A NOTE ABOUT THE R COEFFICIENT, THE INTRACLASS CORRELATION COEFFICIENT AND THEIR ASSOCIATION WITH OUTCOME PREVALENCE

Ariane Murielle MBEKWE YEPNANG
Agnès Caille
Sandra M. Eldridge
Bruno Giraudeau

Université de Tours
INSERM UMR 1246 – SPHERE

Current developments in cluster randomised trials and stepped wedge designs
London, November 21, 2018

Plan

Analyse a cluster randomized trials $\longrightarrow$ Report the ICC[1]

[1] Campbell MK et al., *British Medical Journal*, 328(7441):702–708, 2004

Analyse a cluster randomized trials $\longrightarrow$ Report the ICC[1]

- k: number of clusters
- X_{ij}: binary outcome of the jth individual in the ith cluster
- π: outcome prevalence

[1] Campbell MK et al., *British Medical Journal*, 328(7441):702–708, 2004

Analyse a cluster randomized trials $\longrightarrow$ Report the ICC[1]

- k: number of clusters
- X_{ij}: binary outcome of the jth individual in the ith cluster
- π: outcome prevalence

Common correlation model[2]:

$\rho = \text{cor}(X_{ij}, X_{ij'})$ for all $i = 1, 2, ..., k$ and $j \neq j'$

[1] Campbell MK et al., *British Medical Journal*, 328(7441):702–708, 2004
[2] Eldridge SM et al., *International Statistical Review*, 77(3):378–394, 2009

Analyse a cluster randomized trials $\longrightarrow$ Report the ICC[1]

- k: number of clusters
- X_{ij}: binary outcome of the jth individual in the ith cluster
- π: outcome prevalence

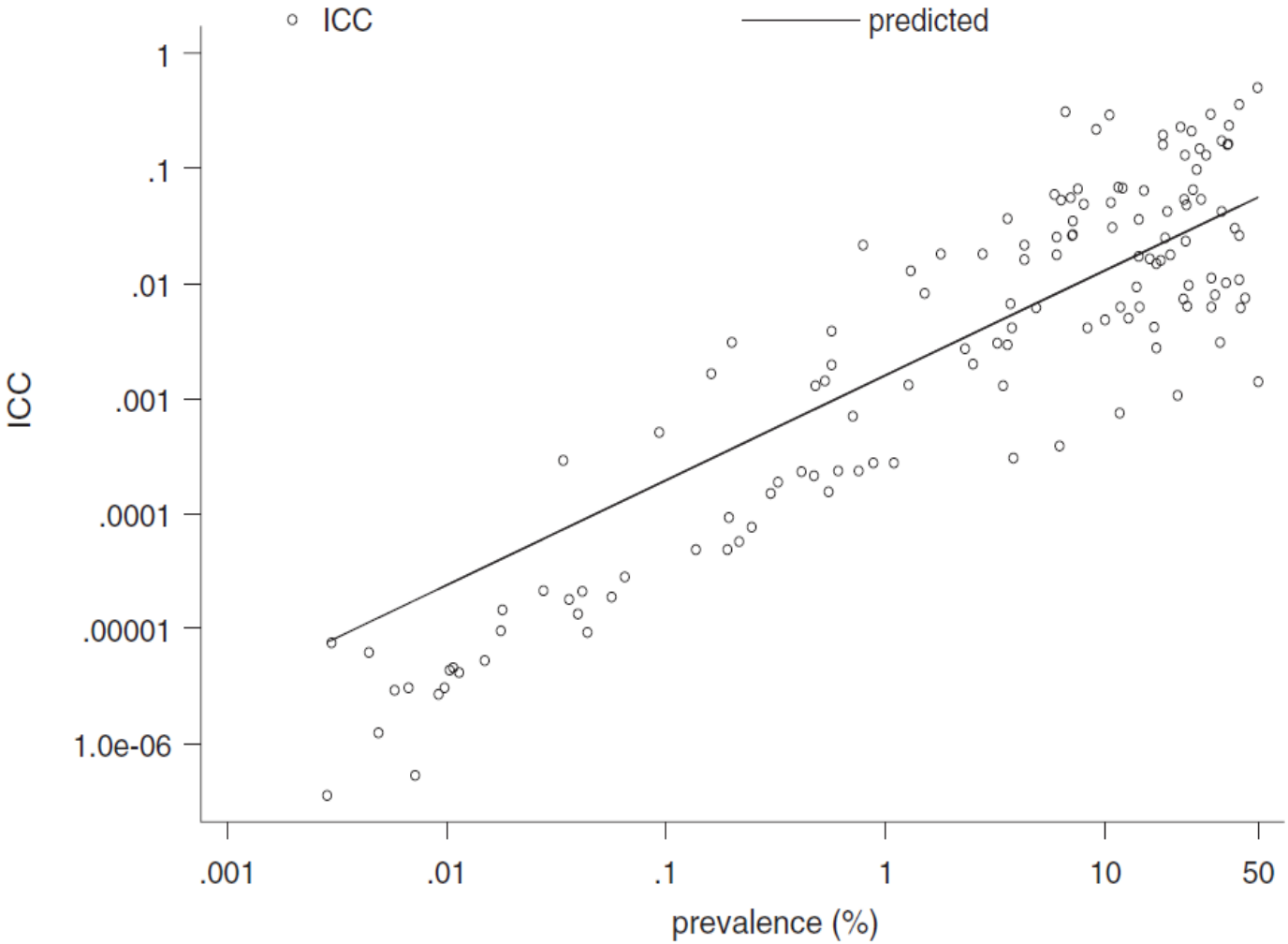
Common correlation model[2]:

$$\rho = \text{cor}(X_{ij}, X_{ij'}) \text{ for all } i = 1, 2, ..., k \text{ and } j \neq j'$$

$$\rho = \frac{\text{P}(X_{ij}=1|X_{ij'}=1)-\pi}{1-\pi}$$

[1] Campbell MK et al., *British Medical Journal*, 328(7441):702–708, 2004
[2] Eldridge SM et al., *International Statistical Review*, 77(3):378–394, 2009

Binary outcome → Association between the outcome prevalence and the ICC[3]



[3]Gulliford M et al., *Journal of clinical epidemiology*, 58(3):246–251, 2005

Plan

- As defined by Rosner[4]:

$$\mathrm{P}(X_{ij} = 1 | X_{ij'} = 1) = R\pi, \quad j, j' = 1, 2 \text{ and } j \neq j'$$

[4] Rosner, *Biometrics*, 105–114, 1982

[5] Crespi et al., *Clinical Trials*, 8(6):687–698, 2011

- As defined by Rosner[4]:

$$P(X_{ij} = 1 | X_{ij'} = 1) = R\pi, \quad j, j' = 1, 2 \text{ and } j \neq j'$$

- As extended by Crespi et al.[5]: $j, j' = 1, 2, ..., n$ and $j \neq j'$

[4] Rosner, *Biometrics*, 105–114, 1982

[5] Crespi et al., *Clinical Trials*, 8(6):687–698, 2011

- As defined by Rosner[4]:

$$P(X_{ij} = 1 | X_{ij'} = 1) = R\pi, \quad j, j' = 1, 2 \text{ and } j \neq j'$$

- As extended by Crespi et al.[5]: $j, j' = 1, 2, ..., n$ and $j \neq j'$
- Interpretation: a quantification of how much or less likely a member of a cluster is to have success given that another member of the cluster has success

[4] Rosner, *Biometrics*, 105–114, 1982

[5] Crespi et al., *Clinical Trials*, 8(6):687–698, 2011

- As defined by Rosner[4]:

$$P(X_{ij} = 1 | X_{ij'} = 1) = R\pi, \quad j, j' = 1, 2 \text{ and } j \neq j'$$

- As extended by Crespi et al.[5]: $j, j' = 1, 2, ..., n$ and $j \neq j'$
- Interpretation: a quantification of how much or less likely a member of a cluster is to have success given that another member of the cluster has success
- $1 \leq R \leq \frac{1}{\pi}$

[4] Rosner, *Biometrics*, 105–114, 1982

[5] Crespi et al., *Clinical Trials*, 8(6):687–698, 2011

- As defined by Rosner[4]:

$$P(X_{ij} = 1 | X_{ij'} = 1) = R\pi, \quad j, j' = 1, 2 \text{ and } j \neq j'$$

- As extended by Crespi et al.[5]: $j, j' = 1, 2, ..., n$ and $j \neq j'$
- Interpretation: a quantification of how much or less likely a member of a cluster is to have success given that another member of the cluster has success
- $1 \leq R \leq \frac{1}{\pi}$
- $R = 1 + \frac{\rho(1-\pi)}{\pi}$

[4] Rosner, *Biometrics*, 105–114, 1982
[5] Crespi et al., *Clinical Trials*, 8(6):687–698, 2011

- R is less influenced by the outcome prevalence than the ICC[6]
- Illustration with an example

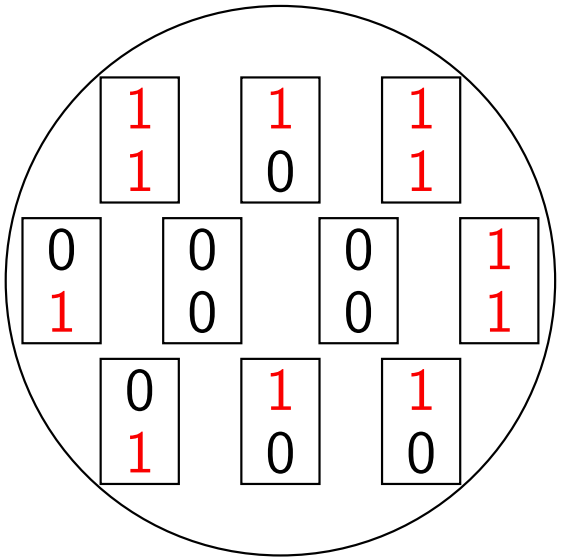
[6] Crespi et al., *Clinical Trials*, 8(6):687–698, 2011

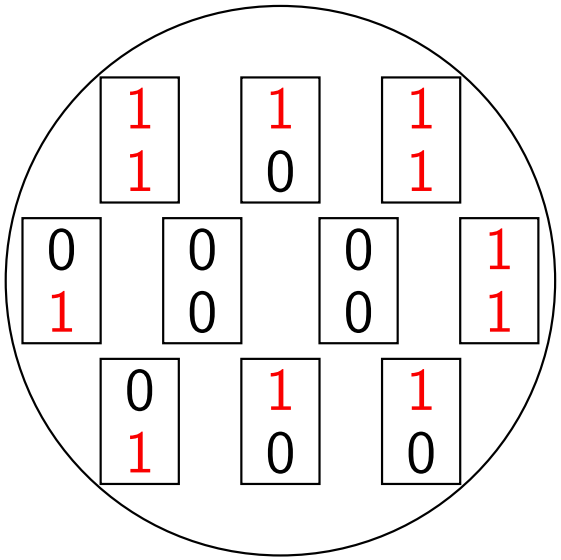
Plan

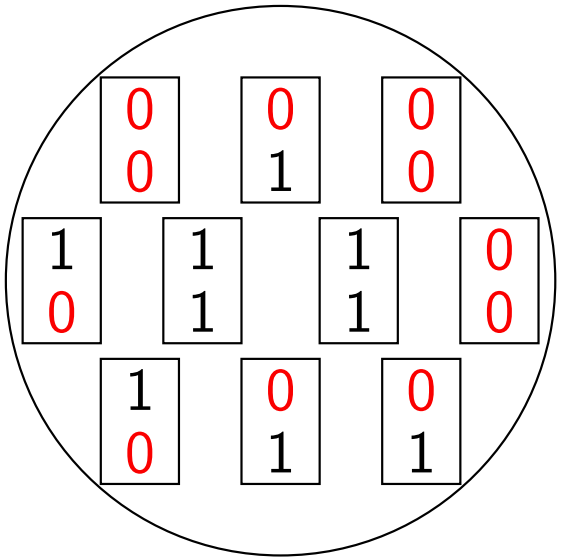
To explore the relation between the ICC, the R coefficient and the outcome prevalence.

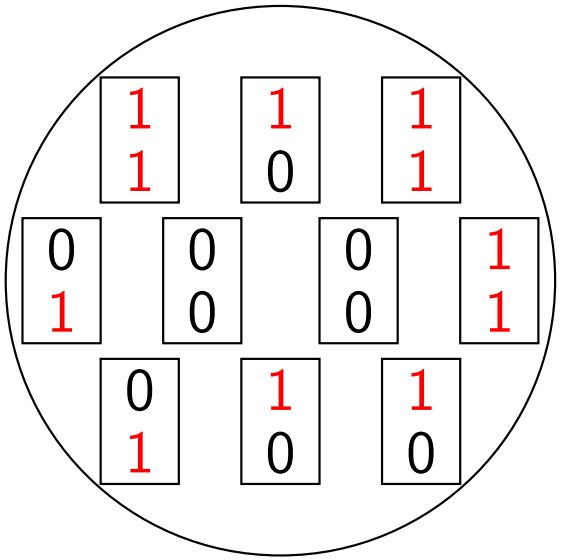
1. Through mathematical developments
2. Through a simulation study

Plan

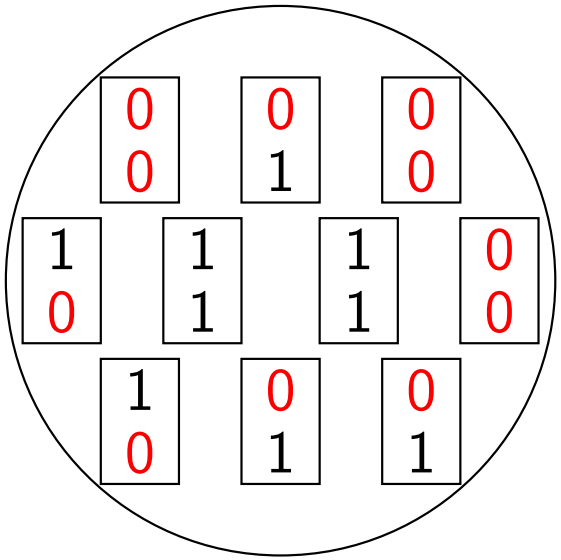

1 1
1 0
1 1
0 1
0 0
0 0
1 1
0 1
1 0
1 0


1 1
1 0
1 1
0 1
0 0
0 0
1 1
0 1
1 0
1 0

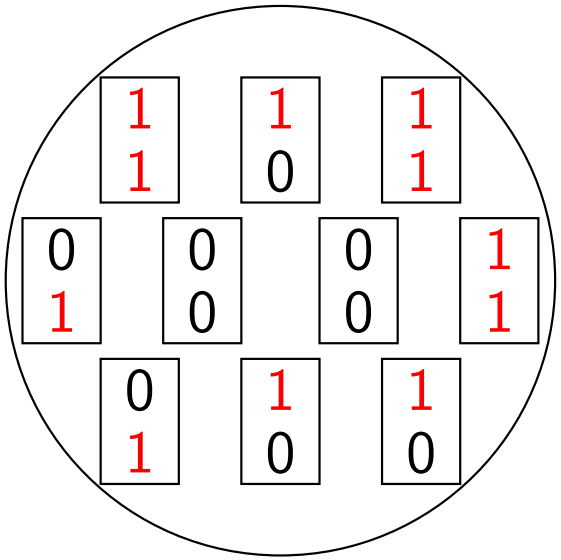

0 0
0 1
0 0
1 0
1 1
1 1
0 0
1 0
0 1
0 1



Prevalence=$\widehat{\pi}$

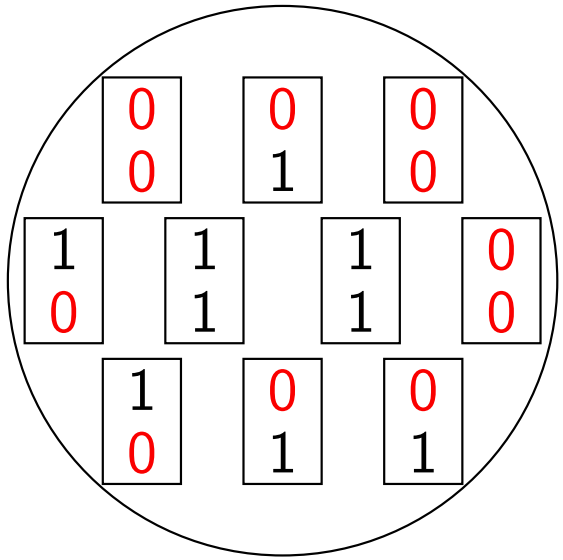


Prevalence=$1 - \widehat{\pi}$



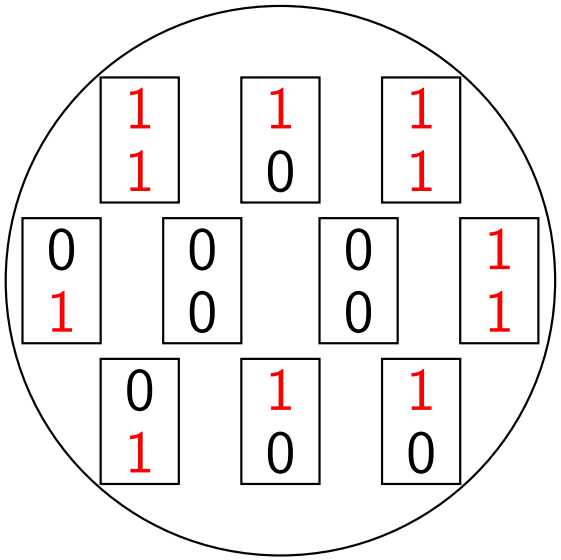
Prevalence$=\widehat{\pi}$

ICC$=\widehat{\rho}$



Prevalence$=1-\widehat{\pi}$

ICC$=\widehat{\rho}$

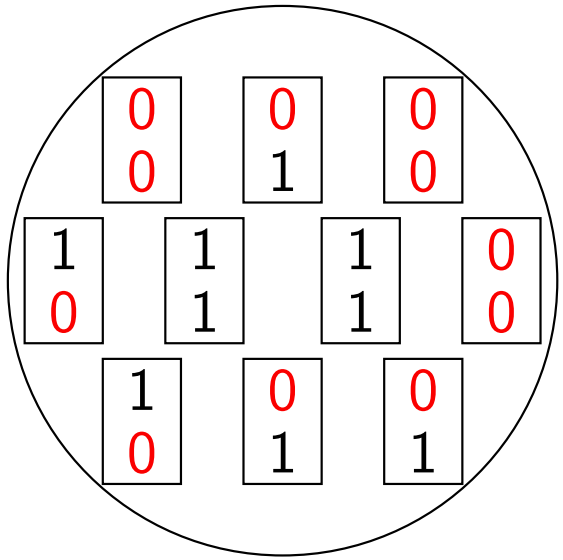


Prevalence=$\widehat{\pi}$

ICC=$\widehat{\rho}$

R coefficient=$\widehat{R}_1$

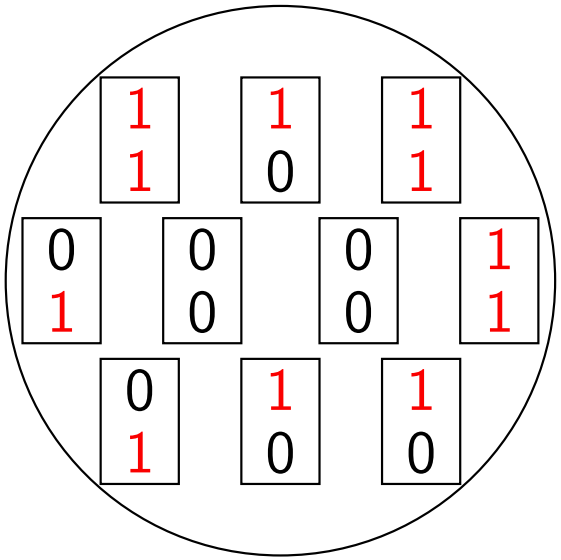
$\widehat{R}_1 \neq \widehat{R}_2$

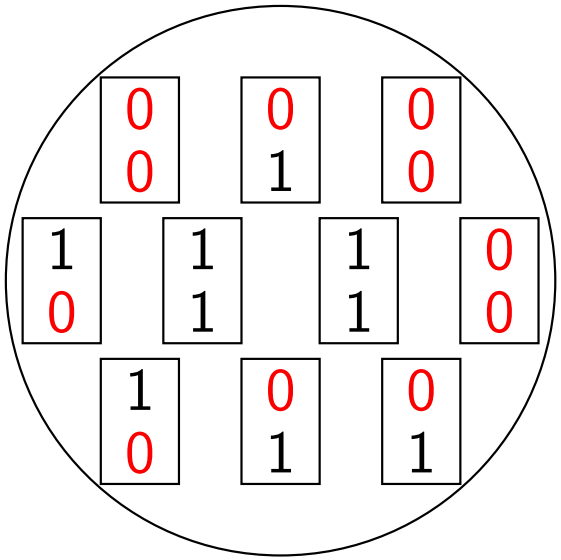


Prevalence=$1 - \widehat{\pi}$

ICC=$\widehat{\rho}$

R coefficient=$\widehat{R}_2$





Prevalence=$\widehat{\pi}$

ICC=$\widehat{\rho}$

R coefficient=$\widehat{R}_1$

$\widehat{R}_1 \neq \widehat{R}_2$

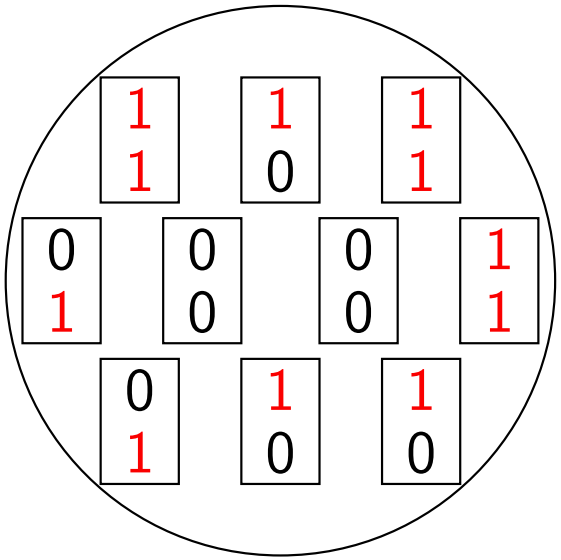
Prevalence=$1 - \widehat{\pi}$

ICC=$\widehat{\rho}$

R coefficient=$\widehat{R}_2$

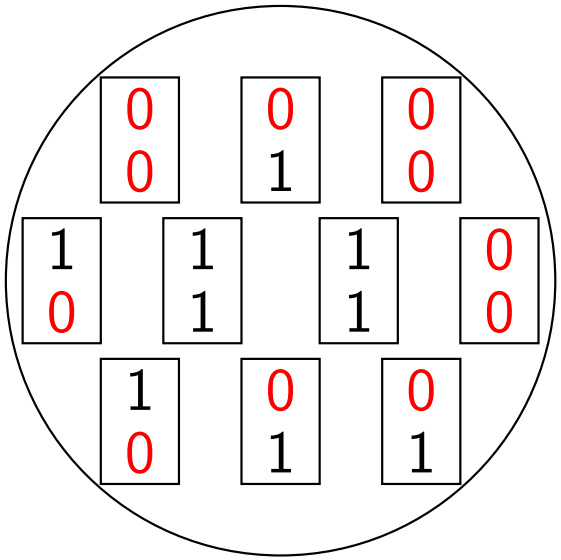
Asymmetry of R around the 0.5 prevalence value

R depends on prevalence



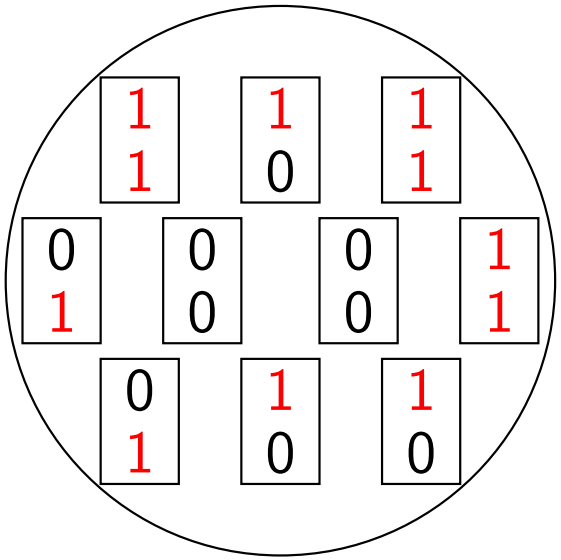
50 clusters

$\widehat{\pi}_1 = \mathbf{0.35}$



50 clusters

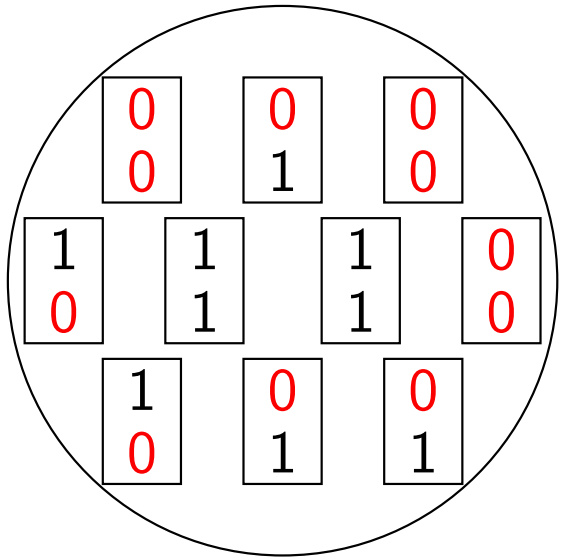
$\widehat{\pi}_2 = \mathbf{0.65}$



50 clusters

$\widehat{\pi}_1 = 0.35$
$\widehat{\rho}_1 = 0.78$
$\widehat{R}_1 = 2.45$

$\widehat{\rho}_1 = \widehat{\rho}_1$
$\widehat{R}_1 \neq \widehat{R}_2$



50 clusters

$\widehat{\pi}_2 = 0.65$
$\widehat{\rho}_2 = 0.78$
$\widehat{R}_2 = 1.42$

Plan

Plan

Data generation

- Generate a continuous outcome with a fixed ICC

Data generation

- Generate a continuous outcome with a fixed ICC
- Dichotomize the continuous outcome to obtain a binary outcome, for a given threshold

Data generation

- Generate a continuous outcome with a fixed ICC
- Dichotomize the continuous outcome to obtain a binary outcome, for a given threshold
- Vary the threshold of dichotomization $\rightarrow$ vary prevalence with a fixed underlying clustering

Data generation

- Generate a continuous outcome with a fixed ICC
- Dichotomize the continuous outcome to obtain a binary outcome, for a given threshold
- Vary the threshold of dichotomization $\rightarrow$ vary prevalence with a fixed underlying clustering
- Generate 10 000 datasets

Data generation

- Generate a continuous outcome with a fixed ICC
- Dichotomize the continuous outcome to obtain a binary outcome, for a given threshold
- Vary the threshold of dichotomization $\rightarrow$ vary prevalence with a fixed underlying clustering
- Generate 10 000 datasets
- Many scenarios : $\begin{cases} \text{Fixed/variable cluster sizes} \\ \text{Various underlying ICCs} \\ \text{Various numbers of cluster} \end{cases}$

Plan

ICC = 0.3

20 clusters: mean = 25, variance = 225

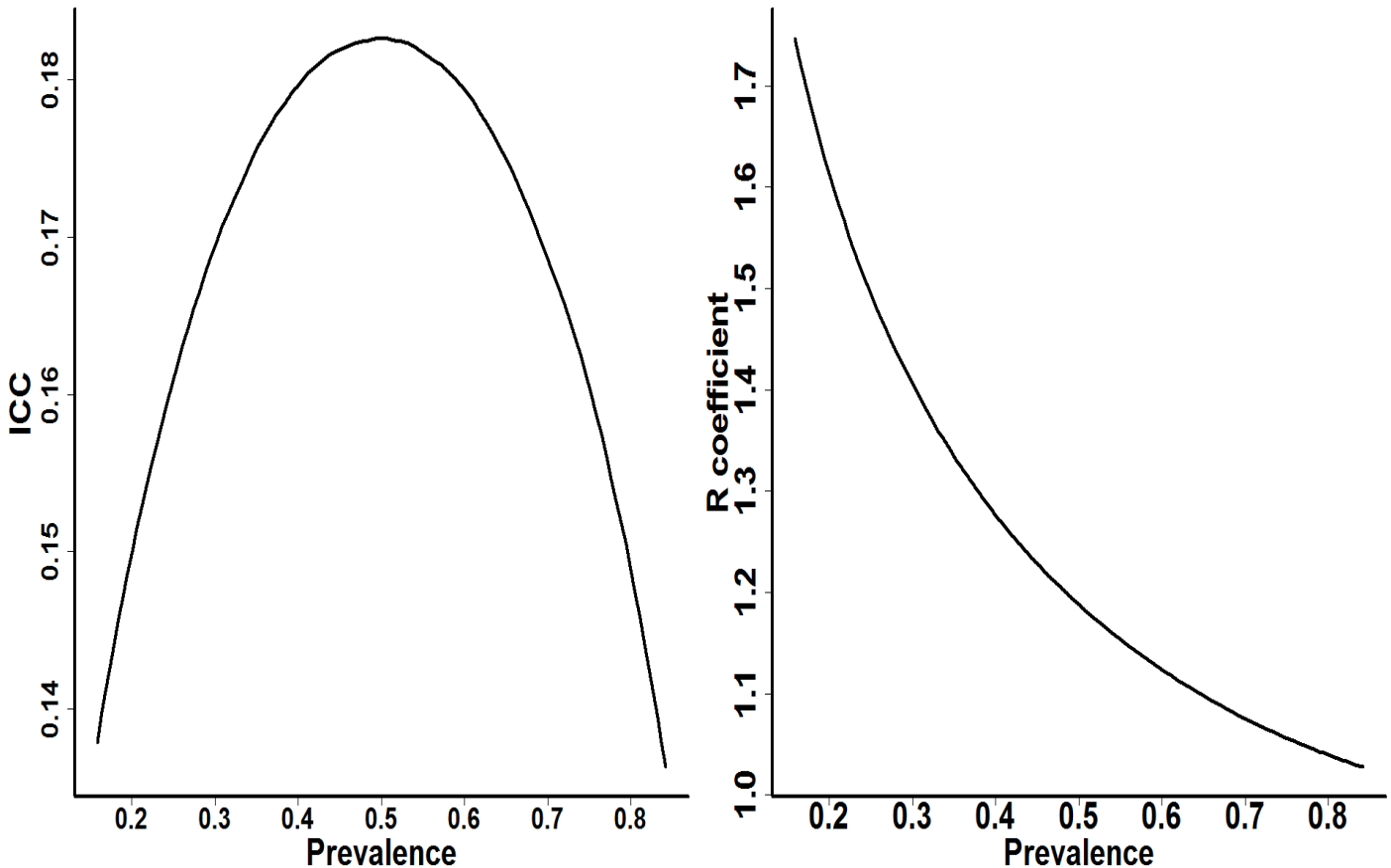


Figure: Mean ICC and R coefficient estimates for binary outcomes as a function of estimated outcome prevalence.

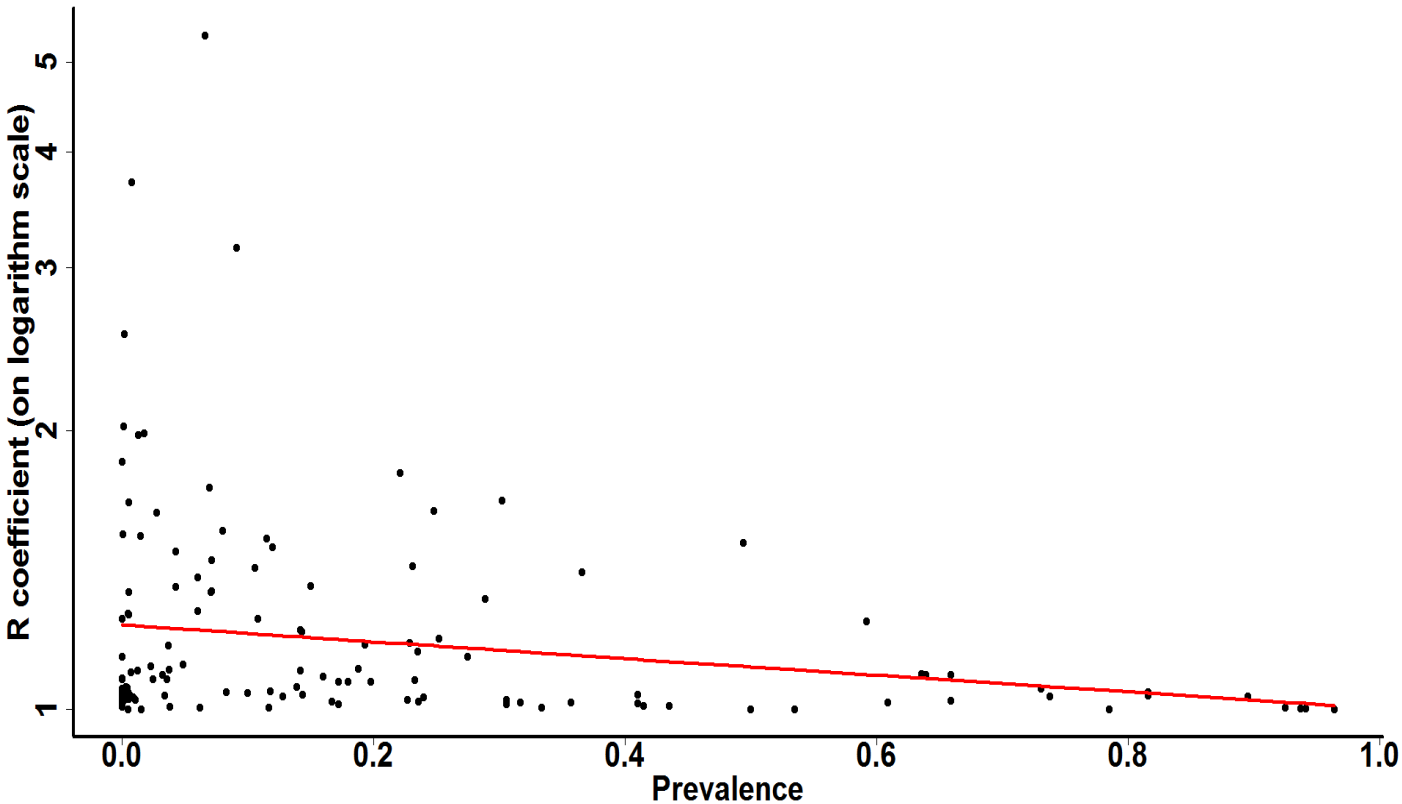
Plan

Data

- Data obtained from a Health Technology Assessment (HTA) review[7]
- 139 ICCs and prevalences for analysis

[7] Ukoumunne O. et al., *Health Technology Assessment*, 3(5):iii–92, 1999

Results



Association between R and prevalence with data from the Health Technology Assessment review.
The estimated slope coefficient was -0.208 and $p = 0.018$.

Plan

- Clusters of size 2: proof that R depends on prevalence
- Clusters of size greater than 2: Simulations $\rightarrow$ R depends on prevalence

- Clusters of size 2: proof that R depends on prevalence
- Clusters of size greater than 2: Simulations $\rightarrow$ R depends on prevalence
- As for the ICC, R coefficient depends on the outcome prevalence
- R coefficient is not the appropriate coefficient
- Future works needed

THANK YOU FOR YOUR ATTENTION