6	19.2	37.2	41.2	69.6
<i>lin-greg</i> (jackknife)	98.0	96.8	98.8	96.4	25.6	42.4	60.0	73.2
<i>npar</i> (analytical)	94.8	93.2	95.6	94.0	91.2	89.2	77.6	86.4
<i>npar</i> (bootstrap)	97.6	96.0	98.4	95.2	94.4	92.0	84.8	87.6
<i>npar-greg</i> (jackknife)	98.0	96.8	98.8	96.8	93.2	92.4	87.6	90.0

TABLE 7  
*Nominal coverage of the 95% CI of nine estimators for the small population  
 $(N^{(2)} = 150,000)$  with a sample size of 2,500.*

Estimator	NULL	XRE	LIN <sub>0</sub>	LIN <sub>1</sub>	QUAD <sub>0</sub>	QUAD <sub>1</sub>	EXP <sub>0</sub>	EXP <sub>1</sub>
<i>psm</i>	94.8	94.8	96.0	96.0	92.0	94.0	94.4	92.8
<i>unw</i>	96.4	91.6	93.6	86.8	1.6	4.0	53.6	52.4
<i>trim</i>	97.2	96.8	91.2	89.6	51.6	52.0	78.8	83.2
<i>xre</i> (analytical)	96.4	94.8	98.4	98.4	44.8	54.8	86.8	88.4
<i>xre</i> (bootstrap)	96.8	94.8	98.4	99.6	46.0	56.0	88.0	90.0
<i>xre-greg</i> (jackknife)	99.2	97.6	97.2	97.2	54.0	61.2	86.8	89.2
<i>lin</i> (analytical)	96.4	94.8	95.2	91.2	20.4	33.6	67.2	76.8
<i>lin</i> (bootstrap)	96.8	94.8	95.6	92.4	20.8	35.6	68.8	78.4
<i>lin-greg</i> (jackknife)	99.2	98.0	98.4	97.2	29.2	44.8	79.6	85.2
<i>npar</i> (analytical)	95.6	96.0	94.4	90.4	89.2	87.6	81.2	87.6
<i>npar</i> (bootstrap)	98.4	97.2	97.6	96.0	91.2	89.2	86.4	91.2
<i>npar-greg</i> (jackknife)	98.0	98.4	99.2	98.0	94.0	91.6	88.8	94.0

TABLE 8  
*Nominal coverage of the 95% CI of nine estimators for the small population  
 $(N^{(2)} = 150,000)$  with a sample size of 500.*

Estimator	NULL	XRE	LIN <sub>0</sub>	LIN <sub>1</sub>	QUAD <sub>0</sub>	QUAD <sub>1</sub>	EXP <sub>0</sub>	EXP <sub>1</sub>
<i>psm</i>	86.4	88.4	88.4	88.0	90.0	88.0	92.0	91.6
<i>unw</i>	94.4	94.0	95.2	94.4	53.2	55.6	84.7	84.0
<i>trim</i>	92.8	96.0	95.2	92.8	88.0	88.8	92.4	92.4
<i>xre</i> (analytical)	94.8	94.4	98.0	98.4	66.0	73.6	90.4	88.8
<i>xre</i> (bootstrap)	95.2	93.6	98.0	98.8	67.2	72.8	89.2	89.2
<i>xre-greg</i> (jackknife)	97.6	97.6	97.6	98.4	82.4	83.6	94.0	96.0
<i>lin</i> (analytical)	94.8	94.4	95.2	94.4	64.8	72.0	88.8	88.8
<i>lin</i> (bootstrap)	94.8	94.4	95.6	95.2	66.4	72.0	89.2	89.2
<i>lin-greg</i> (jackknife)	97.2	97.6	100	99.2	85.6	85.6	97.2	97.6
<i>npar</i> (analytical)	93.2	96.0	94.0	93.6	83.6	84.0	90.8	91.2
<i>npar</i> (bootstrap)	95.2	97.2	95.2	96.4	90.0	88.4	94.8	96.4
<i>npar-greg</i> (jackknife)	96.8	97.6	100	99.2	92.8	91.6	98.0	98.0

TABLE 9  
*Average length ( $\times 100$ ) of the 95% CI of nine estimators for the large population  
 $(N^{(1)} = 6,000,000)$  with a sample size of 25,000.*

Estimator	NULL	XRE	LIN <sub>0</sub>	LIN <sub>1</sub>	QUAD <sub>0</sub>	QUAD <sub>1</sub>	EXP <sub>0</sub>	EXP <sub>1</sub>
<i>psm</i>	2.4	2.4	1.8	1.8	2.3	2.3	2.2	2.2
<i>unw</i>	1.2	1.2	1.2	1.2	1.2	1.2	1.2	1.2
<i>trim</i>	1.5	1.5	1.3	1.3	1.4	1.4	1.4	1.4
<i>xre</i> (analytical)	1.3	1.8	1.9	1.9	2.1	2.1	2.1	2.1
<i>xre</i> (bootstrap)	1.3	1.8	2.1	2.1	2.1	2.1	2.2	2.2
<i>xre-greg</i> (jackknife)	1.5	1.9	1.8	1.8	2.4	2.5	2.1	2.1
<i>lin</i> (analytical)	1.3	1.8	1.1	1.4	1.9	2.0	1.4	1.8
<i>lin</i> (bootstrap)	1.3	1.8	1.1	1.4	1.9	2.0	1.4	1.8
<i>lin-greg</i> (jackknife)	1.5	1.7	1.3	1.4	2.2	2.1	1.7	1.8
<i>npar</i> (analytical)	1.4	1.8	1.2	1.4	2.1	2.1	1.9	2.0
<i>npar</i> (bootstrap)	1.6	1.7	1.3	1.4	2.1	2.1	2.2	2.0
<i>npar-greg</i> (jackknife)	1.8	1.9	1.4	1.6	2.1	2.1	2.0	2.1

TABLE 10  
*Average length ( $\times 100$ ) of the 95% CI of nine estimators for the large population  
 $(N^{(1)} = 6,000,000)$  with a sample size of 5,000.*

Estimator	NULL	XRE	LIN <sub>0</sub>	LIN <sub>1</sub>	QUAD <sub>0</sub>	QUAD <sub>1</sub>	EXP <sub>0</sub>	EXP <sub>1</sub>
<i>psm</i>	5.3	5.3	4.1	4.1	5.1	5.1	5.0	5.0
<i>unw</i>	2.8	2.8	2.7	2.7	2.7	2.7	2.7	2.7
<i>trim</i>	3.3	3.3	2.9	2.9	3.2	3.2	3.2	3.2
<i>xre</i> (analytical)	2.8	3.2	4.2	4.2	3.9	4.0	4.1	4.1
<i>xre</i> (bootstrap)	2.9	3.2	4.5	4.5	3.9	4.0	4.1	4.1
<i>xre-greg</i> (jackknife)	3.4	3.7	3.9	3.9	4.5	4.5	4.0	4.2
<i>lin</i> (analytical)	2.8	3.2	2.4	2.6	3.2	3.5	2.8	3.2
<i>lin</i> (bootstrap)	2.9	3.3	2.4	2.6	3.2	3.5	2.8	3.2
<i>lin-greg</i> (jackknife)	3.4	3.5	2.9	2.9	3.8	3.9	3.3	3.4
<i>npar</i> (analytical)	3.1	3.5	2.6	2.8	4.4	4.4	3.7	3.8
<i>npar</i> (bootstrap)	3.6	3.6	2.9	3.0	4.9	4.8	4.2	4.1
<i>npar-greg</i> (jackknife)	4.0	4.1	3.4	3.4	4.6	4.7	4.4	4.4

TABLE 11  
*Average length ( $\times 100$ ) of the 95% CI of nine estimators for the small population  
 $(N^{(1)} = 150,000)$  with a sample size of 2,500.*

Estimator	NULL	XRE	LIN <sub>0</sub>	LIN <sub>1</sub>	QUAD <sub>0</sub>	QUAD <sub>1</sub>	EXP <sub>0</sub>	EXP <sub>1</sub>
<i>psm</i>	7.4	7.4	5.8	5.8	7.2	7.2	7.1	7.1
<i>unw</i>	3.9	3.9	3.7	3.7	3.8	3.8	3.8	3.8
<i>trim</i>	4.6	4.6	4.1	4.1	4.5	4.5	4.5	4.5
<i>xre</i> (analytical)	3.9	4.2	5.8	5.8	4.8	5.0	5.1	5.2
<i>xre</i> (bootstrap)	4.1	4.4	6.1	6.1	4.9	5.0	5.2	5.3
<i>xre-greg</i> (jackknife)	4.8	5.0	5.3	5.3	5.4	5.6	5.2	5.4
<i>lin</i> (analytical)	3.9	4.2	3.3	3.5	4.1	4.4	3.8	4.1
<i>lin</i> (bootstrap)	4.1	4.3	3.4	3.5	4.2	4.5	3.9	4.2
<i>lin-greg</i> (jackknife)	4.8	4.9	4.1	4.1	4.9	5.1	4.7	4.8
<i>npar</i> (analytical)	4.3	4.7	3.6	3.8	5.9	5.9	4.7	4.9
<i>npar</i> (bootstrap)	5.1	5.2	4.1	4.2	6.7	6.6	5.4	5.5
<i>npar-greg</i> (jackknife)	5.8	5.8	4.6	4.7	6.5	6.5	5.9	6.0

TABLE 12  
*Average length ( $\times 100$ ) of the 95% CI of nine estimators for the small population  
 $(N^{(1)} = 150,000)$  with a sample size of 500.*

Estimator	NULL	XRE	LIN <sub>0</sub>	LIN <sub>1</sub>	QUAD <sub>0</sub>	QUAD <sub>1</sub>	EXP <sub>0</sub>	EXP <sub>1</sub>
<i>psm</i>	13.5	13.5	10.8	10.8	13.1	13.1	12.9	12.9
<i>unw</i>	8.7	8.7	8.4	8.3	8.5	8.5	8.6	8.6
<i>trim</i>	10.3	10.3	9.1	9.1	10.1	10.0	10.0	10.0
<i>xre</i> (analytical)	8.9	9.0	11.2	11.2	9.1	9.2	9.3	9.4
<i>xre</i> (bootstrap)	9.0	9.1	11.1	11.1	9.2	9.3	9.3	9.4
<i>xre-greg</i> (jackknife)	10.8	10.9	11.2	11.1	11.0	11.1	10.8	11.0
<i>lin</i> (analytical)	8.9	9.0	7.4	7.5	8.7	8.9	8.6	8.7
<i>lin</i> (bootstrap)	8.9	9.0	7.5	7.5	8.8	8.9	8.6	8.7
<i>lin-greg</i> (jackknife)	10.8	10.9	11.4	11.4	11.1	11.1	11.1	11.1
<i>npar</i> (analytical)	9.7	9.7	7.9	8.0	10.6	10.7	9.6	9.7
<i>npar</i> (bootstrap)	11.1	11.1	9.0	9.0	12.0	12.0	10.9	10.9
<i>npar-greg</i> (jackknife)	12.4	12.5	12.1	12.4	13.9	14.1	13.0	12.9

*Results for Mean Trend Estimation.* In this Section, extra simulation results are presented for the mean trend estimation. Setup of the simulation study is described in Section 5.3.

The bias results for all simulation conditions are presented in Figures 9–12. In these figures the bias is presented at each time point. The *psm* estimator yields, as expected, unbiased estimates. We clearly observe that the *unw* and *trim* estimators are biased for all simulation settings at all time points. For the underlying models F1 and F2, the *npar* estimator is unbiased. For F3 and F4, however, they are unbiased at the start and end of the time trend, but show some severe bias in the region of the sharp valley. This bias is more pronounced for decreasing sample sizes. The *npar-greg* estimator, on the contrary, remains unbiased for the underlying models F3 and F4. For the smallest sample size (Figure 12) we observe some small bias for the *npar* and *npar-greg* estimators for F1 and F2. From these results, it is thus clear that the GREG adjusted estimator is able to correct for the bias present in the predictions of the weight smoothing model.

Figures 13–16 present the variance of the estimators at each time point. The variance of *psm* is larger when compared to the other estimators. The *npar-greg* estimator exhibits more variability than *unw* and *npar*, and has a comparable variability as the *trim* estimator.

Combining the bias and variance, leads to the MSE results shown in Figures 17–20. For the largest sample size (Figure 17), the *unw* and *trim* estimators yield higher MSE results. For the other sample size settings, we observe that the *npar* and *npar-greg* have a better performance in terms of MSE when compared to the other methods. For all time points, the *npar* estimator has lower MSE values for F1 and F2 when compared with *npar-greg*. For the functions with the sharp valley, F3 and F4, the mean squared error of *npar* is smaller than the mean squared error of *npar-greg*, except at those time points where the sharp valley occurs. The MSE of the *npar* estimator shows a steep increase in that region. Thus, for cases such as scenarios F3 and F4 the *npar* estimator is not preferred, but the *npar-greg* estimator is recommended.

Results of the nominal coverage and average length of the 95% point-wise confidence intervals are presented in Figures 21–28. In each figure, we show the results of the nominal coverage and average length split up by underlying function and sample size. The *unw* and *trim* estimators have very poor coverages resulting from the bias associated with the estimates. The *psm* slightly underestimates the true coverage at almost all time points. For the small sample size (Figures 27 and 28) this undercoverage is even substantial. For F1 and F2, the coverage results of the *npar* and *npar-greg* are close to

the actual coverage level. Both the analytical and bootstrap based CIs of the *npar* estimator have good coverage values. For F3 and F4, the coverage of the *npar* CIs based on the analytical variance shows a drop in the sharp valley area. The bootstrap variance based CIs attain better nominal coverages. However, this comes at the cost of wider CIs in the sharp valley area. For the smallest sample size (Figures 27 and 28), the CIs of the *npar* estimator based on the bootstrap approach still underestimate the actual coverage. The jackknife based CIs of the *npar-greg* estimator achieve good coverage for all simulation settings for all time points. The average lengths of the CIs of the *npar-greg* are comparable with those of *trim* and are substantially smaller when compared to *psm*.

Overall, the *npar* estimator performs well for mean trend estimation. However, when bias is present in the NPAR weight smoothing model, the *npar-greg* estimator is preferred for mean trend estimation.

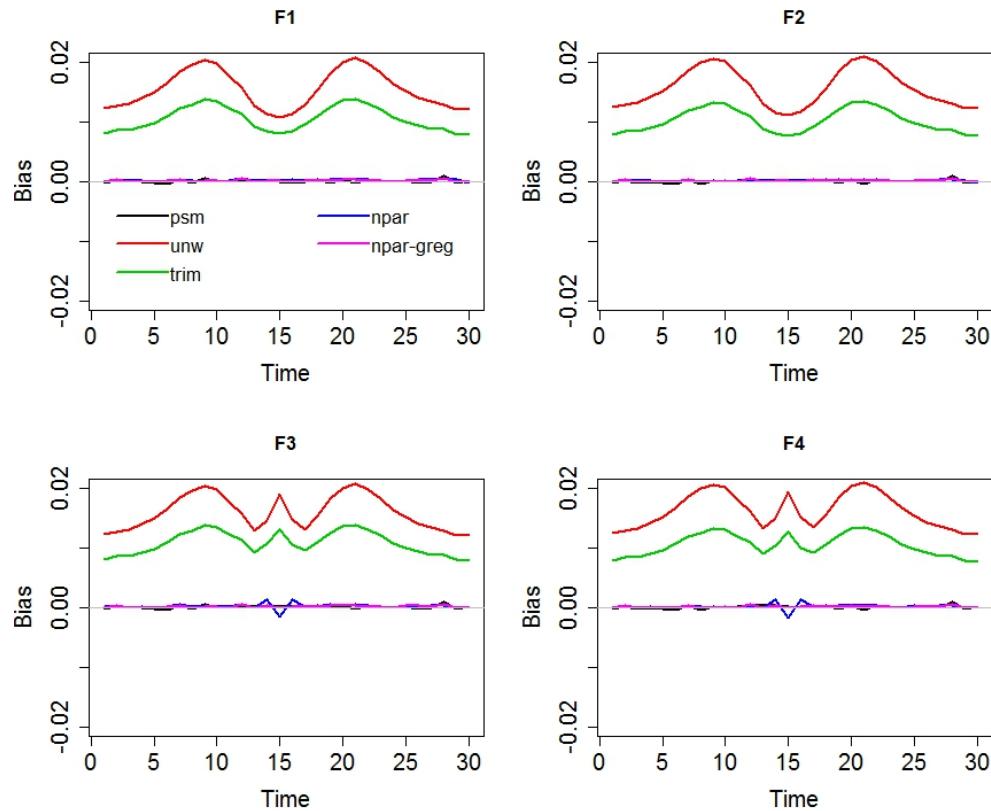


FIG 9. Bias at each timepoint for population models F1, F2, F3 and F4 with population size  $N^{(1)} = 6,000,000$  and sample size 25,000.

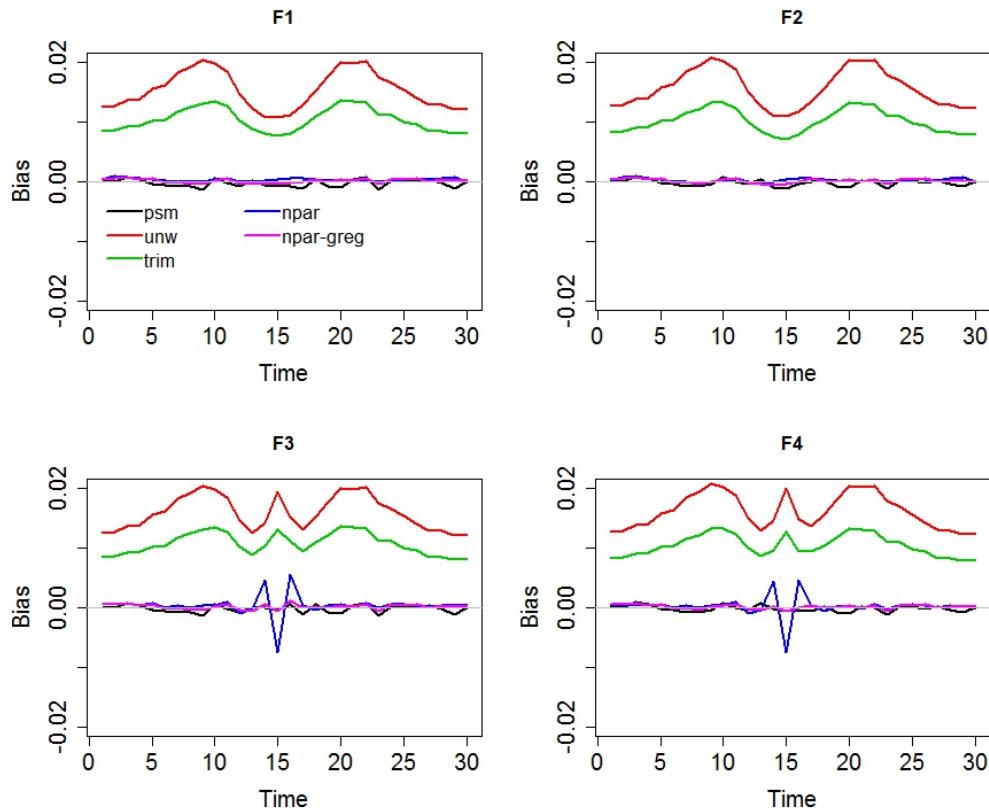


FIG 10. Bias at each timepoint for population models  $F_1$ ,  $F_2$ ,  $F_3$  and  $F_4$  with population size  $N^{(1)} = 6,000,000$  and sample size 5,000.

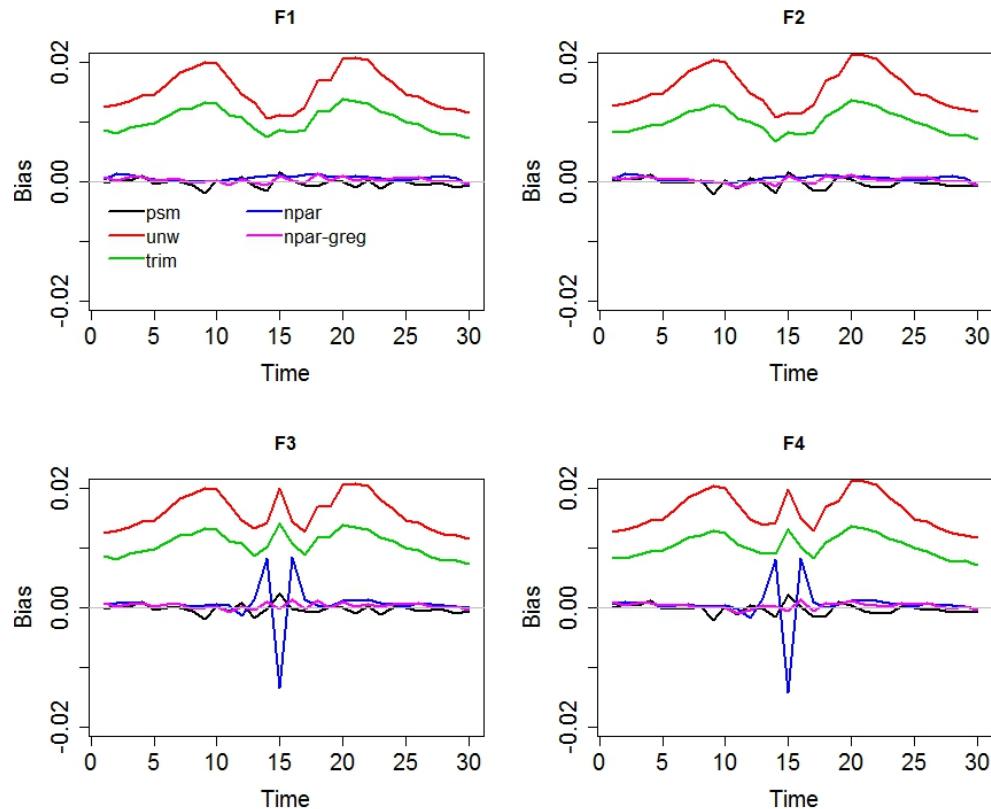


FIG 11. Bias at each timepoint for population models  $F_1$ ,  $F_2$ ,  $F_3$  and  $F_4$  with population size  $N^{(2)} = 150,000$  and sample size 2,500.

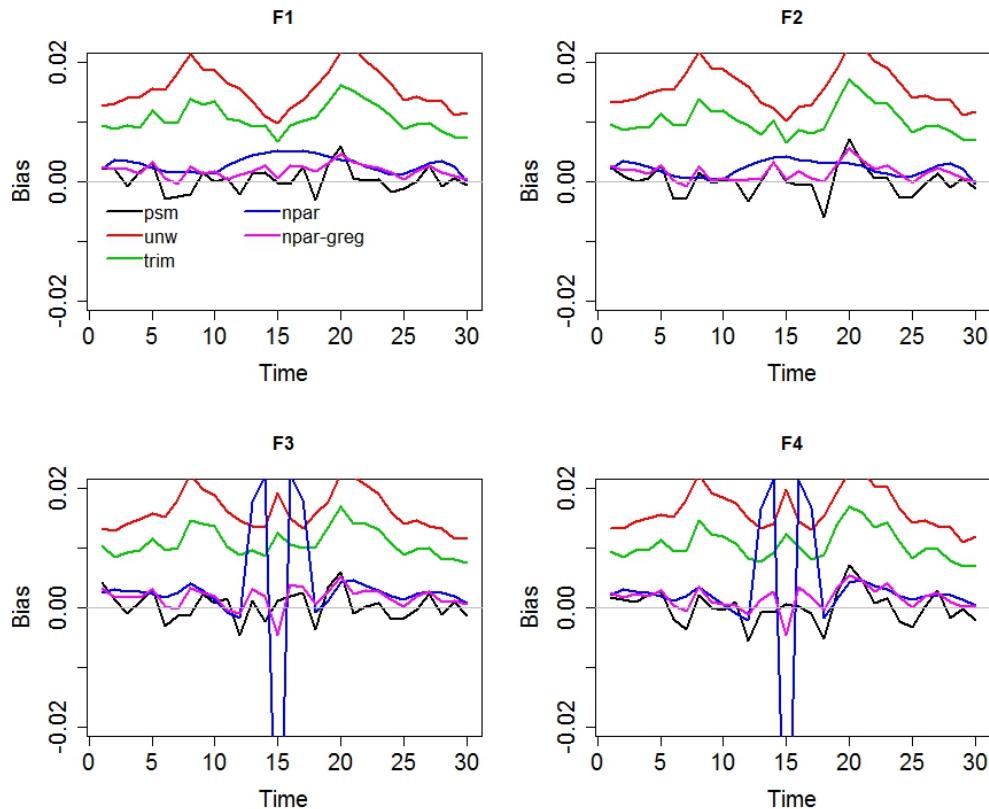


FIG 12. Bias at each timepoint for population models  $F_1$ ,  $F_2$ ,  $F_3$  and  $F_4$  with population size  $N^{(2)} = 150,000$  and sample size 500.

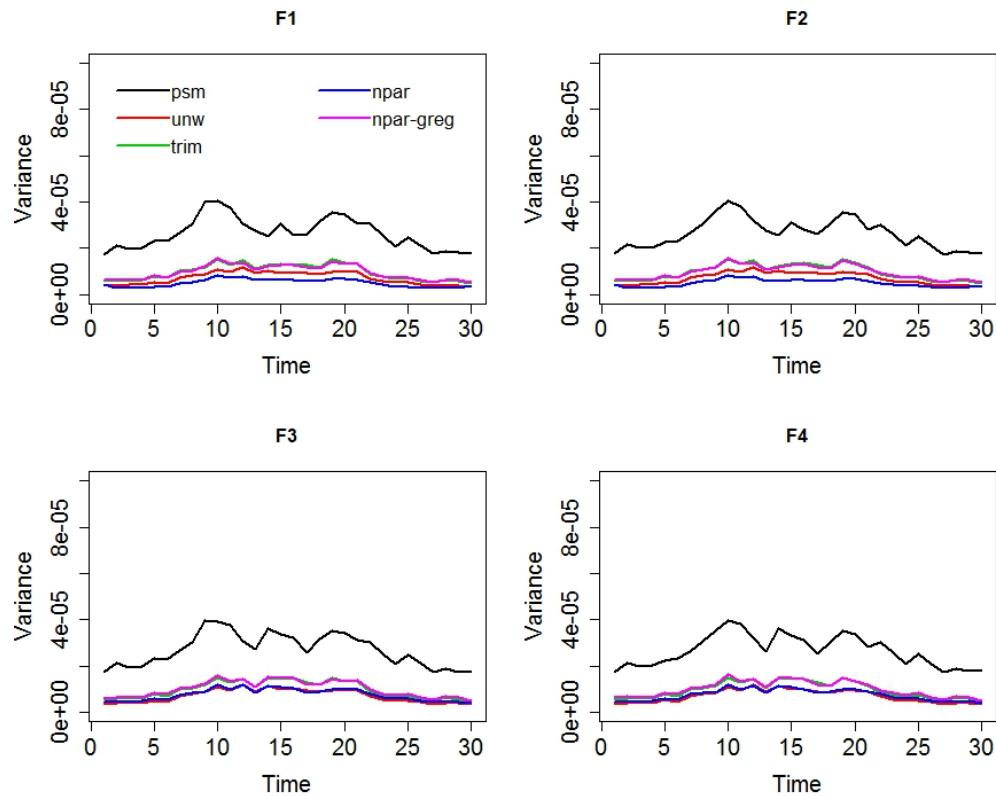


FIG 13. Variance at each timepoint for population models  $F_1$ ,  $F_2$ ,  $F_3$  and  $F_4$  with population size  $N^{(1)} = 6,000,000$  and sample size 25,000.

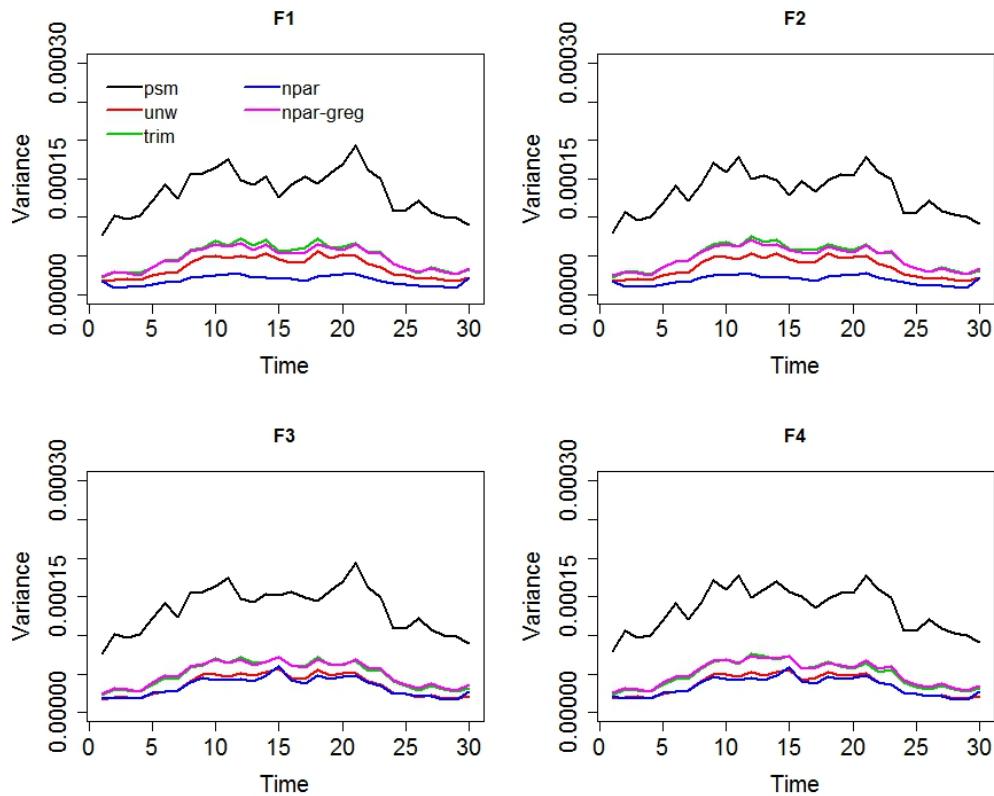


FIG 14. Variance at each timepoint for population models  $F_1$ ,  $F_2$ ,  $F_3$  and  $F_4$  with population size  $N^{(1)} = 6,000,000$  and sample size 5,000.

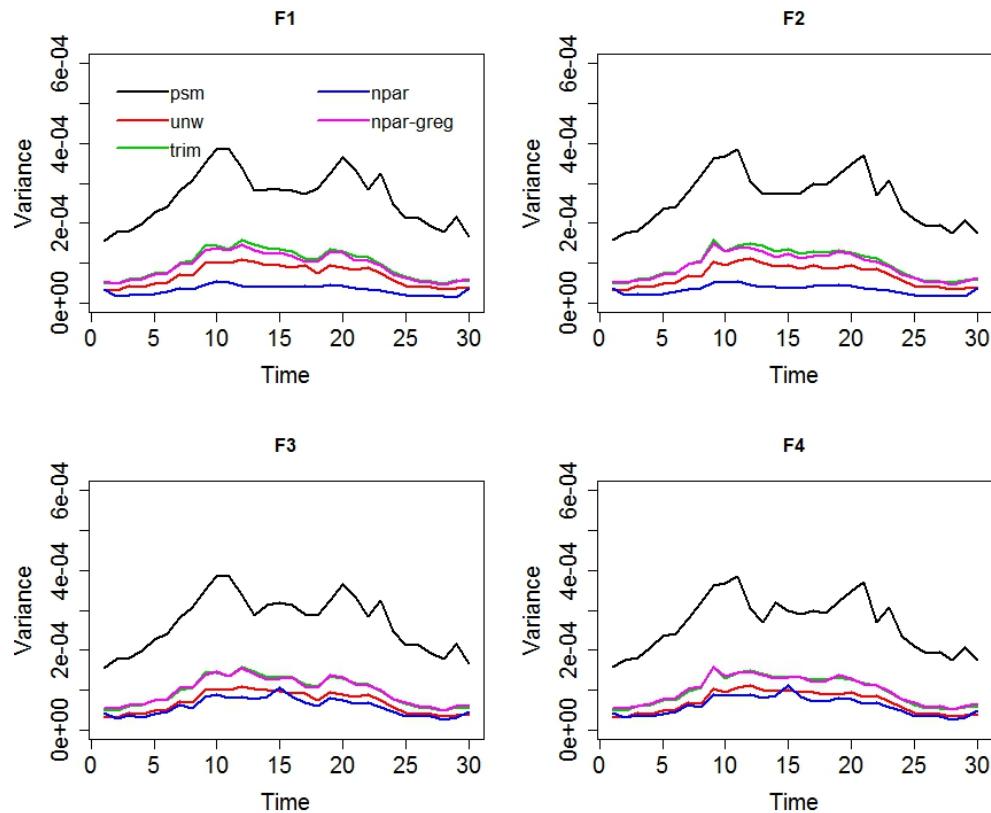


FIG 15. Variance at each timepoint for population models  $F_1$ ,  $F_2$ ,  $F_3$  and  $F_4$  with population size  $N^{(2)} = 150,000$  and sample size 2,500.

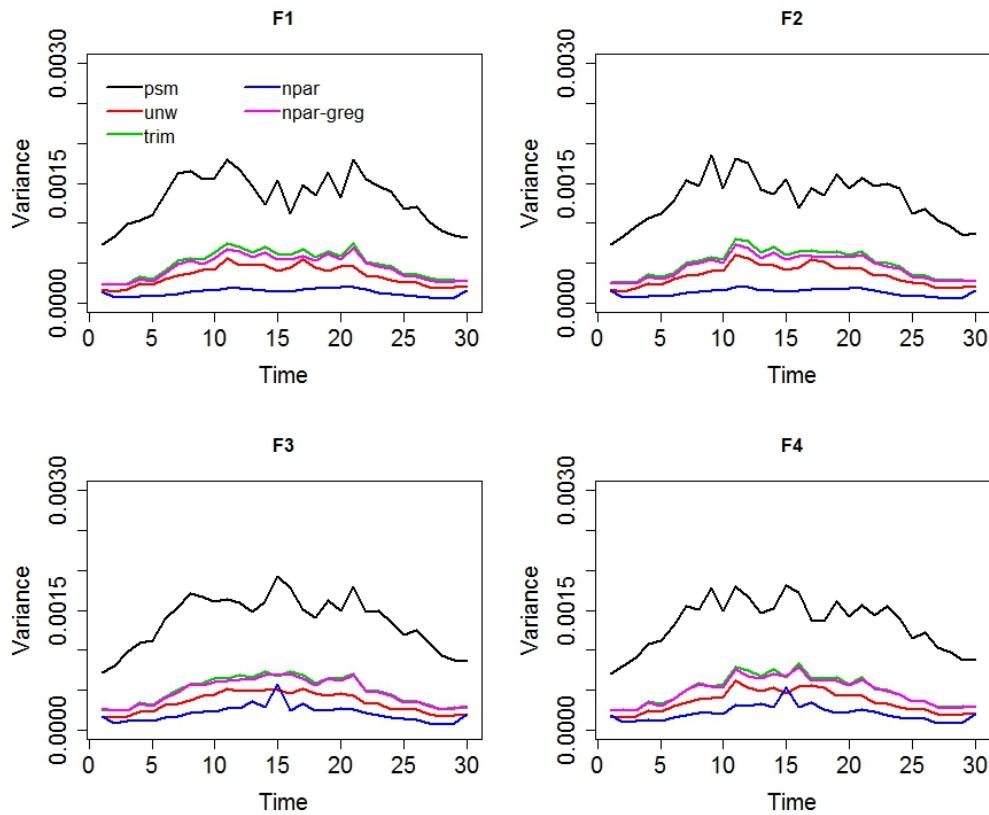


FIG 16. Variance at each timepoint for population models  $F_1$ ,  $F_2$ ,  $F_3$  and  $F_4$  with population size  $N^{(2)} = 150,000$  and sample size 500.

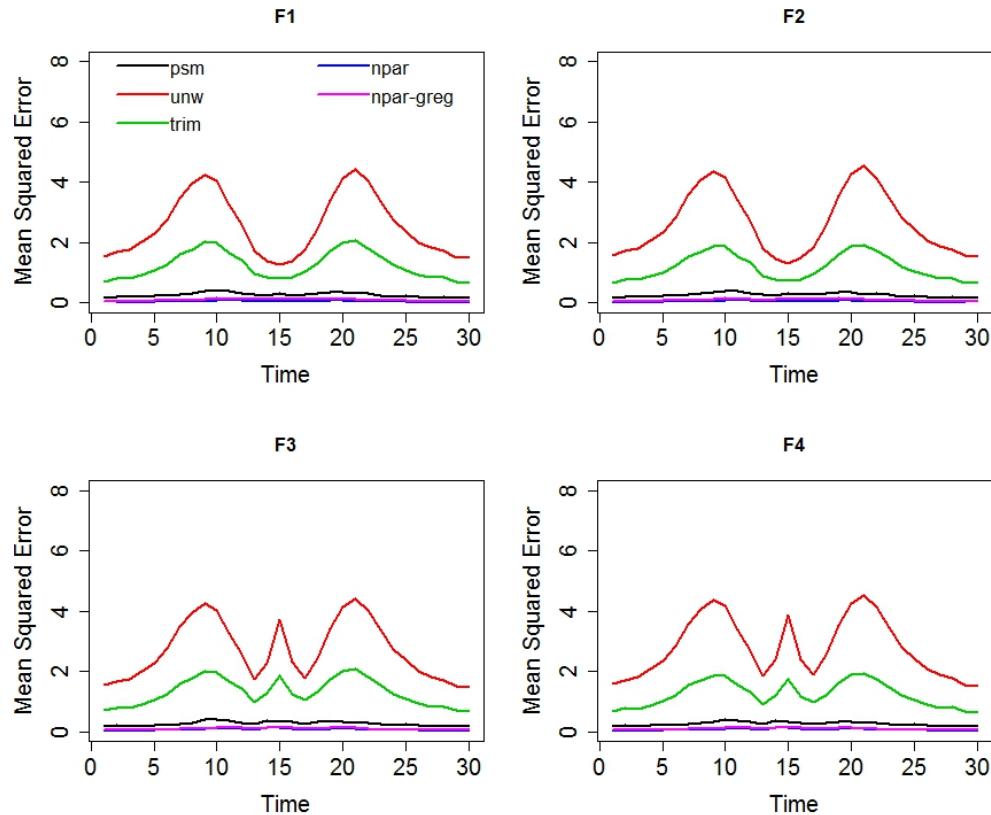


FIG 17. Mean squared errors  $\times 10^4$  at each timepoint for population models  $F_1$ ,  $F_2$ ,  $F_3$  and  $F_4$  with population size  $N^{(1)} = 6,000,000$  and sample size 25,000.

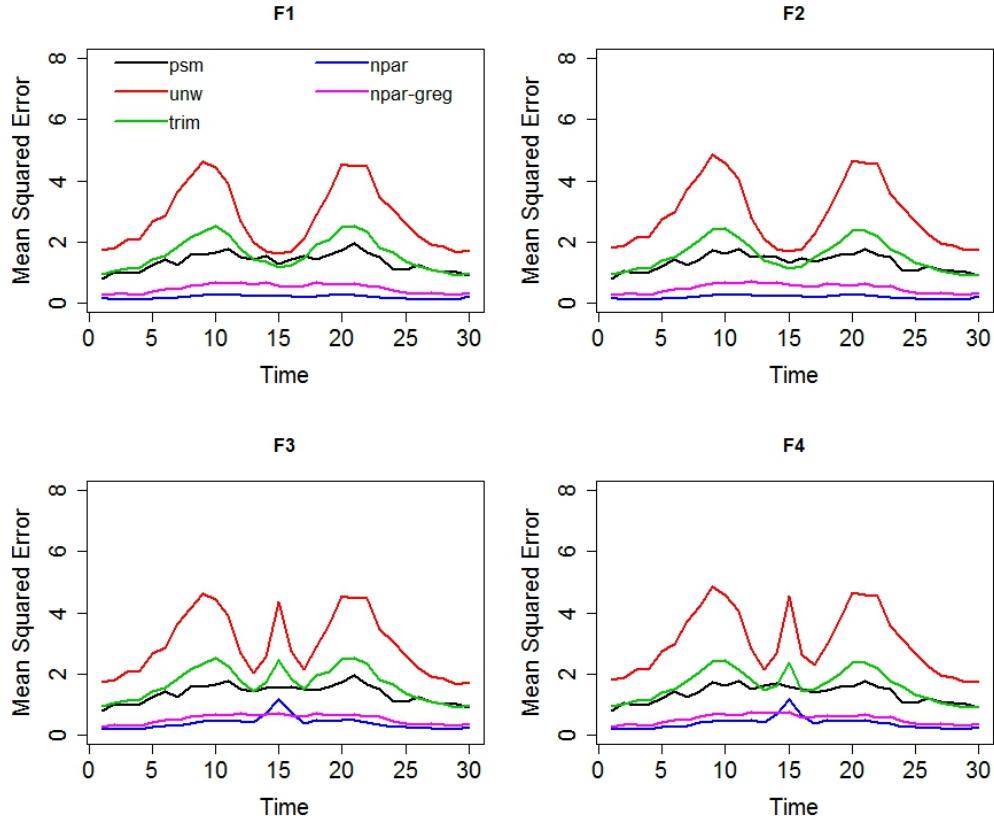


FIG 18. Mean squared errors  $\times 10^4$  at each timepoint for population models  $F_1$ ,  $F_2$ ,  $F_3$  and  $F_4$  with population size  $N^{(1)} = 6,000,000$  and sample size 5,000.

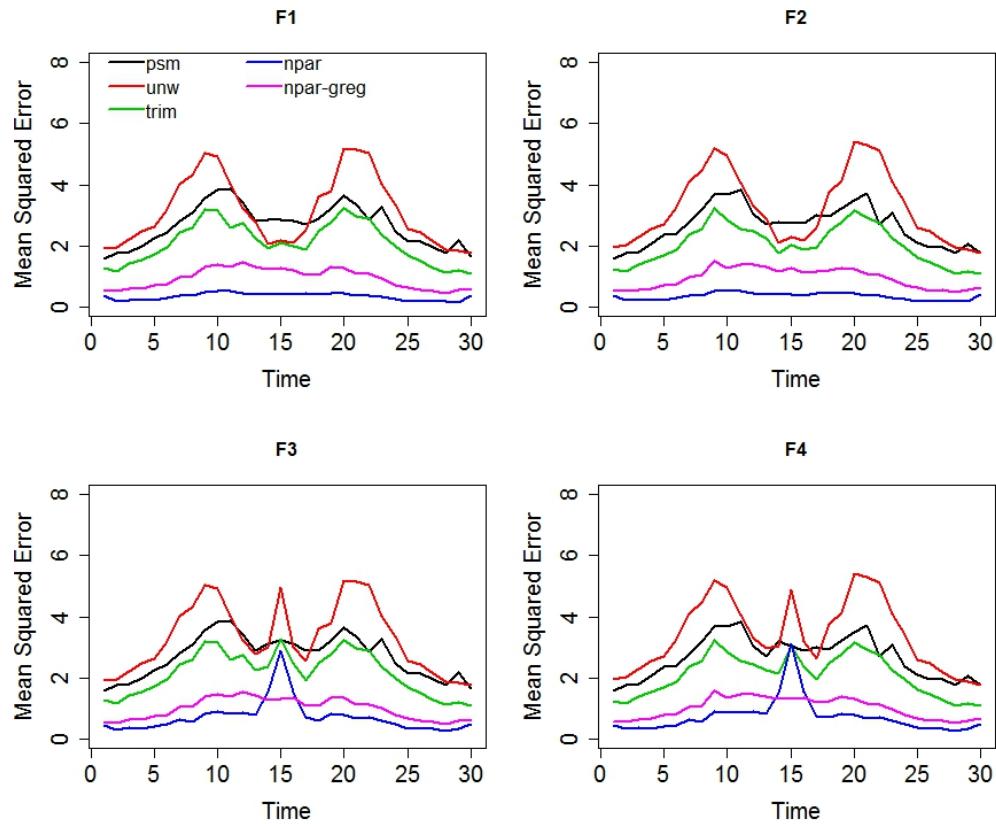


FIG 19. Mean squared errors  $\times 10^4$  at each timepoint for population models  $F_1$ ,  $F_2$ ,  $F_3$  and  $F_4$  with population size  $N^{(2)} = 150,000$  and sample size 2,500.

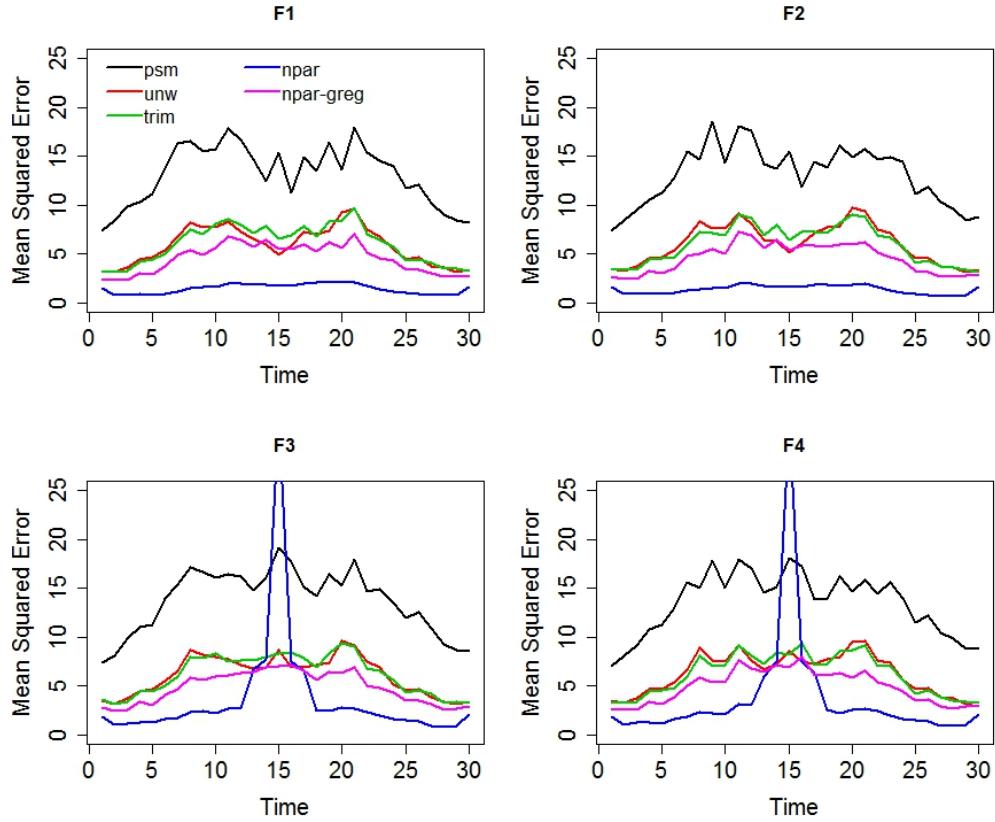


FIG 20. Mean squared errors  $\times 10^4$  at each timepoint for population models  $F_1$ ,  $F_2$ ,  $F_3$  and  $F_4$  with population size  $N^{(2)} = 150,000$  and sample size 500.

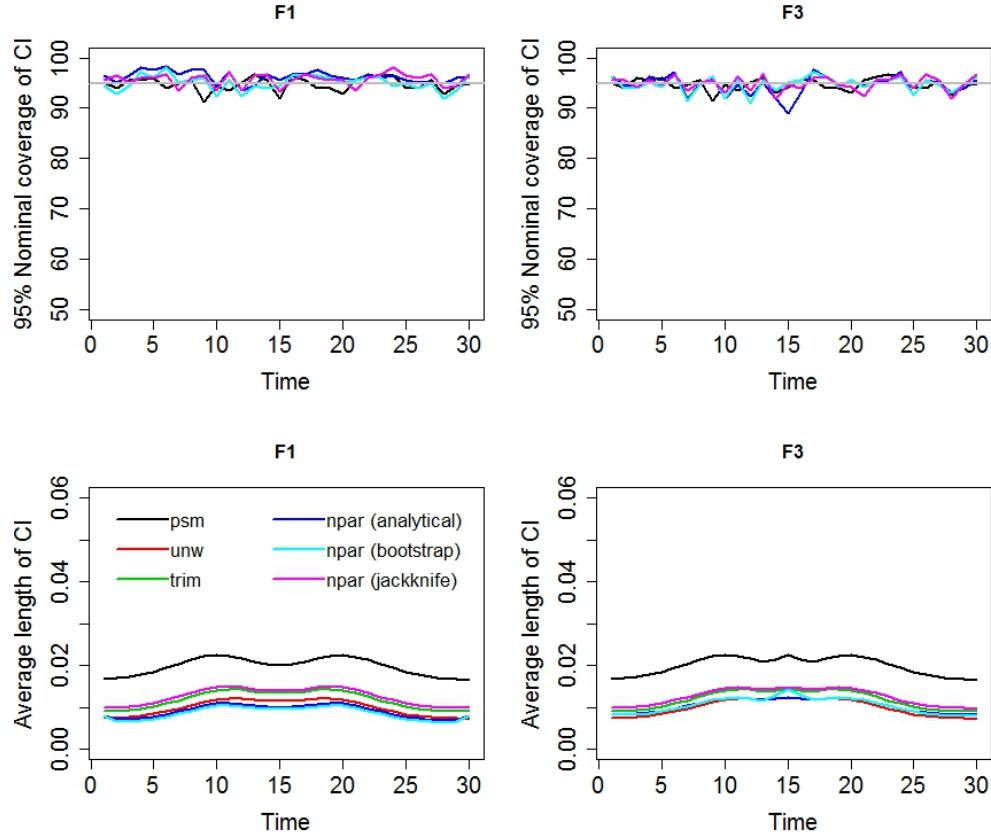


FIG 21. Nominal coverage and average length of the 95% point-wise confidence intervals at each timepoint for population models F1 and F3 with population size  $N^{(1)} = 6,000,000$  and sample size 25,000. The nominal coverage of some estimators is smaller than the range of the y-axis used and are, therefore, not depicted on the figure.

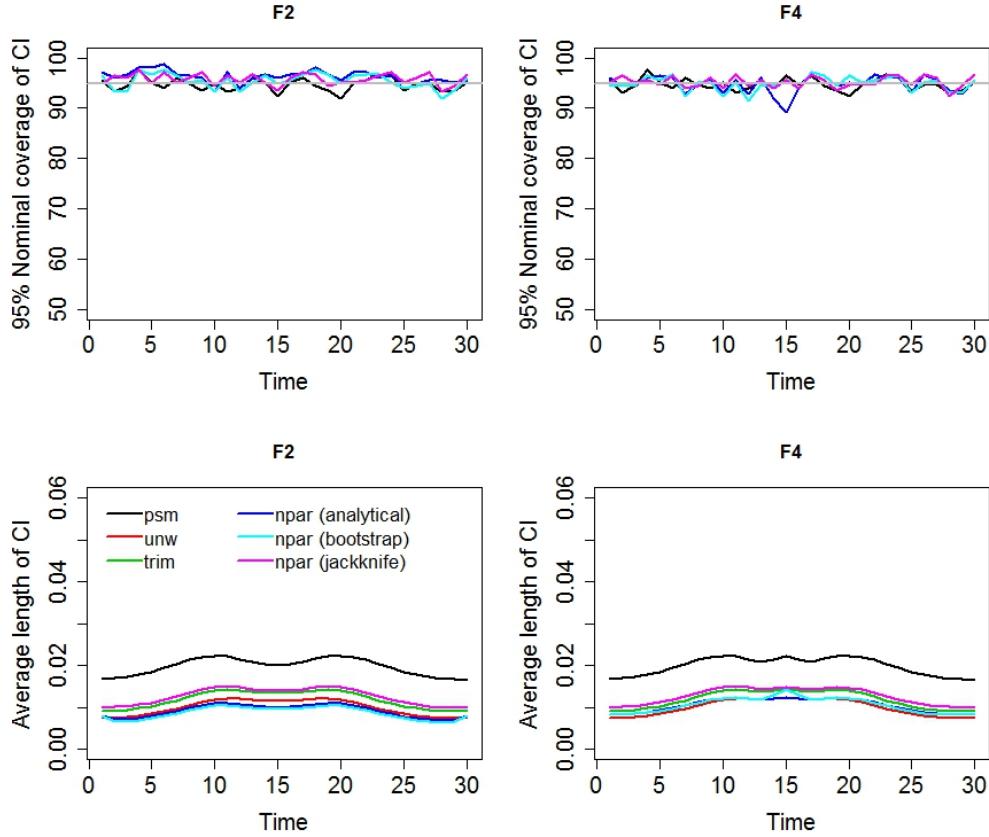


FIG 22. Nominal coverage and average length of the 95% point-wise confidence intervals at each timepoint for population models F2 and F4 with population size  $N^{(1)} = 6,000,000$  and sample size 25,000. The nominal coverage of some estimators is smaller than the range of the y-axis used and are, therefore, not depicted on the figure.

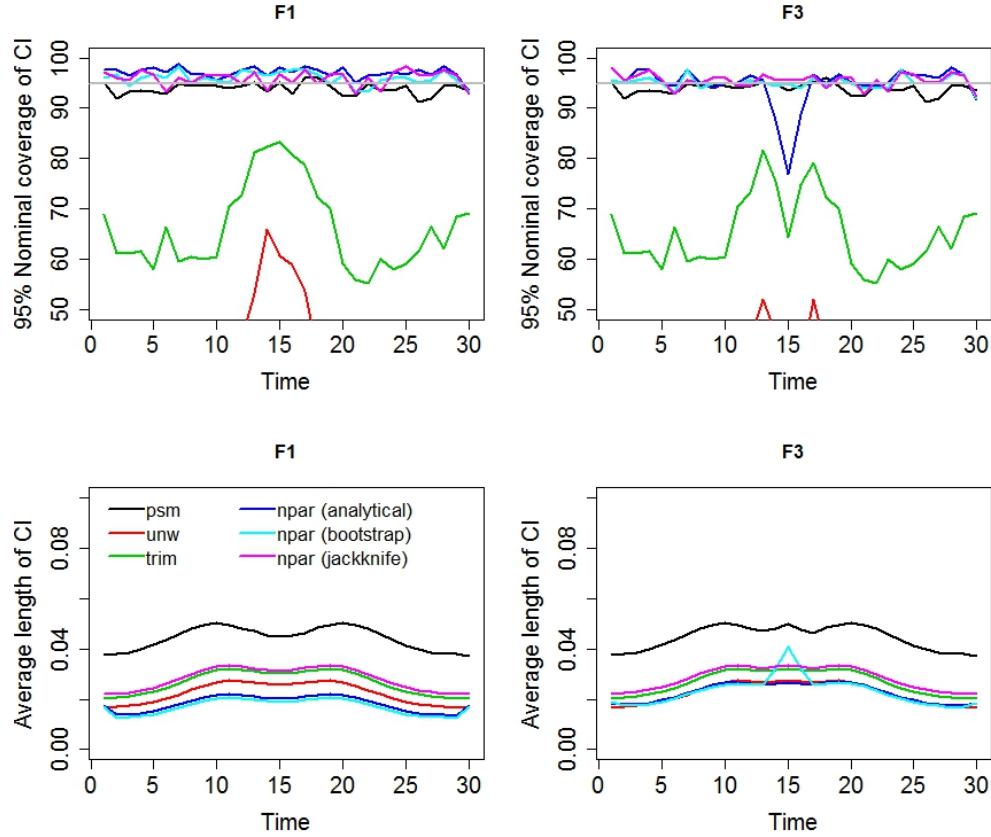


FIG 23. Nominal coverage and average length of the 95% point-wise confidence intervals at each timepoint for population models F1 and F3 with population size  $N^{(1)} = 6,000,000$  and sample size 5,000. The nominal coverage of some estimators is smaller than the range of the y-axis used and are, therefore, not depicted on the figure.

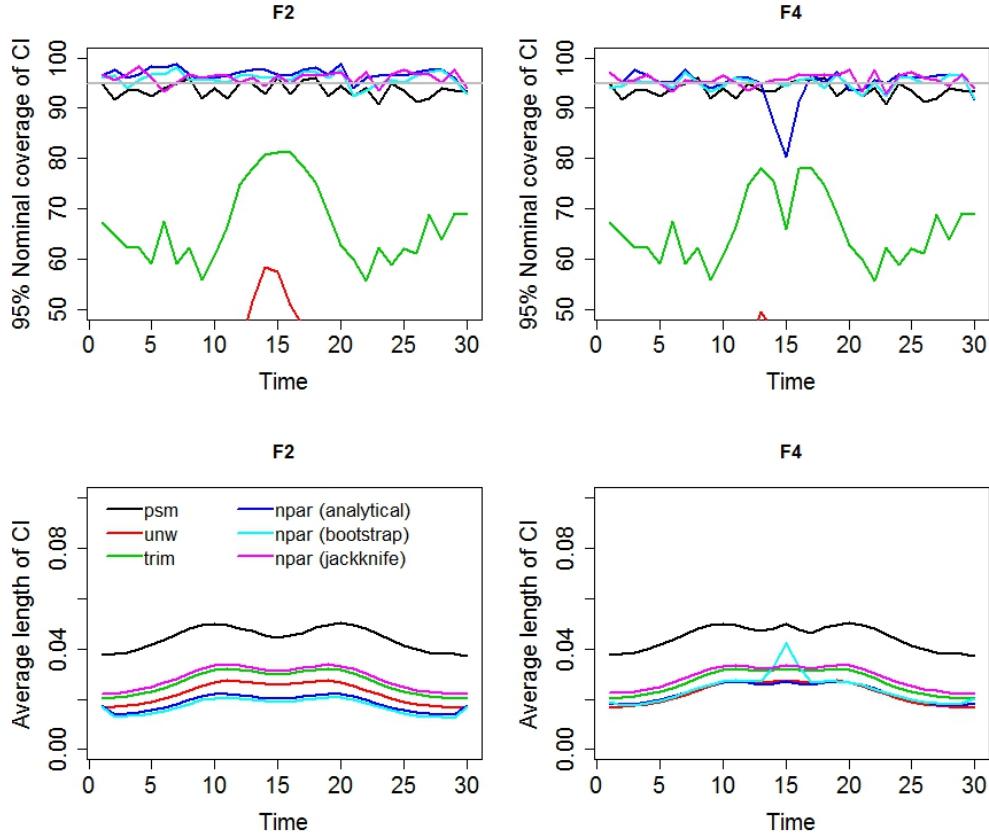


FIG 24. Nominal coverage and average length of the 95% point-wise confidence intervals at each timepoint for population models F2 and F4 with population size  $N^{(1)} = 6,000,000$  and sample size 5,000. The nominal coverage of some estimators is smaller than the range of the y-axis used and are, therefore, not depicted on the figure.

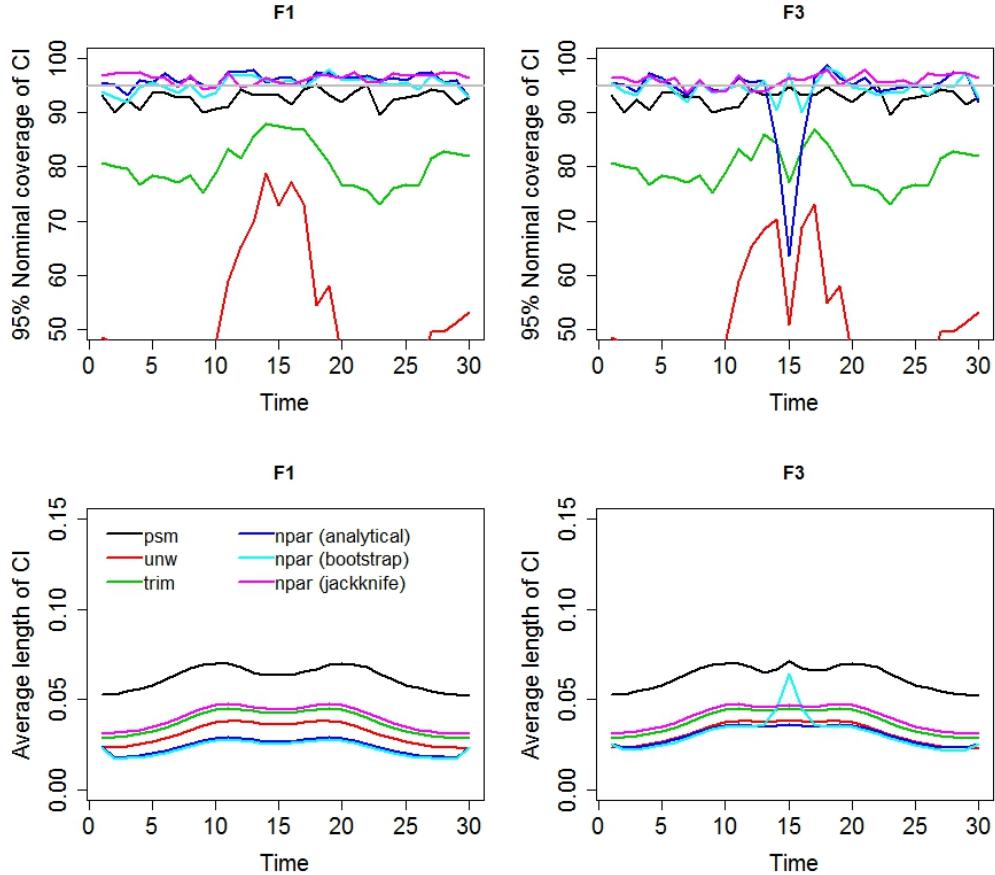


FIG 25. Nominal coverage and average length of the 95% point-wise confidence intervals at each timepoint for population models F1 and F3 with population size  $N^{(1)} = 150,000$  and sample size 2,500. The nominal coverage of some estimators is smaller than the range of the y-axis used and are, therefore, not depicted on the figure.

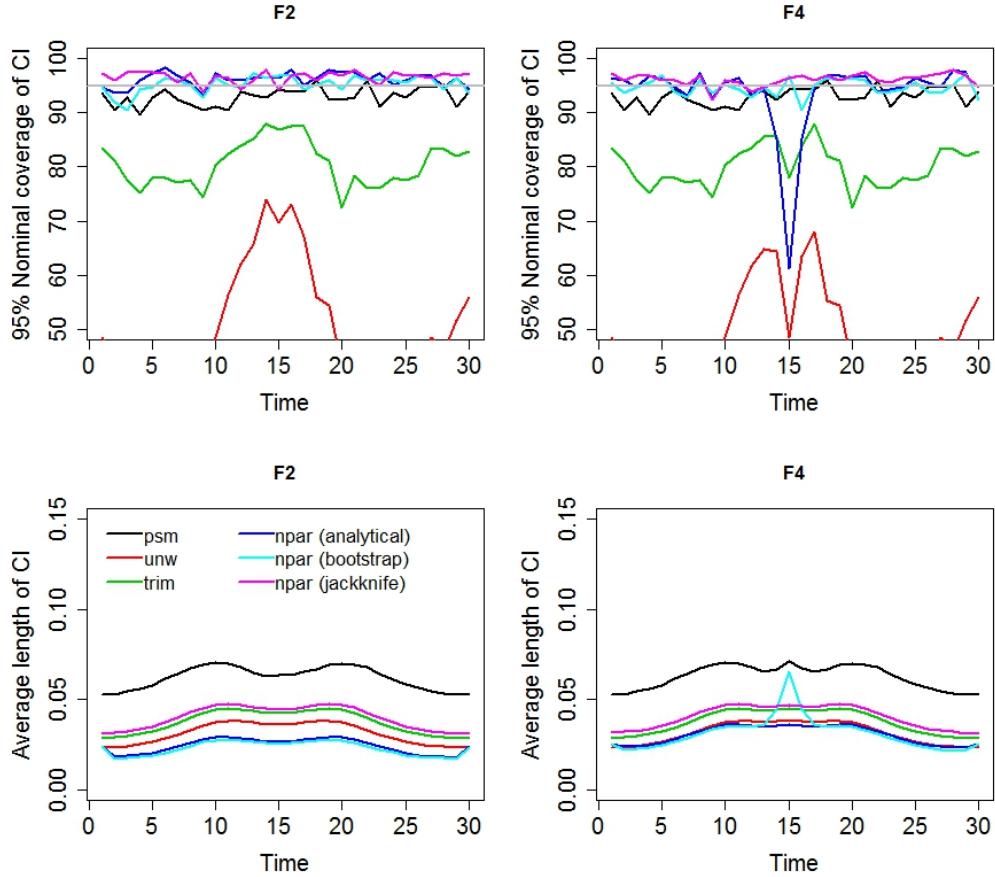


FIG 26. Nominal coverage and average length of the 95% point-wise confidence intervals at each timepoint for population models F2 and F4 with population size  $N^{(1)} = 150,000$  and sample size 2,500. The nominal coverage of some estimators is smaller than the range of the y-axis used and are, therefore, not depicted on the figure.

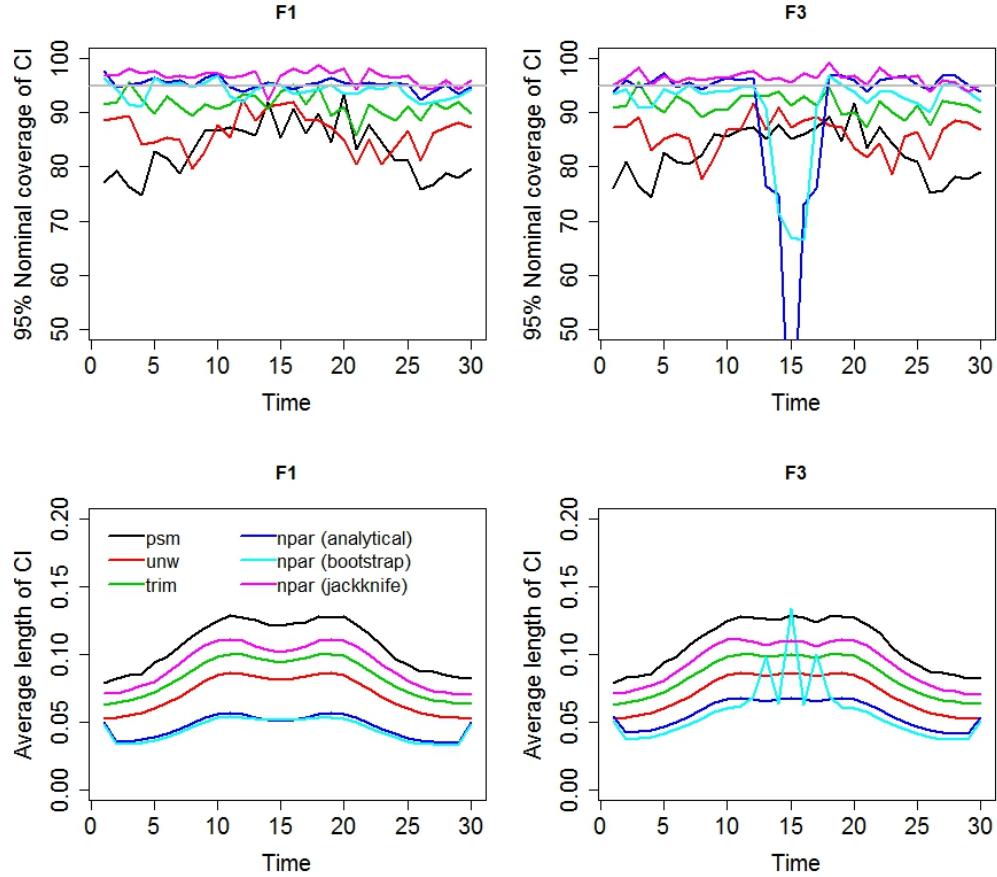


FIG 27. Nominal coverage and average length of the 95% point-wise confidence intervals at each timepoint for population models F1 and F3 with population size  $N^{(1)} = 150,000$  and sample size 500. The nominal coverage of some estimators is smaller than the range of the y-axis used and are, therefore, not depicted on the figure.

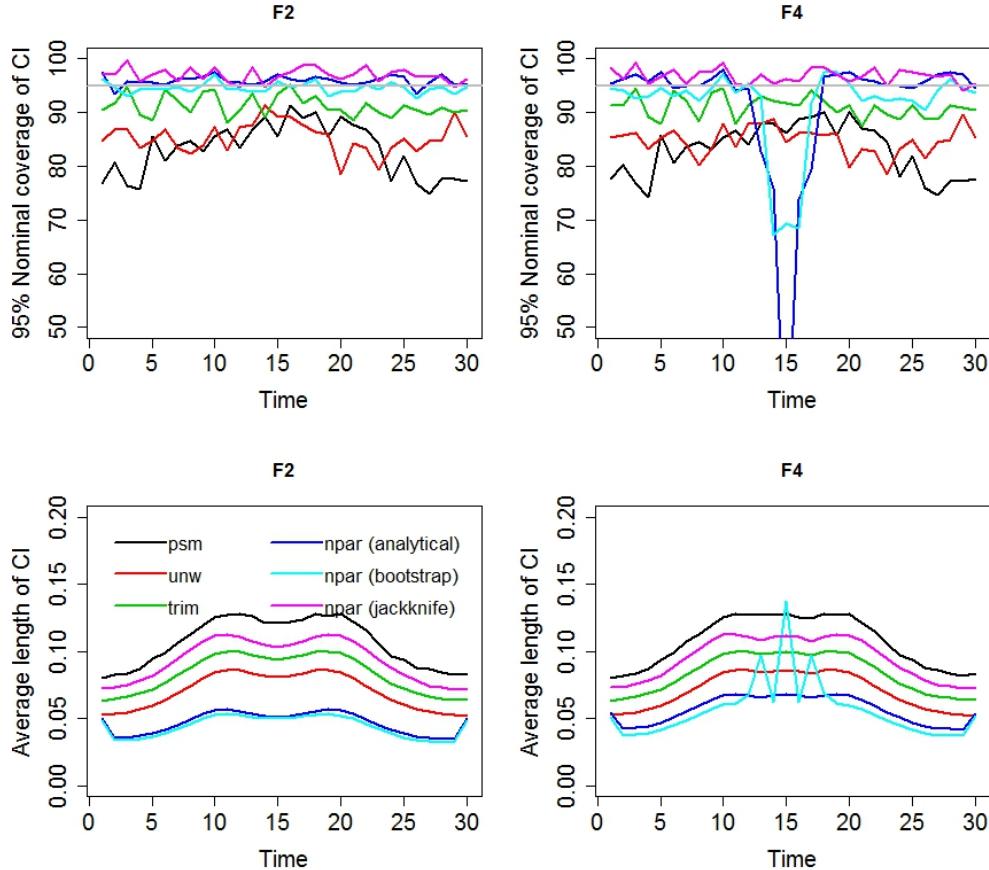


FIG 28. Nominal coverage and average length of the 95% point-wise confidence intervals at each timepoint for population models F2 and F4 with population size  $N^{(1)} = 150,000$  and sample size 500. The nominal coverage of some estimators is smaller than the range of the y-axis used and are, therefore, not depicted on the figure.

## Appendix E.

*Additional simulation results with respect to varying values of  $w_0$ .* In this section, we present some simulation results to investigate the impact of the value of  $w_0$  in the GREG adjusted estimators (5) and (6). For the prevalence, we used similar settings as described in Section 5.1. We performed the simulation for the population with size 6,000,000. Ten populations were randomly generated and ten samples of size 5,000 were taken in each population. The value of  $w_0$  used in the *trim* estimator and the GREG-adjusted estimators are 2, 3, 5 or 10. The bias and variance results are presented in Figures 29–36 and Tables 13–16 below. As expected, smaller values of  $w_0$  yield estimates that have a larger bias, but smaller variance, the opposite is true for larger values of  $w_0$ . In terms of MSE, no clear conclusion on the optimal value for  $w_0$  can be given since this depends on the true underlying population.

For trend estimation, similar settings as in Section 5.3 are used. Again a population size of 6,000,000 and a sample size of 5,000 are used. One hundred replications are considered. Only the figures with respect to the mean squared error are given (Figures 37–40). Similar results as for the prevalence estimation are observed. It is also observed that for increasing values of  $w_0$  the MSE of the trend estimators *trim* and *npar-greg* approximates the MSE of the *psm* estimator.

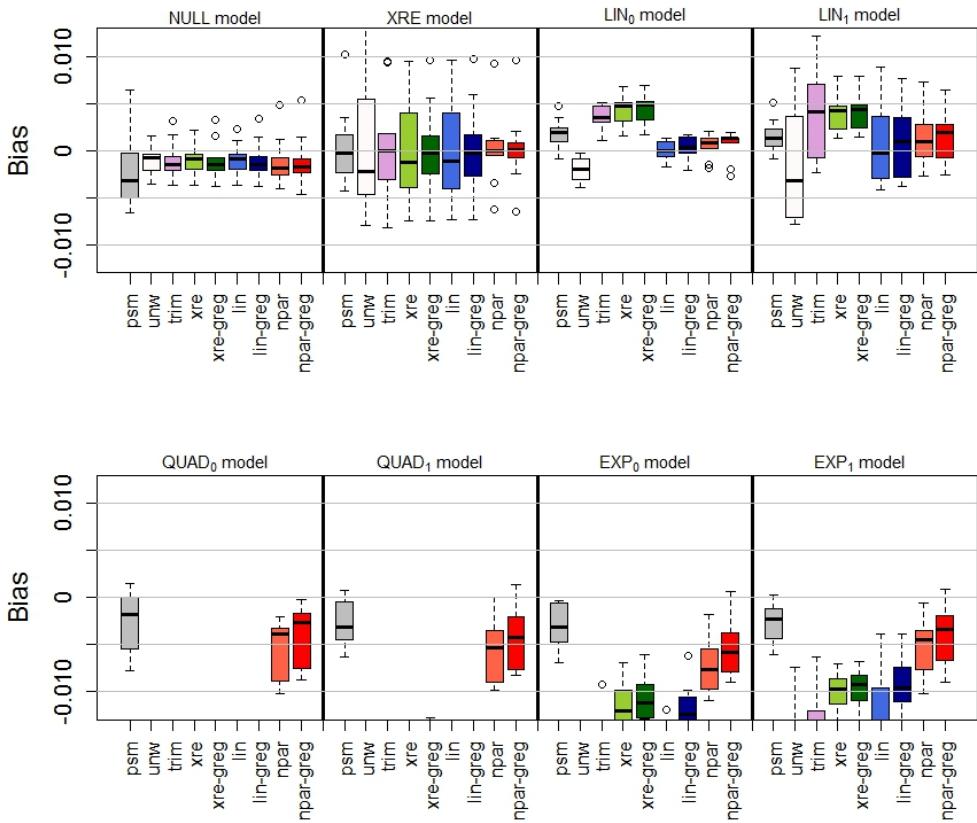


FIG 29. Bias results of nine estimators and eight population models for the large population  $N^{(1)} = 6,000,000$  with a sample size of 5,000 and  $w_0 = 2$ . Boxplots of the average bias associated with the 10 simulated populations are shown. The bias of some estimators exceeds the range of the y-axis used and are, therefore, not depicted on the figure.

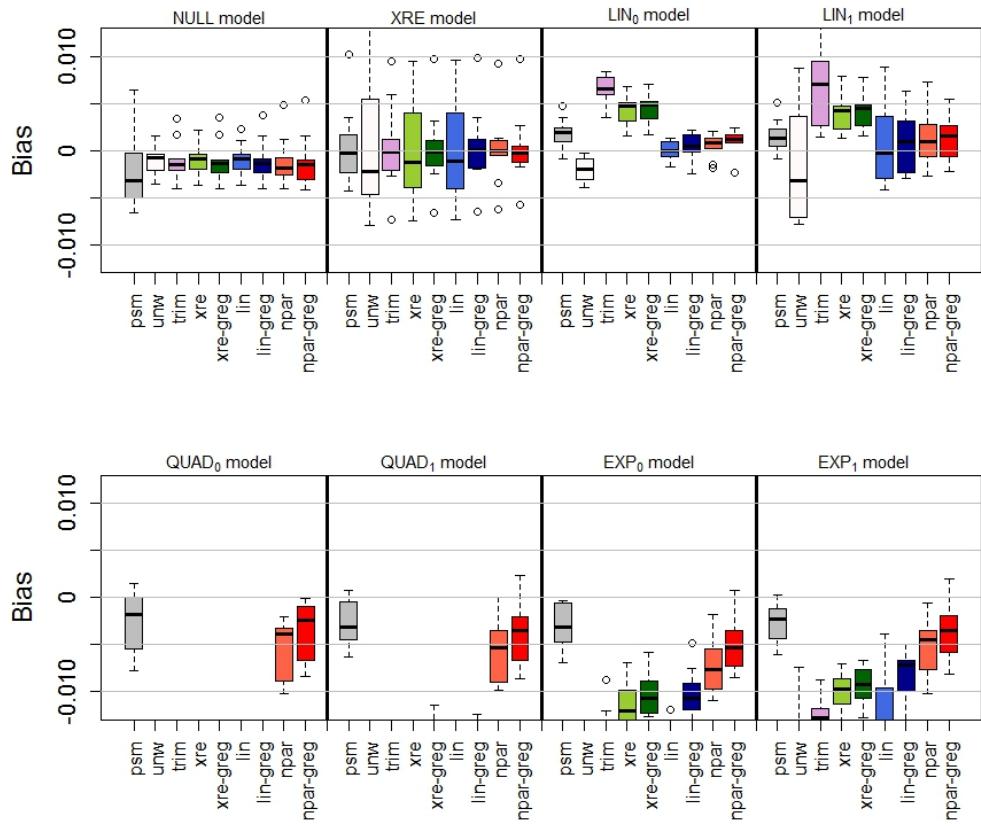


FIG 30. Bias results of nine estimators and eight population models for the large population  $N^{(1)} = 6,000,000$  with a sample size of 5,000 and  $w_0 = 3$ . Boxplots of the average bias associated with the 10 simulated populations are shown. The bias of some estimators exceeds the range of the y-axis used and are, therefore, not depicted on the figure.

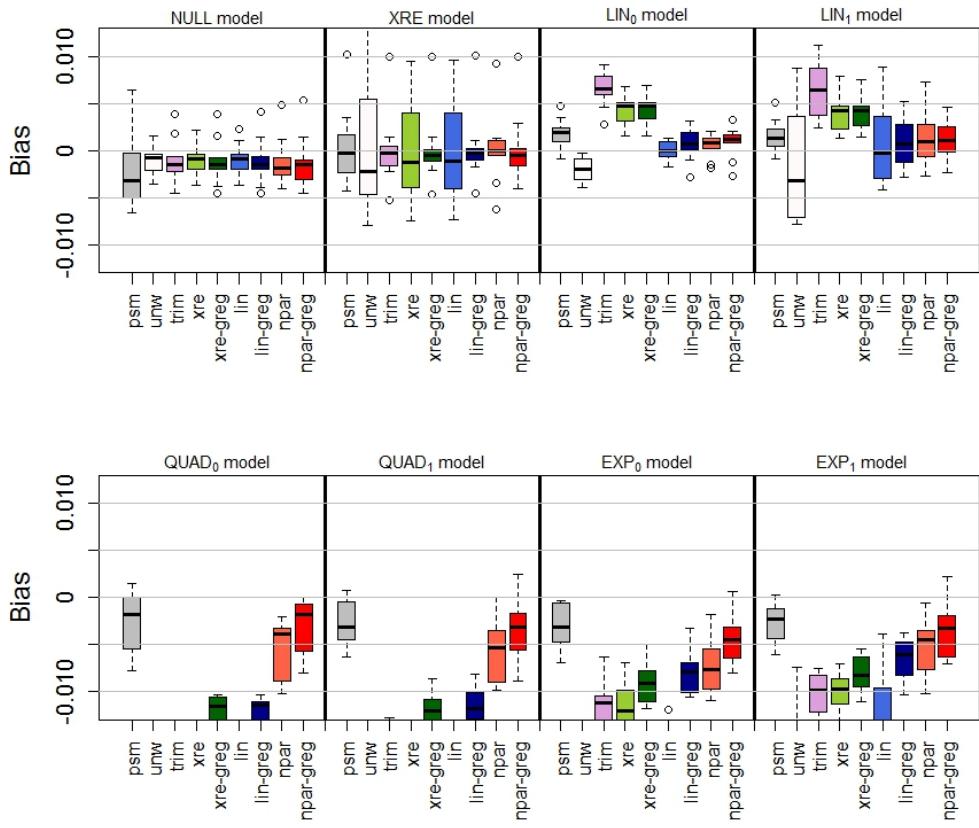


FIG 31. Bias results of nine estimators and eight population models for the large population  $N^{(1)} = 6,000,000$  with a sample size of 5,000 and  $w_0 = 5$ . Boxplots of the average bias associated with the 10 simulated populations are shown. The bias of some estimators exceeds the range of the y-axis used and are, therefore, not depicted on the figure.

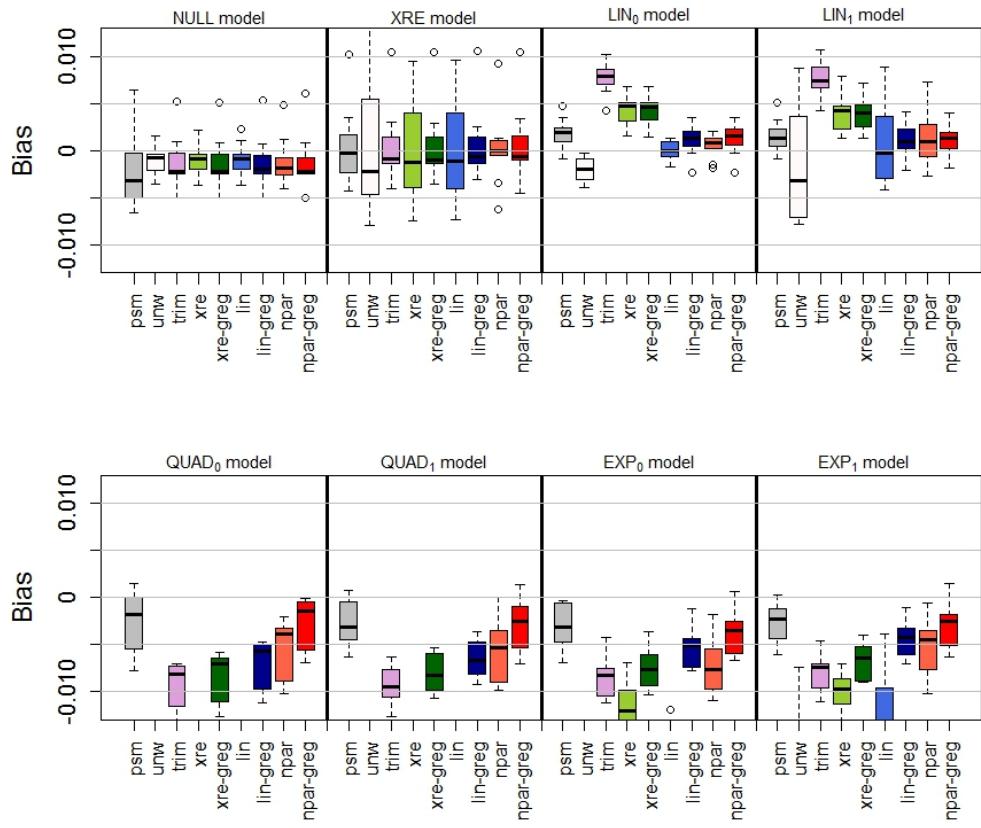


FIG 32. Bias results of nine estimators and eight population models for the large population  $N^{(1)} = 6,000,000$  with a sample size of 5,000 and  $w_0 = 10$ . Boxplots of the average bias associated with the 10 simulated populations are shown. The bias of some estimators exceeds the range of the y-axis used and are, therefore, not depicted on the figure.

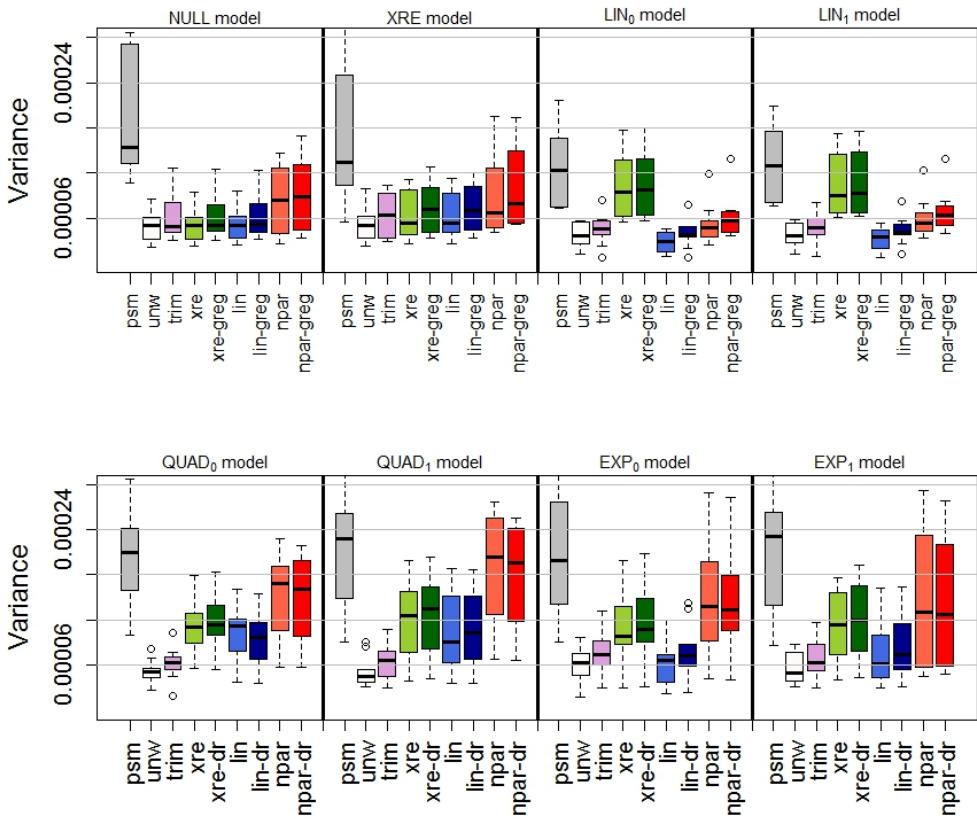


FIG 33. Variance results of nine estimators and eight population models for the large population  $N^{(1)} = 6,000,000$  with a sample size of 5,000 and  $w_0 = 2$ . Boxplots of the average bias associated with the 10 simulated populations are shown. The variance of some estimators exceeds the range of the y-axis used and are, therefore, not depicted on the figure.

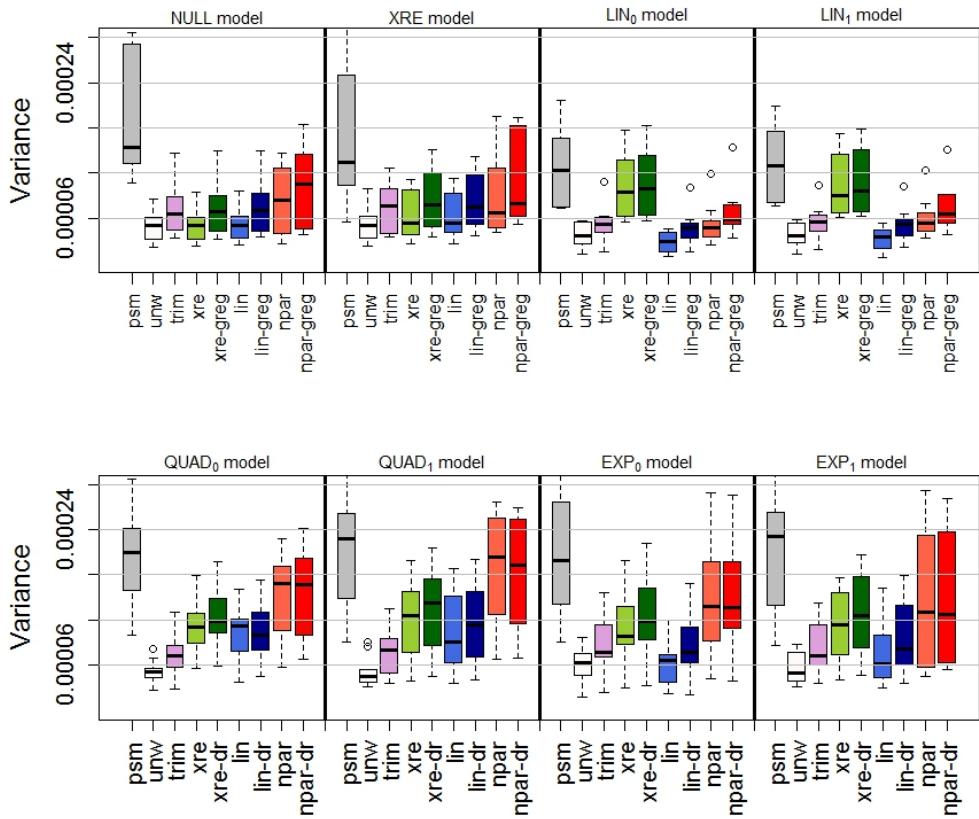


FIG 34. Variance results of nine estimators and eight population models for the large population  $N^{(1)} = 6,000,000$  with a sample size of 5,000 and  $w_0 = 3$ . Boxplots of the average bias associated with the 10 simulated populations are shown. The variance of some estimators exceeds the range of the y-axis used and are, therefore, not depicted on the figure.

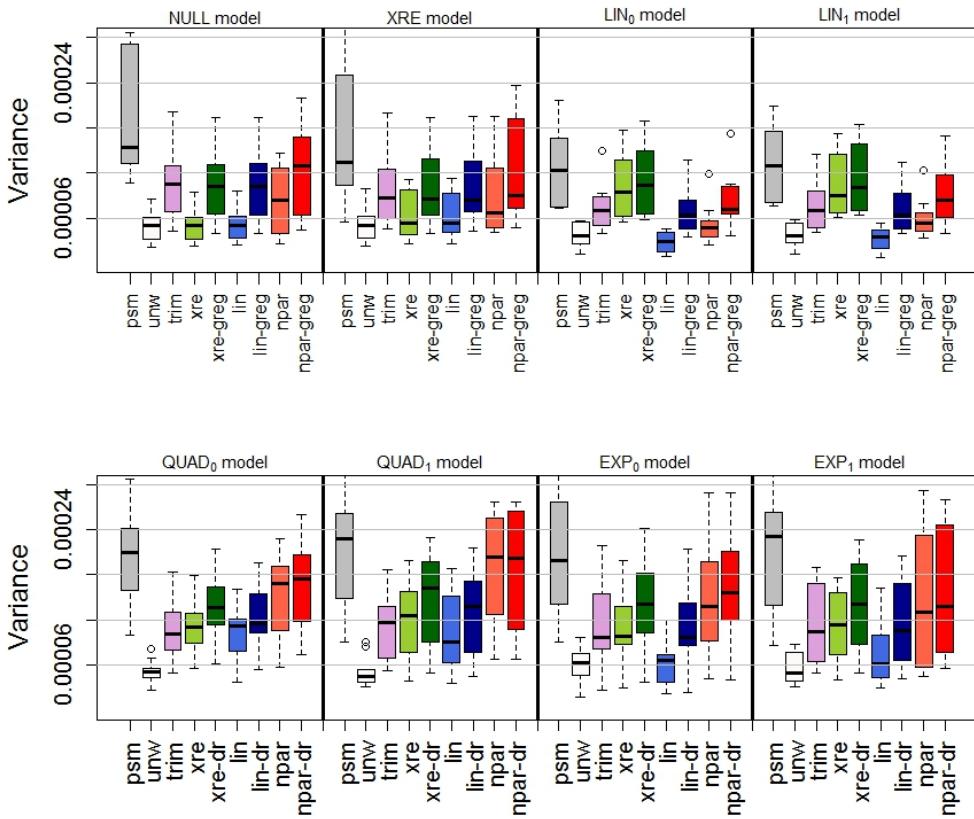


FIG 35. Variance results of nine estimators and eight population models for the large population  $N^{(1)} = 6,000,000$  with a sample size of 5,000 and  $w_0 = 5$ . Boxplots of the average bias associated with the 10 simulated populations are shown. The variance of some estimators exceeds the range of the y-axis used and are, therefore, not depicted on the figure.

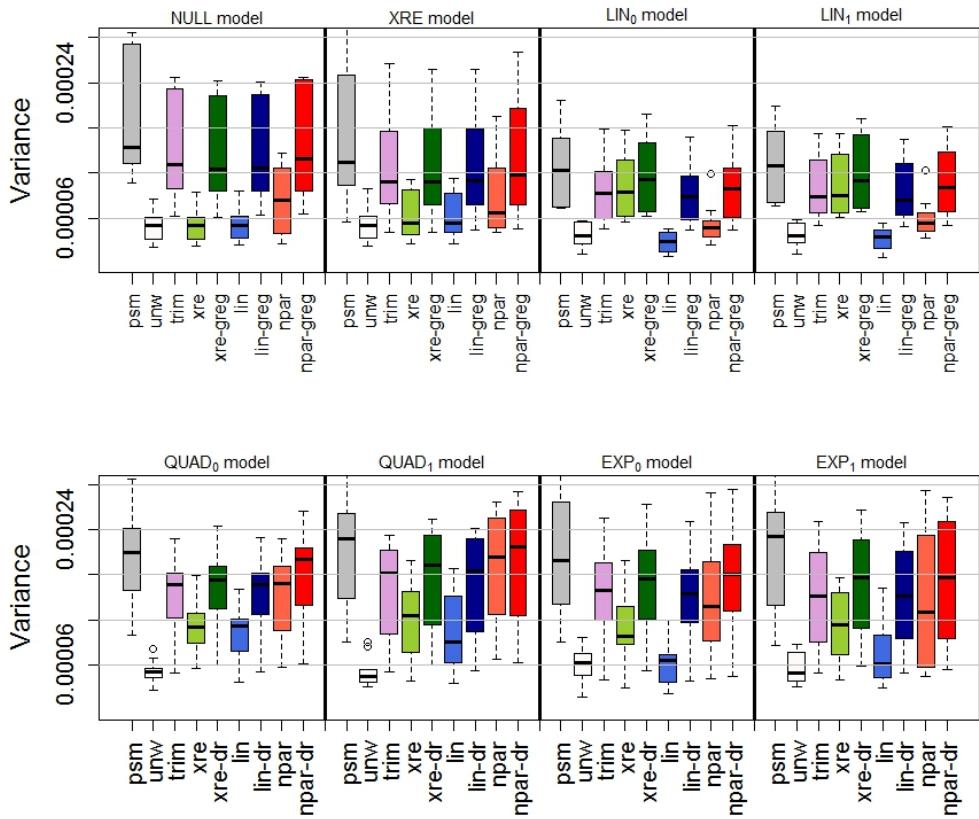


FIG 36. Variance results of nine estimators and eight population models for the large population  $N^{(1)} = 6,000,000$  with a sample size of 5,000 and  $w_0 = 10$ . Boxplots of the average bias associated with the 10 simulated populations are shown. The variance of some estimators exceeds the range of the y-axis used and are, therefore, not depicted on the figure.

TABLE 13

*Mean squared error ( $\times 10^4$ ) of nine estimators ( $w_0 = 2$ ) and eight population models for the large population  $N^{(1)} = 6,000,000$  with a sample size of 5,000.*

Estimator	NULL	XRE	LIN <sub>0</sub>	LIN <sub>1</sub>	QUAD <sub>0</sub>	QUAD <sub>1</sub>	EXP <sub>0</sub>	EXP <sub>1</sub>
<i>psm</i>	1.90	2.11	1.22	1.54	2.00	2.47	2.06	2.33
<i>unw</i>	0.51	1.40	0.41	1.23	16.89	16.92	4.74	5.03
<i>trim</i>	0.64	1.36	0.56	1.30	7.87	7.93	3.11	3.10
<i>xre</i>	0.53	0.63	1.19	1.20	5.33	4.51	2.41	2.17
<i>xre-greg</i>	0.66	0.74	1.22	1.23	4.24	3.75	2.33	2.11
<i>lin</i>	0.54	0.63	0.29	0.33	8.76	5.99	3.52	2.32
<i>lin-greg</i>	0.67	0.74	0.4	0.45	5.22	4.09	2.20	1.69
<i>npar</i>	0.84	0.92	0.53	0.62	1.85	2.18	2.04	1.69
<i>npar-greg</i>	0.95	1.01	0.63	0.70	1.66	1.95	1.77	1.55

TABLE 14

*Mean squared error ( $\times 10^4$ ) of nine estimators ( $w_0 = 3$ ) and eight population models for the large population  $N^{(1)} = 6,000,000$  with a sample size of 5,000.*

Estimator	NULL	XRE	LIN <sub>0</sub>	LIN <sub>1</sub>	QUAD <sub>0</sub>	QUAD <sub>1</sub>	EXP <sub>0</sub>	EXP <sub>1</sub>
<i>psm</i>	1.90	2.11	1.22	1.54	2.00	2.47	2.06	2.33
<i>unw</i>	0.51	1.40	0.41	1.23	16.89	16.92	4.74	5.03
<i>trim</i>	0.75	1.34	0.91	1.54	6.2	6.34	2.95	2.95
<i>xre</i>	0.53	0.63	1.19	1.20	5.33	4.51	2.41	2.17
<i>xre-greg</i>	0.79	0.85	1.25	1.26	3.75	3.42	2.32	2.13
<i>lin</i>	0.54	0.63	0.29	0.33	8.76	5.99	3.52	2.32
<i>lin-greg</i>	0.79	0.84	0.48	0.53	4.11	3.40	1.93	1.60
<i>npar</i>	0.84	0.92	0.53	0.62	1.85	2.18	2.04	1.69
<i>npar-greg</i>	1.06	1.09	0.70	0.77	1.69	1.94	1.76	1.59

TABLE 15

*Mean squared error ( $\times 10^4$ ) of nine estimators ( $w_0 = 5$ ) and eight population models for the large population  $N^{(1)} = 6,000,000$  with a sample size of 5,000.*

Estimator	NULL	XRE	LIN <sub>0</sub>	LIN <sub>1</sub>	QUAD <sub>0</sub>	QUAD <sub>1</sub>	EXP <sub>0</sub>	EXP <sub>1</sub>
<i>psm</i>	1.90	2.11	1.22	1.54	2.00	2.47	2.06	2.33
<i>unw</i>	0.51	1.40	0.41	1.23	16.89	16.92	4.74	5.03
<i>trim</i>	1.06	1.50	1.12	1.56	3.58	3.90	2.31	2.43
<i>xre</i>	0.53	0.63	1.19	1.20	5.33	4.51	2.41	2.17
<i>xre-greg</i>	1.07	1.02	1.28	1.29	2.95	2.86	2.22	2.10
<i>lin</i>	0.54	0.63	0.29	0.33	8.76	5.99	3.52	2.32
<i>lin-greg</i>	1.07	1.02	0.68	0.71	2.84	2.61	1.72	1.58
<i>npar</i>	0.84	0.92	0.53	0.62	1.85	2.18	2.04	1.69
<i>npar-greg</i>	1.30	1.22	0.85	0.90	1.77	1.97	1.80	1.70

TABLE 16

*Mean squared error ( $\times 10^4$ ) of nine estimators ( $w_0 = 10$ ) and eight population models for the large population  $N^{(1)} = 6,000,000$  with a sample size of 5,000.*

Estimator	NULL	XRE	LIN <sub>0</sub>	LIN <sub>1</sub>	QUAD <sub>0</sub>	QUAD <sub>1</sub>	EXP <sub>0</sub>	EXP <sub>1</sub>
<i>psm</i>	1.90	2.11	1.22	1.54	2.00	2.47	2.06	2.33
<i>unw</i>	0.51	1.40	0.41	1.23	16.89	16.92	4.74	5.03
<i>trim</i>	1.45	1.79	1.52	1.84	2.29	2.68	2.16	2.31
<i>xre</i>	0.53	0.63	1.19	1.20	5.33	4.51	2.41	2.17
<i>xre-greg</i>	1.45	1.31	1.34	1.35	2.33	2.42	2.24	2.16
<i>lin</i>	0.54	0.63	0.29	0.33	8.76	5.99	3.52	2.32
<i>lin-greg</i>	1.45	1.31	0.95	0.99	2.03	2.11	1.80	1.74
<i>npar</i>	0.84	0.92	0.53	0.62	1.85	2.18	2.04	1.69
<i>npar-greg</i>	1.57	1.41	1.04	1.09	1.89	2.06	1.92	1.87

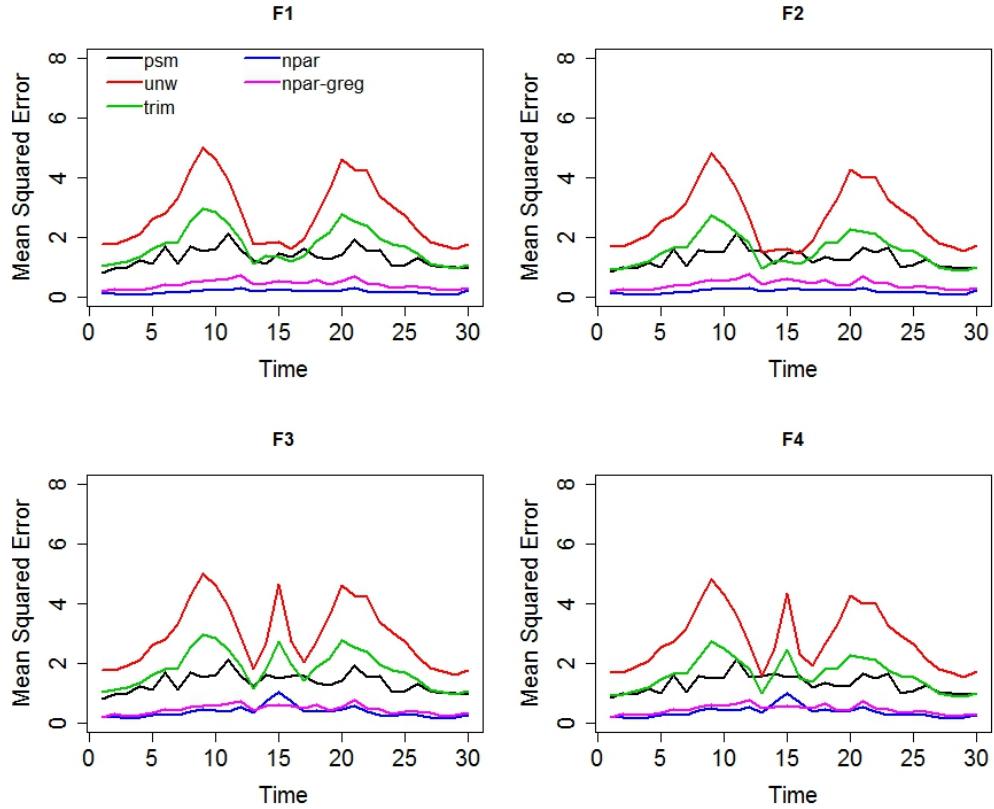


FIG 37. Mean squared errors  $\times 10^4$  at each timepoint for population models  $F_1$ ,  $F_2$ ,  $F_3$  and  $F_4$  with population size  $N^{(1)} = 6,000,000$  and sample size 5,000. The value of  $w_0$  is 2.

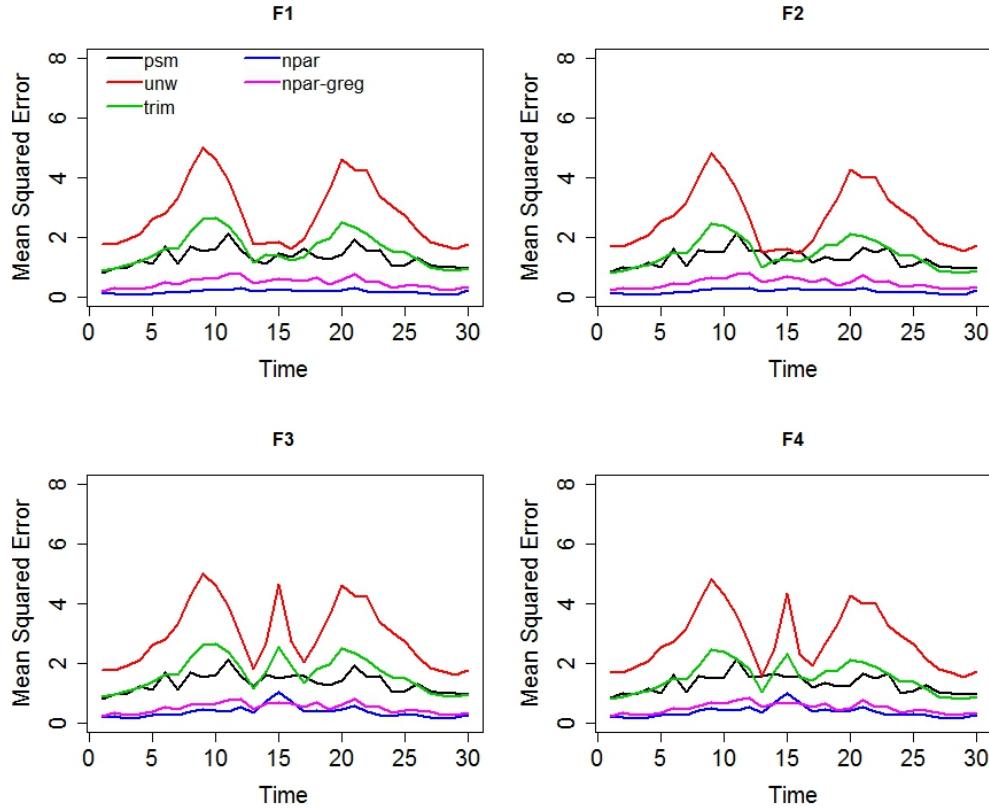


FIG 38. Mean squared errors  $\times 10^4$  at each timepoint for population models  $F_1$ ,  $F_2$ ,  $F_3$  and  $F_4$  with population size  $N^{(1)} = 6,000,000$  and sample size 5,000. The value of  $w_0$  is 3.

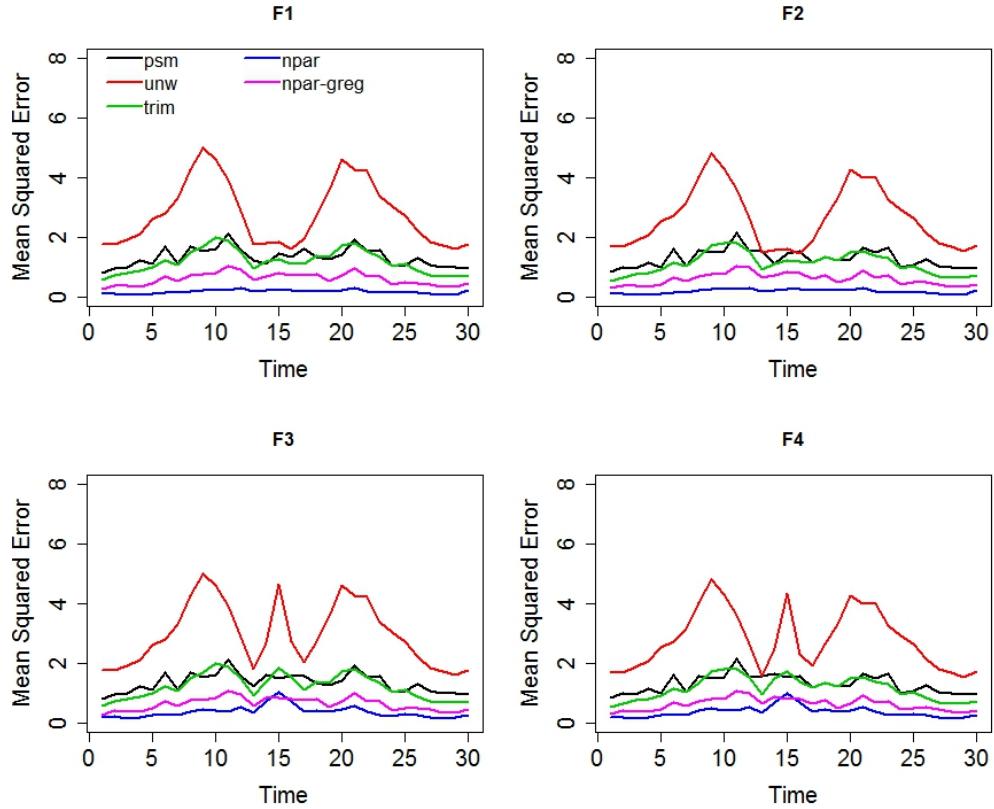


FIG 39. Mean squared errors  $\times 10^4$  at each timepoint for population models  $F_1$ ,  $F_2$ ,  $F_3$  and  $F_4$  with population size  $N^{(1)} = 6,000,000$  and sample size 5,000. The value of  $w_0$  is 5.

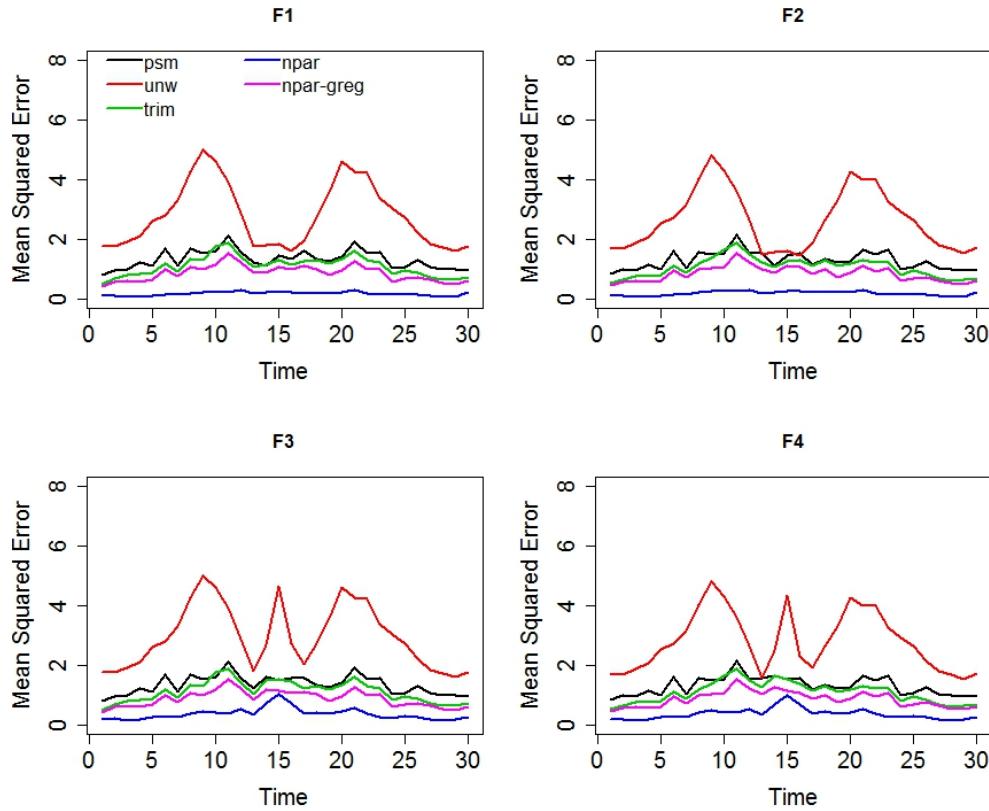


FIG 40. Mean squared errors  $\times 10^4$  at each timepoint for population models  $F_1$ ,  $F_2$ ,  $F_3$  and  $F_4$  with population size  $N^{(1)} = 6,000,000$  and sample size 5,000. The value of  $w_0$  is 10.

*Additional GIS results with respect to varying values of  $w_0$ .* We also investigate the Great Influenza Survey (see Section 2 and Section 6) with other values of  $w_0$ . Again, we used values of 2, 3, 5 and 10 for  $w_0$ . Results with regard to the prevalence estimation are given in Tables 17–20. For the *trim* estimator the influence of  $w_0$  on the obtained estimates is large, namely 5.49% for  $w_0 = 2$  and 6.66% for  $w_0 = 10$ . For the *lin-greg* and *npar-greg* estimators the influence is smaller with estimates ranging from 6.80% for  $w_0 = 2$  to 7.10% for  $w_0 = 10$ .

In Figures 41–44 we present the results of the incidence estimation for different values of  $w_0$ . The influence of  $w_0$  is higher when compared to the prevalence results. For example, the estimated incidence of the *npar-greg* estimator with a value of  $w_0 = 10$  would not be useful for practical usage because it is too variable.

TABLE 17  
*Estimates and 95% confidence intervals (CI) of the overall prevalence of ILI during the 2010-2011 influenza season based on nine estimators. The value of  $w_0$  is 2.*

Method	Estimate	Analytical CI	Bootstrap CI	Jackknife CI
<i>psm</i>	7.10%	[5.31-9.45%]	-	-
<i>unw</i>	5.12%	[4.52-5.80%]	-	-
<i>trim</i>	5.49%	[4.81-6.26%]	-	-
<i>xre</i>	5.96%	[4.95-7.17%]	[5.00-7.10%]	-
<i>xre-greg</i>	6.10%	-	-	[4.84-7.57%]
<i>lin</i>	6.88%	[5.69-8.30%]	[5.76-8.21%]	-
<i>lin-greg</i>	6.80%	-	-	[5.67-8.14%]
<i>npar</i>	6.88%	[5.69-8.30%]	[5.61-8.41%]	-
<i>npar-greg</i>	6.80%	-	-	[5.67-8.14%]

TABLE 18  
*Estimates and 95% confidence intervals (CI) of the overall prevalence of ILI during the 2010-2011 influenza season based on nine estimators. The value of  $w_0$  is 3.*

Method	Estimate	Analytical CI	Bootstrap CI	Jackknife CI
<i>psm</i>	7.10%	[5.31-9.45%]	-	-
<i>unw</i>	5.12%	[4.52-5.80%]	-	-
<i>trim</i>	5.61%	[4.88-6.46%]	-	-
<i>xre</i>	5.96%	[4.95-7.17%]	[5.00-7.10%]	-
<i>xre-greg</i>	6.16%	-	-	[4.92-7.68%]
<i>lin</i>	6.88%	[5.69-8.30%]	[5.76-8.21%]	-
<i>lin-greg</i>	6.82%	-	-	[5.66-8.20%]
<i>npar</i>	6.88%	[5.69-8.30%]	[5.61-8.41%]	-
<i>npar-greg</i>	6.82%	-	-	[5.66-8.20%]

TABLE 19

*Estimates and 95% confidence intervals (CI) of the overall prevalence of ILI during the 2010-2011 influenza season based on nine estimators. The value of  $w_0$  is 5.*

Method	Estimate	Analytical CI	Bootstrap CI	Jackknife CI
<i>psm</i>	7.10%	[5.31-9.45%]	-	-
<i>unw</i>	5.12%	[4.52-5.80%]	-	-
<i>trim</i>	6.00%	[5.09-7.07%]	-	-
<i>xre</i>	5.96%	[4.95-7.17%]	[5.00-7.10%]	-
<i>xre-greg</i>	6.37%	-	-	[5.10-7.94%]
<i>lin</i>	6.88%	[5.69-8.30%]	[5.76-8.21%]	-
<i>lin-greg</i>	6.87%	-	-	[5.64-8.35%]
<i>npar</i>	6.88%	[5.69-8.30%]	[5.61-8.41%]	-
<i>npar-greg</i>	6.87%	-	-	[5.64-8.35%]

TABLE 20

*Estimates and 95% confidence intervals (CI) of the overall prevalence of ILI during the 2010-2011 influenza season based on nine estimators. The value of  $w_0$  is 10.*

Method	Estimate	Analytical CI	Bootstrap CI	Jackknife CI
<i>psm</i>	7.10%	[5.31-9.45%]	-	-
<i>unw</i>	5.12%	[4.52-5.80%]	-	-
<i>trim</i>	6.66%	[5.40-8.17%]	-	-
<i>xre</i>	5.96%	[4.95-7.17%]	[5.00-7.10%]	-
<i>xre-greg</i>	6.74%	-	-	[5.36-8.43%]
<i>lin</i>	6.88%	[5.69-8.30%]	[5.76-8.21%]	-
<i>lin-greg</i>	7.10%	-	-	[5.70-8.82%]
<i>npar</i>	6.88%	[5.69-8.30%]	[5.61-8.41%]	-
<i>npar-greg</i>	7.10%	-	-	[5.70-8.82%]

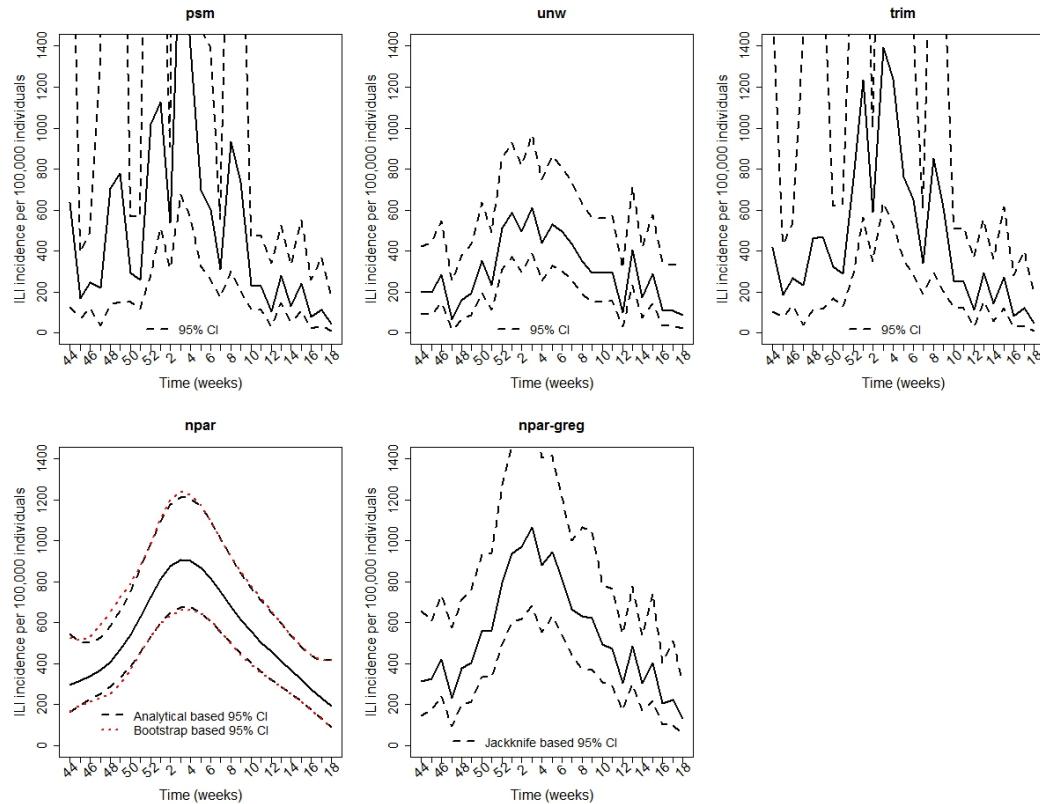


FIG 41. Estimated incidence trends with corresponding 95% pointwise confidence intervals from the 2010/2011 Great Influenza Survey (GIS) using five estimation methods. The value of  $w_0$  is 2.

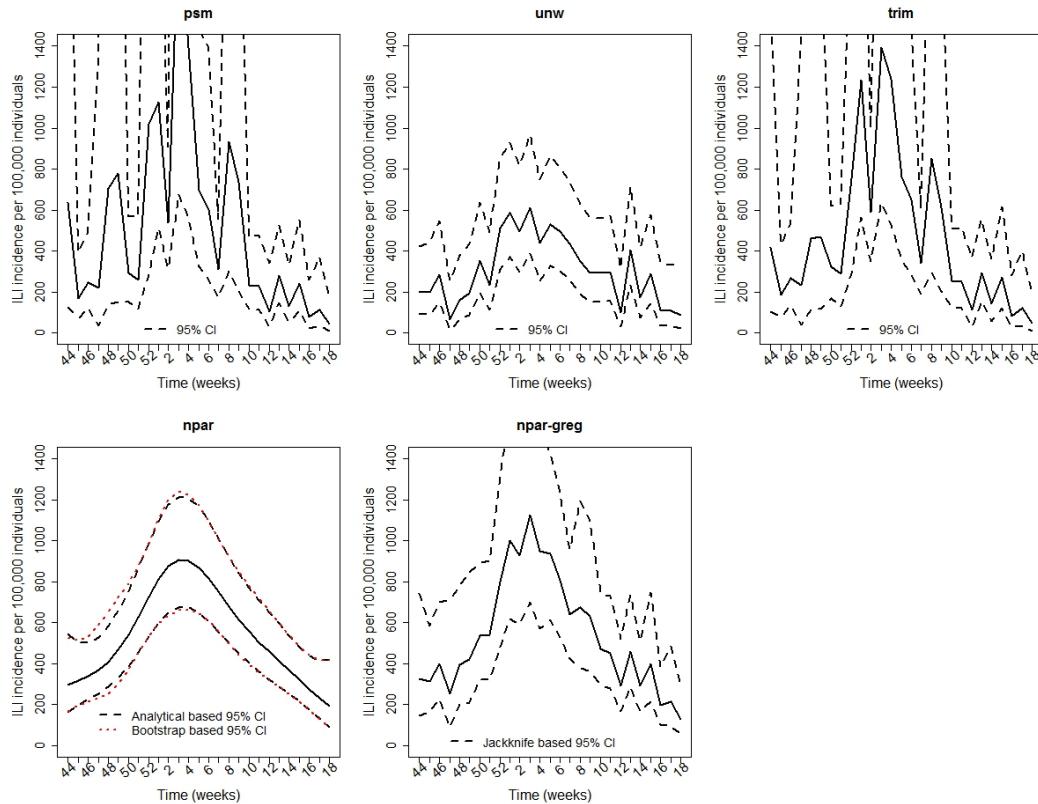


FIG 42. Estimated incidence trends with corresponding 95% pointwise confidence intervals from the 2010/2011 Great Influenza Survey (GIS) using five estimation methods. The value of  $w_0$  is 3.

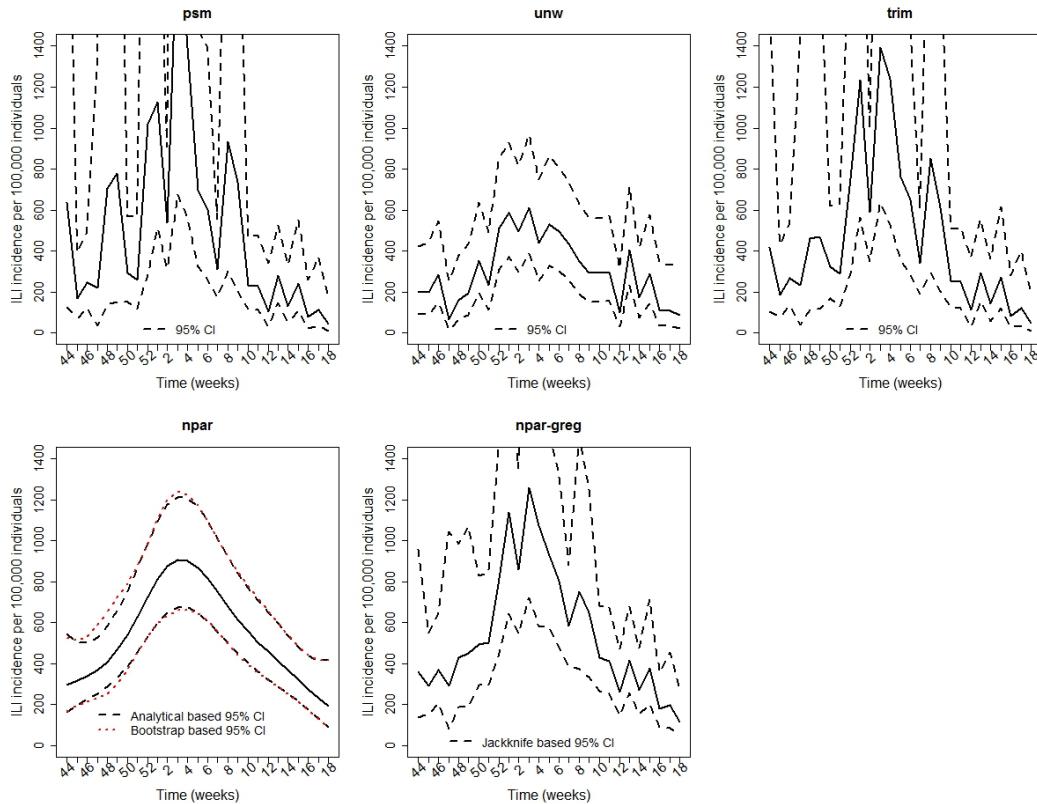


FIG 43. Estimated incidence trends with corresponding 95% pointwise confidence intervals from the 2010/2011 Great Influenza Survey (GIS) using five estimation methods. The value of  $w_0$  is 5.

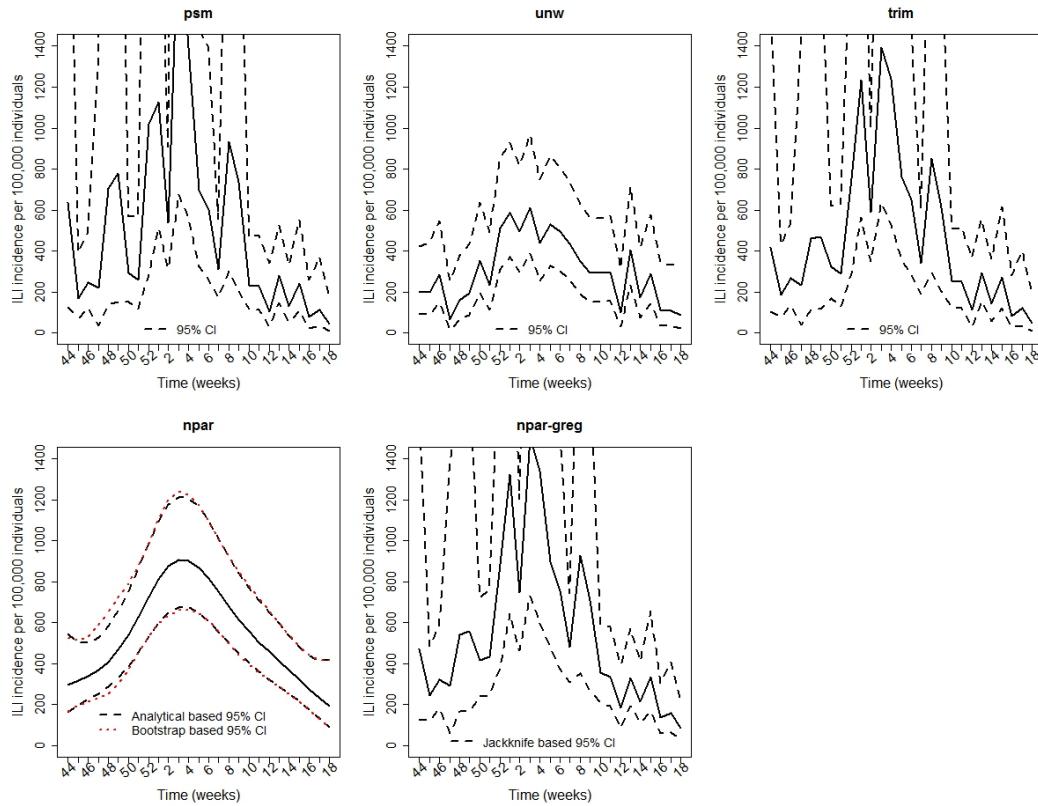


FIG 44. Estimated incidence trends with corresponding 95% pointwise confidence intervals from the 2010/2011 Great Influenza Survey (GIS) using five estimation methods. The value of  $w_0$  is 10.

*Additional simulation results for small sample size.* The smallest sample size considered in the simulation study is 500. In this section, we investigate in more detail the performance of the proposed methodology for samples with an even smaller sample size. The population consists of 18 strata with strata and sample sizes given in Table 21. The total population size is 200,000 and the total sample size is 150. The highest post-stratification weight is 30. Note that the population and sample settings used in the paper needed to be adjusted to obtain large post-stratification weights for smaller sample sizes. Population values are generated from the same eight models as described in Section 5.1 of the paper. The population means of the generated data sets are again estimated using nine estimators.

Table 22 summarizes the MSE results for 100 simulated datasets. We observe that the *npar* and *npar-greg* estimators still perform adequately but do not perform the best in any setting. Interestingly, none of the considered estimators is optimal. Thus, we can conclude that for small sample sizes the estimators *npar* and *npar-greg* deteriorate and more care needs to be taken when using these estimators. For smaller sample sizes it is much harder to fit a non-parametric regression model to the data such that variability of the estimator increases which explains why the *npar* and *npar-greg* estimators do not perform as well as in the simulation study in the paper. The fact that for smaller sample sizes the estimators *npar* and *npar-greg* deteriorate is a limitation of our proposed work. Therefore, care needs to be taken when using the proposed methods with small sample sizes.

TABLE 21  
*Population and sample sizes in the 18 strata for the population used in the simulation study.*

Stratum $h$	1	2	3	4	5	6	7	8	9
$\bar{N}_h$	20000	15000	5000	5000	5000	2500	2500	2500	2500
$n_h$	2	3	3	5	7	10	15	20	20
Stratum $h$	10	11	12	13	14	15	16	17	18
$\bar{N}_h$	2500	2500	5000	5000	10000	10000	25000	40000	40000
$n_1$	15	15	10	10	5	4	3	2	1

TABLE 22

*Mean squared error ( $\times 10^4$ ) of nine estimators and eight population models for a sample size of 150. Based on 100 simulated datasets.*

Estimator	NULL	XRE	LIN <sub>0</sub> <sup>†</sup>	LIN <sub>1</sub>	QUAD <sub>0</sub>	QUAD <sub>1</sub>	EXP <sub>0</sub>	EXP <sub>1</sub>
<i>psm</i>	173.8	150.2	114.0	117.3	187.2	188.9	185.4	187.4
<i>unw</i>	17.9	16.3	213.2	215.0	38.2	38.5	21.7	21.1
<i>trim</i>	49.0	31.6	110.8	112.6	51.2	49.3	45.5	45.6
<i>xre</i>	18.7	13.9	160.5	163.8	36.5	36.4	24.3	24.1
<i>xre-greg</i>	34.6	21.8	125.9	127.2	42.7	41.1	41.1	42.6
<i>lin</i>	30.7	29.7	15.9	17.3	92.3	97.6	35.4	37.7
<i>lin-greg</i>	46.3	35.5	26.8	28.8	80.2	80.1	48.1	50.4
<i>npar</i>	72.2	40.1	45.1	46.3	101.4	99.8	70.7	70.9
<i>npar-greg</i>	90.5	48.4	54.8	57.8	108.3	104.8	86.1	86.6

*Model fits and results on additional spline basis functions.* The simulation study results show that the *npar* estimator yields smaller MSE values than the *npar-greg* in the NULL, XRE, and LIN scenarios. However, the *npar-greg* performs better than the *npar* in the QUAD and EXP scenarios. This may suggest that the NPAR weight smoothing model does not fit the data generated by the QUAD and EXP models well. To investigate this we perform a small simulation study using different types of splines and make some plots to check how the model fitted the simulated data and the GIS data.

We investigate the impact of different types of splines. The types of splines we considered are the approximating thin plate spline (used in the paper), the linear truncated splines and cubic B-spline (with third order difference penalty). Simulation results using these splines basis functions indicate that the results are qualitatively similar over the different types of splines. In Table 23, we present the MSE results for the simulation study for  $N^{(1)}$  with a sample size of 5,000 (see Table 2 in the paper). We observe that the results between the different types of splines do not differ much. Also the results about the nominal coverage (not shown) are very similar.

In Figure 45, we present a plot of how the model fits the data for two simulated data sets using the EXP<sub>0</sub> scenario. We observe that the model fits the data adequately. We also observe that there is little difference in the fits using the different types of splines.

Finally, In Figure 46 we present the model fit of the NPAR weight smoothing model to the GIS data (see Section 6 in the paper). It is observed that the model fits the data adequately.

TABLE 23  
*Mean squared errors ( $\times 10^4$ ) of the simulation study for the large population  $N^{(1)} = 6,000,000$  with a sample size of 5,000. <sup>†</sup>Based on 249 simulated datasets.*

Estimator	NULL	XRE	LIN <sub>0</sub> <sup>†</sup>	LIN <sub>1</sub>	QUAD <sub>0</sub>	QUAD <sub>1</sub>	EXP <sub>0</sub>	EXP <sub>1</sub>
<i>psm</i>	1.70	2.85	1.16	1.73	1.88	2.92	1.82	2.90
<i>unw</i>	0.46	1.83	0.43	1.54	16.09	17.52	4.16	5.42
<i>trim</i>	0.68	1.92	0.86	1.68	5.61	6.50	2.41	3.26
<i>xre</i>	0.47	0.61	1.01	1.04	4.64	4.11	1.83	1.75
<i>xre-greg</i>	0.67	0.84	1.06	1.09	3.14	2.90	1.73	1.68
<i>lin</i>	0.48	0.61	0.30	0.35	8.35	5.59	2.97	2.04
<i>lin-greg</i>	0.67	0.83	0.50	0.53	3.66	2.96	1.49	1.32
<i>npar</i>	0.65	0.90	0.48	0.57	1.63	1.71	1.74	1.49
<i>npar-greg</i>	0.85	1.08	0.66	0.71	1.52	1.57	1.49	1.40
<i>npar</i>	0.65	0.93	0.50	0.55	1.50	1.59	1.71	1.49
<i>npar-greg</i>	0.85	1.10	0.67	0.69	1.43	1.49	1.49	1.40
<i>npar</i>	0.64	0.89	0.45	0.54	1.67	1.67	1.71	1.43
<i>npar-greg</i>	0.83	1.05	0.62	0.68	1.51	1.52	1.44	1.33

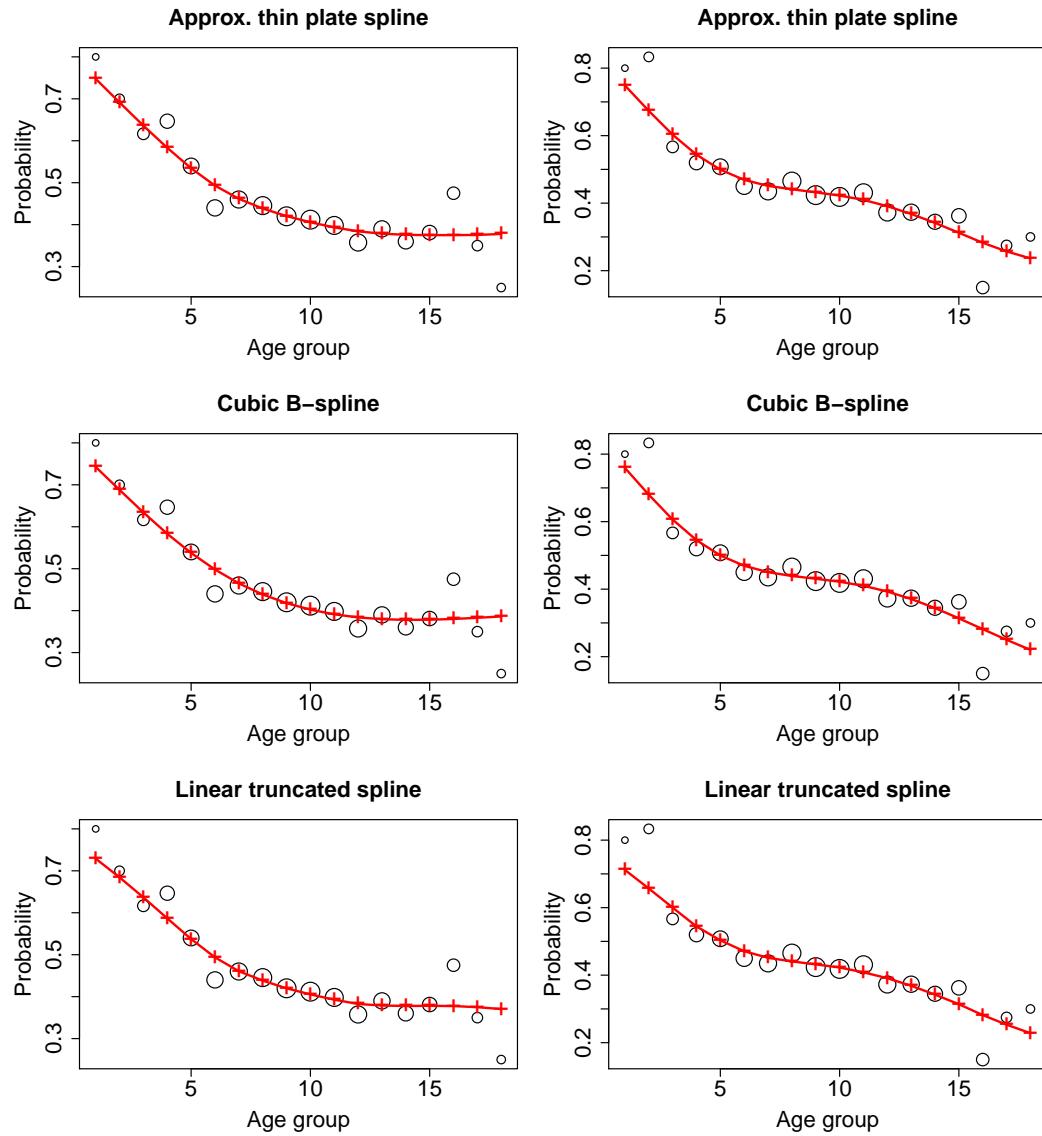


FIG 45. Model fit of three types of splines to two simulated datasets. The dots represent the simulated data where the size of the dot is proportional to sample size in the post-strata. The red crosses represent the model fits.

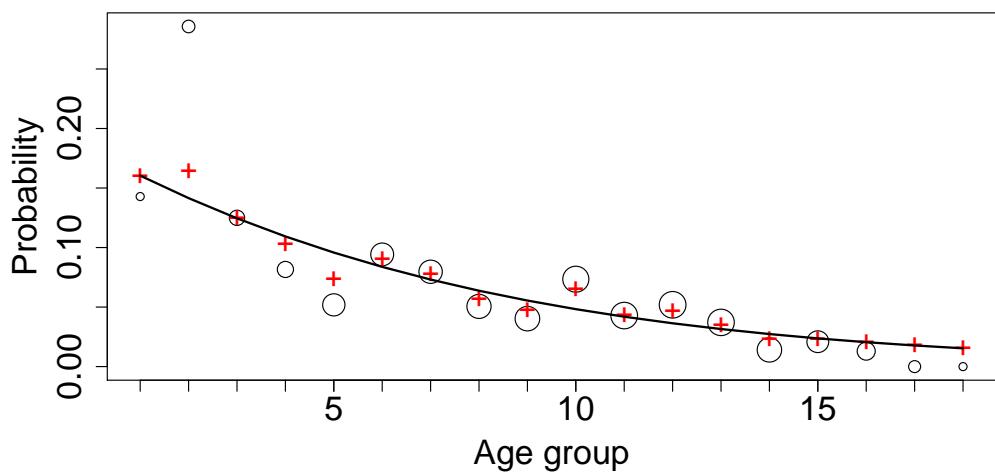


FIG 46. Model fit of the NPAR weight smoothing model to the 2010-2011 GIS data. The dots represent the GIS data where the size of the dot is proportional to the sample size in the age group. The black line represents the model fit without the random effects and the red crosses with the random effects.

**References.**

- HASTIE, T. J. and TIBSHIRANI, R. J. (1990). *Generalized Additive Models*. London: Chapman and Hall.
- SAS INSTITUTE INC. (2008). *The GLIMMIX procedure*. Cary, NC: SAS Institute Inc.
- MOLENBERGHS, G. and VERBEKE, G. (2005). *Models for Discrete Longitudinal Data*. New York: Springer.
- RUPPERT, D., WAND, M. P. and CARROLL, R. J. (2003). *Semiparametric Regression*. Cambridge: University Press.
- SEARLE, S. R. (1982). *Matrix Algebra useful for Statistics*. New York: Wiley-Interscience.
- R CORE TEAM (2014). R: A Language and Environment for Statistical Computing R Foundation for Statistical Computing, Vienna, Austria.
- WOLFINGER, R. D. and O'CONNELL, M. (1993). A pseudo-likelihood approach. *Journal of Statistical Computation and Simulation* **48** 233-243.