

# Covariate-adjusted generalized pairwise comparisons in small samples

S. Jaspers, J. Verbeek, O. Thas

*Data Science Institute - Hasselt University*

September 26, 2024



[WWW.UHASSELT.BE/DSI](http://WWW.UHASSELT.BE/DSI)

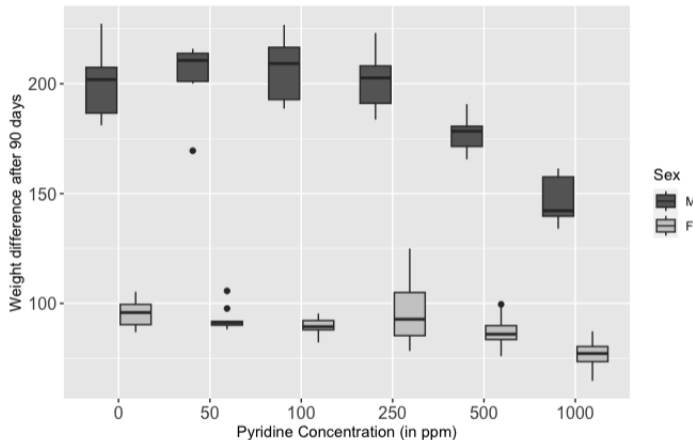


- 1 Introduction
- 2 Methodology
- 3 Simulation Study
- 4 Data Application
- 5 Discussion

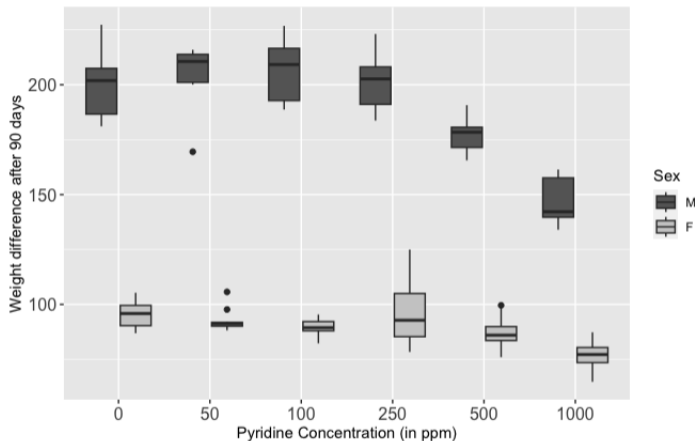
---

# Introduction

# Motivating example: Pyridine data

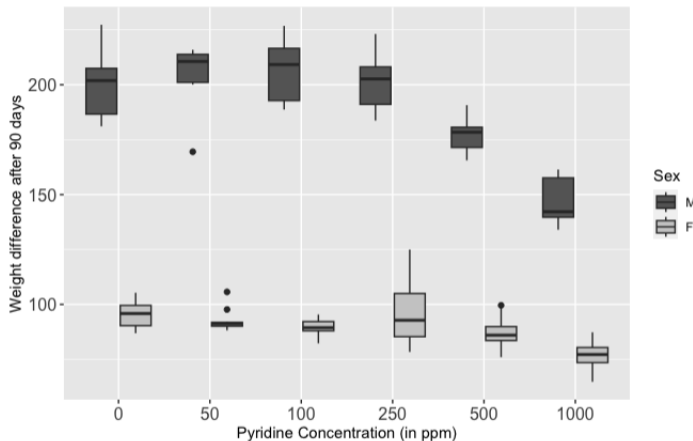


# Motivating example: Pyridine data



- ▶ N = 120 rats (60 male and 60 female) were randomized to six different dose levels

# Motivating example: Pyridine data



- ▶  $N = 120$  rats (60 male and 60 female) were randomized to six different dose levels
- ▶  $n_{ij} = 10$  animals per dose/sex

- ▶ GPC effect sizes: related to the probability that the outcome of a subject in one group ( $A = 1$ ) is favourable compared to the outcome of a subject from another group ( $A = 0$ ).
- ▶ E.g.:  $P(Y < Y^* | A = 0, A^* = 1)$ .
- ▶ When ties are possible, a popular adaptation is the effect size

$$\Theta = P(Y < Y^* | A = 0, A^* = 1) + 0.5P(Y = Y^* | A = 0, A^* = 1) := P(Y \preceq Y^* | A = 0, A^* = 1),$$

which will be referred to as the probabilistic index (PI).

- ▶ In presence of covariates: conditional PI

$$P(Y \preceq Y^* | A = 0, A^* = 1, \mathbf{X} = \mathbf{X}^*)$$

---

## Methodology



- ▶ PI can be estimated using Probabilistic Index Model (PIM) as introduced by Thas et al. (2012):

$$g(P(Y \preceq Y^* | A, A^*, \mathbf{X}, \mathbf{X}^*)) = \beta_0 + \beta_A(A^* - A) + \beta'_X(\mathbf{X}^* - \mathbf{X}),$$

where  $g(\cdot)$  is a link function.

- ▶ PIM expressed in terms of pseudo-observations  $I_{ij}$ :

$$E \{I(Y_j \preceq Y_j^*) | A, A^*, \mathbf{X}, \mathbf{X}^*\} = P(Y_j \preceq Y_j^* | A, A^*, \mathbf{X}, \mathbf{X}^*) = g^{-1}(\mathbf{Z}'\boldsymbol{\beta}),$$

with  $\mathbf{Z}'_{ij} = (A_j^* - A_i, (\mathbf{X}^*_{j} - \mathbf{X}_i)')$  and  $I_{ij} = I(Y_i \preceq Y_j^*) = I(Y_i < Y_j^*) + 0.5I(Y_i = Y_j^*)$

- ▶ Parameter estimates based on solving estimating equation

$$U_n(\boldsymbol{\beta}) = \sum_{(i,j) \in \mathcal{I}_n} \tilde{\mathbf{A}}(\mathbf{Z}_{ij}; \boldsymbol{\beta}) \{I_{ij} - g^{-1}(\mathbf{Z}'_{ij}\boldsymbol{\beta})\} = \mathbf{0},$$

$$\tilde{\mathbf{A}}(\mathbf{Z}_{ij}; \boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}} g^{-1}(\mathbf{Z}'_{ij}\boldsymbol{\beta}) V^{-1}(g^{-1}(\mathbf{Z}'_{ij}\boldsymbol{\beta})),$$

where  $V^{-1}(g^{-1}(\mathbf{Z}'_{ij}\boldsymbol{\beta})) = g^{-1}(\mathbf{Z}'_{ij}\boldsymbol{\beta})(1 - g^{-1}(\mathbf{Z}'_{ij}\boldsymbol{\beta}))$  is equal to the variance of the pseudo-observations  $I_{ij}$

- ▶ Parameter estimates based on solving estimating equation

$$U_n(\boldsymbol{\beta}) = \sum_{(i,j) \in \mathcal{I}_n} \tilde{\mathbf{A}}(\mathbf{Z}_{ij}; \boldsymbol{\beta}) \{I_{ij} - g^{-1}(\mathbf{Z}'_{ij}\boldsymbol{\beta})\} = \mathbf{0},$$

$$\tilde{\mathbf{A}}(\mathbf{Z}_{ij}; \boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}} g^{-1}(\mathbf{Z}'_{ij}\boldsymbol{\beta}) V^{-1}(g^{-1}(\mathbf{Z}'_{ij}\boldsymbol{\beta})),$$

where  $V^{-1}(g^{-1}(\mathbf{Z}'_{ij}\boldsymbol{\beta})) = g^{-1}(\mathbf{Z}'_{ij}\boldsymbol{\beta})(1 - g^{-1}(\mathbf{Z}'_{ij}\boldsymbol{\beta}))$  is equal to the variance of the pseudo-observations  $I_{ij}$

- ▶ Problem: Asymptotic estimation theory is not applicable if the sample size is small ( $< 50$ )

- ▶ Parameter estimates based on solving estimating equation

$$U_n(\boldsymbol{\beta}) = \sum_{(i,j) \in \mathcal{I}_n} \tilde{\mathbf{A}}(\mathbf{Z}_{ij}; \boldsymbol{\beta}) \{I_{ij} - g^{-1}(\mathbf{Z}'_{ij}\boldsymbol{\beta})\} = \mathbf{0},$$

$$\tilde{\mathbf{A}}(\mathbf{Z}_{ij}; \boldsymbol{\beta}) = \frac{\partial}{\partial \boldsymbol{\beta}} g^{-1}(\mathbf{Z}'_{ij}\boldsymbol{\beta}) V^{-1}(g^{-1}(\mathbf{Z}'_{ij}\boldsymbol{\beta})),$$

where  $V^{-1}(g^{-1}(\mathbf{Z}'_{ij}\boldsymbol{\beta})) = g^{-1}(\mathbf{Z}'_{ij}\boldsymbol{\beta})(1 - g^{-1}(\mathbf{Z}'_{ij}\boldsymbol{\beta}))$  is equal to the variance of the pseudo-observations  $I_{ij}$

- ▶ Problem: Asymptotic estimation theory is not applicable if the sample size is small ( $< 50$ )
- ▶ Huge improvement by Amorim et al. (2015), who introduced a bias-reduced adjusted jackknife empirical likelihood procedure

# PIM within GEE framework

Pseudo-observations	Cluster1	Cluster2	Cluster3	Pseudo-covariates
$Y_1^{GEE} = I_{11} = I_{y_{11} < y_{21}}$	1	1	1	$X_1^{GEE} = z_{11}$
$Y_2^{GEE} = I_{12} = I_{y_{11} < y_{22}}$	1	2	2	$X_2^{GEE} = z_{12}$
$Y_3^{GEE} = I_{13} = I_{y_{11} < y_{23}}$	1	3	3	$X_3^{GEE} = z_{13}$
$Y_4^{GEE} = I_{21} = I_{y_{12} < y_{21}}$	2	1	4	$X_4^{GEE} = z_{21}$
$Y_5^{GEE} = I_{22} = I_{y_{12} < y_{22}}$	2	2	5	$X_5^{GEE} = z_{22}$
$Y_6^{GEE} = I_{23} = I_{y_{12} < y_{23}}$	2	3	6	$X_6^{GEE} = z_{23}$
$Y_7^{GEE} = I_{31} = I_{y_{13} < y_{21}}$	3	1	7	$X_7^{GEE} = z_{31}$
$Y_8^{GEE} = I_{32} = I_{y_{13} < y_{22}}$	3	2	8	$X_8^{GEE} = z_{32}$
$Y_9^{GEE} = I_{33} = I_{y_{13} < y_{23}}$	3	3	9	$X_9^{GEE} = z_{33}$

Pseudo-observations	Cluster1	Cluster2	Cluster3	Pseudo-covariates
$Y_1^{GEE} = I_{11} = I_{y_{11} < y_{21}}$	1	1	1	$X_1^{GEE} = z_{11}$
$Y_2^{GEE} = I_{12} = I_{y_{11} < y_{22}}$	1	2	2	$X_2^{GEE} = z_{12}$
$Y_3^{GEE} = I_{13} = I_{y_{11} < y_{23}}$	1	3	3	$X_3^{GEE} = z_{13}$
$Y_4^{GEE} = I_{21} = I_{y_{12} < y_{21}}$	2	1	4	$X_4^{GEE} = z_{21}$
$Y_5^{GEE} = I_{22} = I_{y_{12} < y_{22}}$	2	2	5	$X_5^{GEE} = z_{22}$
$Y_6^{GEE} = I_{23} = I_{y_{12} < y_{23}}$	2	3	6	$X_6^{GEE} = z_{23}$
$Y_7^{GEE} = I_{31} = I_{y_{13} < y_{21}}$	3	1	7	$X_7^{GEE} = z_{31}$
$Y_8^{GEE} = I_{32} = I_{y_{13} < y_{22}}$	3	2	8	$X_8^{GEE} = z_{32}$
$Y_9^{GEE} = I_{33} = I_{y_{13} < y_{23}}$	3	3	9	$X_9^{GEE} = z_{33}$

- ▶ Two non-nested levels of clustering

Pseudo-observations	Cluster1	Cluster2	Cluster3	Pseudo-covariates
$Y_1^{GEE} = I_{11} = I_{y_{11} < y_{21}}$	1	1	1	$X_1^{GEE} = z_{11}$
$Y_2^{GEE} = I_{12} = I_{y_{11} < y_{22}}$	1	2	2	$X_2^{GEE} = z_{12}$
$Y_3^{GEE} = I_{13} = I_{y_{11} < y_{23}}$	1	3	3	$X_3^{GEE} = z_{13}$
$Y_4^{GEE} = I_{21} = I_{y_{12} < y_{21}}$	2	1	4	$X_4^{GEE} = z_{21}$
$Y_5^{GEE} = I_{22} = I_{y_{12} < y_{22}}$	2	2	5	$X_5^{GEE} = z_{22}$
$Y_6^{GEE} = I_{23} = I_{y_{12} < y_{23}}$	2	3	6	$X_6^{GEE} = z_{23}$
$Y_7^{GEE} = I_{31} = I_{y_{13} < y_{21}}$	3	1	7	$X_7^{GEE} = z_{31}$
$Y_8^{GEE} = I_{32} = I_{y_{13} < y_{22}}$	3	2	8	$X_8^{GEE} = z_{32}$
$Y_9^{GEE} = I_{33} = I_{y_{13} < y_{23}}$	3	3	9	$X_9^{GEE} = z_{33}$

- ▶ Two non-nested levels of clustering
- ▶ Correct inference for GEEs with non-nested clusters provided by Miglioretti and Heagerty (2004)

Pseudo-observations	Cluster1	Cluster2	Cluster3	Pseudo-covariates
$Y_1^{GEE} = I_{11} = I_{y_{11} < y_{21}}$	1	1	1	$X_1^{GEE} = z_{11}$
$Y_2^{GEE} = I_{12} = I_{y_{11} < y_{22}}$	1	2	2	$X_2^{GEE} = z_{12}$
$Y_3^{GEE} = I_{13} = I_{y_{11} < y_{23}}$	1	3	3	$X_3^{GEE} = z_{13}$
$Y_4^{GEE} = I_{21} = I_{y_{12} < y_{21}}$	2	1	4	$X_4^{GEE} = z_{21}$
$Y_5^{GEE} = I_{22} = I_{y_{12} < y_{22}}$	2	2	5	$X_5^{GEE} = z_{22}$
$Y_6^{GEE} = I_{23} = I_{y_{12} < y_{23}}$	2	3	6	$X_6^{GEE} = z_{23}$
$Y_7^{GEE} = I_{31} = I_{y_{13} < y_{21}}$	3	1	7	$X_7^{GEE} = z_{31}$
$Y_8^{GEE} = I_{32} = I_{y_{13} < y_{22}}$	3	2	8	$X_8^{GEE} = z_{32}$
$Y_9^{GEE} = I_{33} = I_{y_{13} < y_{23}}$	3	3	9	$X_9^{GEE} = z_{33}$

- ▶ Two non-nested levels of clustering
- ▶ Correct inference for GEEs with non-nested clusters provided by Miglioretti and Heagerty (2004)
- ▶ Several small-sample corrections available as well (e.g. MacKinnon, 1985; Fay and Graubard, 2001; Morel et al., 2003, ...)



- ▶ Following Miglioretti and Heagerty (2004), fit 3 working independence GEE models, clustering on Cluster1, Cluster2 and Cluster3

$$g(\mu_{kj}) = \mathbf{X}_{kj} \boldsymbol{\alpha}.$$

$$U(\boldsymbol{\alpha}) = \sum_{k=1}^K \mathbf{D}'_k \mathbf{V}_k^{-1} (\mathbf{Y}_k^{GEE} - \boldsymbol{\mu}_k) = \mathbf{0},$$

$$\mathbf{V}_{LZ}^{GEE} = \left( \sum_{k=1}^K \mathbf{D}'_k \mathbf{V}_k^{-1} \mathbf{D} \right)^{-1} \mathbf{M}_{LZ} \left( \sum_{k=1}^K \mathbf{D}'_k \mathbf{V}_k^{-1} \mathbf{D} \right)^{-1},$$

$$\mathbf{M}_{LZ} = \sum_{k=1}^K \mathbf{D}'_k \mathbf{V}_k^{-1} \text{Cov}(\mathbf{Y}_k^{GEE})_{LZ} \mathbf{V}_k^{-1} \mathbf{D}.$$

## Proposed method (Steps 2 and 3)

---

- ▶ Apply small-sample correction to the three variance-covariance matrices  $V_1^{GEE}$ ,  $V_2^{GEE}$ ,  $V_3^{GEE}$

## Proposed method (Steps 2 and 3)

- ▶ Apply small-sample correction to the three variance-covariance matrices  $V_1^{GEE}$ ,  $V_2^{GEE}$ ,  $V_3^{GEE}$ 
  - ▶ 8 adjustments considered, most promising results from Morel et al. (2003)
    - $M_{MBN} = \frac{n^*-1}{n^*-p} \frac{K-1}{K-p} M_{LZ}$ , where  $n^* = \sum_{k=1}^K n_k$
    - $V_{MBN}^{GEE} = \mathbf{B}_{LZ} M_{MBN} \mathbf{B}_{LZ} + \hat{\delta} \hat{\chi} \mathbf{B}_{LZ}$ , where  $\hat{\delta} = \min(0.5, \frac{p}{K-p})$  and  $\hat{\chi} = \max(1; \frac{\text{trace}\{\mathbf{B}_{LZ} M_{LZ}\}}{p})$ .

- ▶ Apply small-sample correction to the three variance-covariance matrices  $V_1^{GEE}$ ,  $V_2^{GEE}$ ,  $V_3^{GEE}$ 
  - ▶ 8 adjustments considered, most promising results from Morel et al. (2003)
    - $M_{MBN} = \frac{n^*-1}{n^*-p} \frac{K-1}{K-p} M_{LZ}$ , where  $n^* = \sum_{k=1}^K n_k$
    - $V_{MBN}^{GEE} = B_{LZ} M_{MBN} B_{LZ} + \hat{\delta} \hat{\chi} B_{LZ}$ , where  $\hat{\delta} = \min(0.5, \frac{p}{K-p})$  and  $\hat{\chi} = \max(1; \frac{\text{trace}\{B_{LZ} M_{LZ}\}}{p})$ .
- ▶ Combine into final variance-covariance matrix through

$$V^{GEE} = V_{MBN,1}^{GEE} + V_{MBN,2}^{GEE} - V_{MBN,3}^{GEE}$$

## Proposed method (Steps 2 and 3)

- ▶ Apply small-sample correction to the three variance-covariance matrices  $V_1^{GEE}$ ,  $V_2^{GEE}$ ,  $V_3^{GEE}$ 
  - ▶ 8 adjustments considered, most promising results from Morel et al. (2003)
    - $M_{MBN} = \frac{n^*-1}{n^*-p} \frac{K-1}{K-p} M_{LZ}$ , where  $n^* = \sum_{k=1}^K n_k$
    - $V_{MBN}^{GEE} = B_{LZ} M_{MBN} B_{LZ} + \hat{\delta} \hat{\chi} B_{LZ}$ , where  $\hat{\delta} = \min(0.5, \frac{p}{K-p})$  and  $\hat{\chi} = \max(1; \frac{\text{trace}\{B_{LZ} M_{LZ}\}}{p})$ .
- ▶ Combine into final variance-covariance matrix through

$$V^{GEE} = V_{MBN,1}^{GEE} + V_{MBN,2}^{GEE} - V_{MBN,3}^{GEE}$$

Extra: separation issues can be dealt with when logit link is used, currently only with MBN adjustment

---

## Simulation Study

▶ Model:

$$Y_i = \alpha_1 X_{1i} + \alpha_2 X_{2i} + \cdots + \alpha_p X_{pi} + \epsilon_i, \quad i = 1, \dots, n,$$

$\epsilon_i | X_i$  are IID  $\mathcal{N}(0, 1)$

$X_1$  is a balanced group indicator (e.g. treatment)

▶ Model:

$$Y_i = \alpha_1 X_{1i} + \alpha_2 X_{2i} + \cdots + \alpha_p X_{pi} + \epsilon_i, \quad i = 1, \dots, n,$$

$\epsilon_i | X_i$  are IID  $\mathcal{N}(0, 1)$

$X_1$  is a balanced group indicator (e.g. treatment)

▶ Corresponding PIMs:

$$\Phi^{-1}(\mathbb{P}\{Y \preceq Y^* | X_1, X_1^*, X_2, X_2^*, \dots, X_p, X_p^*\}) = \beta_1(X_1^* - X_1) + \beta_2(X_2^* - X_2) + \cdots + \beta_p(X_p^* - X_p),$$



▶ Model:

$$Y_i = \alpha_1 X_{1i} + \alpha_2 X_{2i} + \cdots + \alpha_p X_{pi} + \epsilon_i, \quad i = 1, \dots, n,$$

$\epsilon_i | X_i$  are IID  $\mathcal{N}(0, 1)$

$X_1$  is a balanced group indicator (e.g. treatment)

▶ Corresponding PIMs:

$$\Phi^{-1}(\mathbb{P}\{Y \preceq Y^* | X_1, X_1^*, X_2, X_2^*, \dots, X_p, X_p^*\}) = \beta_1(X_1^* - X_1) + \beta_2(X_2^* - X_2) + \cdots + \beta_p(X_p^* - X_p),$$

$$\Phi^{-1}(\mathbb{P}\{Y \preceq Y^* | X_1 < X_1^*, X_2, X_2^*, \dots, X_p, X_p^*\}) = \delta_1 + \delta_2(X_2^* - X_2) + \cdots + \delta_p(X_p^* - X_p)$$

▶ Model:

$$Y_i = \alpha_1 X_{1i} + \alpha_2 X_{2i} + \cdots + \alpha_p X_{pi} + \epsilon_i, \quad i = 1, \dots, n,$$

$\epsilon_i | X_i$  are IID  $\mathcal{N}(0, 1)$

$X_1$  is a balanced group indicator (e.g. treatment)

▶ Corresponding PIMs:

$$\Phi^{-1}(\mathbb{P}\{Y \preceq Y^* | X_1, X_1^*, X_2, X_2^*, \dots, X_p, X_p^*\}) = \beta_1(X_1^* - X_1) + \beta_2(X_2^* - X_2) + \cdots + \beta_p(X_p^* - X_p),$$

$$\Phi^{-1}(\mathbb{P}\{Y \preceq Y^* | X_1 < X_1^*, X_2, X_2^*, \dots, X_p, X_p^*\}) = \delta_1 + \delta_2(X_2^* - X_2) + \cdots + \delta_p(X_p^* - X_p)$$

▶  $n = 14, 16, 20, 24, 30$  observations

▶ 1000 simulation runs with

▶  $\alpha_1 = 0, 0.5, 1$  ( $\beta_1 = \delta_1 = \alpha_1/\sqrt{2}$ )

▶  $p = 2, 4, 6$

▶ Model:

$$Y_i = \alpha_1 X_{1i} + \alpha_2 X_{2i} + \cdots + \alpha_p X_{pi} + \epsilon_i, \quad i = 1, \dots, n,$$

$\epsilon_i | X_i$  are IID  $\mathcal{N}(0, 1)$

$X_1$  is a balanced group indicator (e.g. treatment)

▶ Corresponding PIMs:

$$\Phi^{-1}(\mathbb{P}\{Y \preceq Y^* | X_1, X_1^*, X_2, X_2^*, \dots, X_p, X_p^*\}) = \beta_1(X_1^* - X_1) + \beta_2(X_2^* - X_2) + \cdots + \beta_p(X_p^* - X_p),$$

$$\Phi^{-1}(\mathbb{P}\{Y \preceq Y^* | X_1 < X_1^*, X_2, X_2^*, \dots, X_p, X_p^*\}) = \delta_1 + \delta_2(X_2^* - X_2) + \cdots + \delta_p(X_p^* - X_p)$$

▶  $n = 14, 16, 20, 24, 30$  observations

▶ 1000 simulation runs with

▶  $\alpha_1 = 0, 0.5, 1$  ( $\beta_1 = \delta_1 = \alpha_1/\sqrt{2}$ )

▶  $p = 2, 4, 6$

▶ Coverage of 95% confidence intervals of treatment parameter

# Coverage of the 95% confidence intervals ( $p = 2$ )

$p$	$\alpha_1$	N	PIM	BR-AJEL	AJEL	MBN	Pan	WL	GST	KC	MD	MK	FG
2	0	14	87.50	95.07	94.83	95.43	89.18	96.03	94.11	96.03	93.99	93.63	91.11
		16	90.47	96.10	95.87	96.44	92.31	97.01	95.64	97.24	95.75	94.95	93.34
		20	92.84	97.12	96.47	96.15	92.63	97.01	96.05	97.33	96.26	95.83	94.44
		24	92.23	97.03	95.60	95.19	92.74	96.01	95.30	96.22	95.19	94.48	93.25
		30	92.65	96.37	94.86	95.17	92.15	95.27	94.36	96.58	95.17	94.76	94.06
	0.5	14	86.83	93.73	94.10	96.24	89.34	96.61	95.48	95.73	94.73	94.35	92.60
		16	90.14	96.88	95.38	96.38	90.26	96.63	96.00	96.88	95.38	94.76	94.01
		20	91.39	95.81	96.15	96.38	92.41	96.49	94.90	97.06	96.04	94.90	92.98
		24	92.61	96.79	95.82	95.50	93.04	96.04	95.29	97.00	95.50	95.18	94.11
		30	92.04	96.38	95.45	95.24	92.24	95.86	94.83	97.21	95.24	94.62	93.59
	1	14	83.64	91.44	93.43	95.72	86.54	96.79	94.50	97.09	93.27	92.66	93.12
		16	87.01	94.26	94.71	96.53	90.33	96.68	94.11	95.77	93.35	93.05	91.84
		20	89.90	95.02	95.44	94.74	91.84	95.30	94.05	95.85	93.78	93.36	92.39
		24	92.32	95.91	95.04	94.80	92.81	95.79	94.92	97.15	94.42	94.05	93.06
		30	91.41	95.19	94.73	94.39	92.10	94.62	93.36	96.22	94.62	93.93	92.44

# Coverage of the 95% confidence intervals ( $p = 4$ )

p	$\alpha_1$	N	PIM	BR-AJEL	AJEL	MBN	Pan	WL	GST	KC	MD	MK	FG
4	0	14	78.90	81.60	90.12	94.59	78.59	97.92	97.82	95.11	95.84	96.15	88.15
		16	81.99	88.00	90.84	95.02	82.10	97.05	97.25	95.12	96.64	96.24	88.30
		20	88.32	93.71	93.60	96.24	86.60	96.65	95.43	95.53	96.24	96.04	90.76
		24	88.33	94.06	92.76	95.57	88.43	95.47	95.27	94.97	95.37	94.87	90.14
		30	90.47	94.38	93.48	94.88	89.97	94.58	94.18	95.59	95.39	94.98	91.88
	0.5	14	78.75	76.64	89.53	95.14	78.44	96.83	97.89	94.50	95.14	95.67	88.69
		16	82.77	85.54	90.77	94.87	82.87	97.64	97.23	94.36	95.69	95.69	89.03
		20	87.64	91.59	93.52	95.74	86.12	96.76	96.15	94.63	95.74	95.44	90.68
		24	86.75	92.57	91.97	95.48	87.25	94.98	93.88	93.78	94.98	94.38	88.65
		30	89.44	93.56	93.56	93.96	89.44	94.16	93.76	93.76	94.47	93.76	91.05
	1	14	76.19	72.54	88.15	96.01	78.63	96.12	97.67	95.24	95.02	93.91	90.48
		16	80.23	80.34	89.27	96.49	82.04	97.24	97.02	94.90	94.26	93.84	90.12
		20	83.81	86.70	92.16	94.85	83.40	96.39	95.05	92.89	92.68	92.16	87.22
		24	84.99	91.37	91.78	94.66	85.41	94.96	93.94	91.47	93.11	92.91	87.26
		30	89.04	93.55	93.24	94.16	88.63	94.36	94.16	92.01	93.55	93.34	89.75

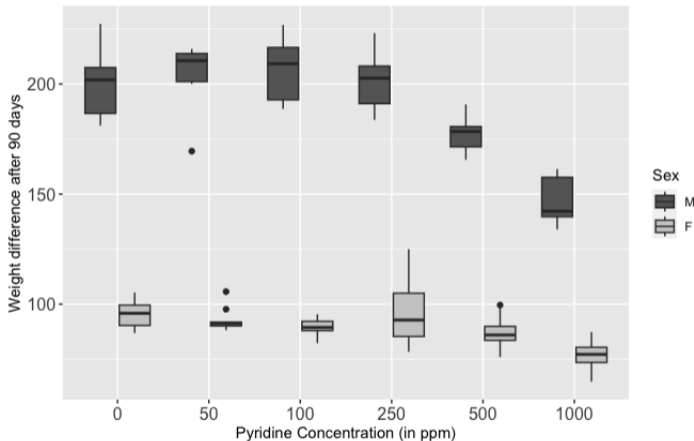
# Coverage of the 95% confidence intervals ( $p = 6$ )

$p$	$\alpha_1$	N	PIM	BR-AJEL	AJEL	MBN	Pan	WL	GST	KC	MD	MK	FG
6	0	14	70.00	55.36	79.05	94.29	74.05	88.57	99.52	97.14	98.57	99.52	92.74
		16	77.79	66.63	86.80	94.74	72.53	97.00	99.46	96.35	98.50	99.14	90.02
		20	79.43	78.31	87.68	90.84	75.56	96.33	96.64	93.08	95.82	96.44	84.32
		24	83.15	88.09	90.01	93.54	80.32	95.56	95.36	94.15	95.86	95.86	85.97
		30	88.40	93.10	92.70	95.20	85.10	95.40	95.30	95.40	96.30	95.80	89.50
	0.5	14	68.56	48.61	81.44	94.32	73.78	86.89	99.19	95.48	98.49	99.77	92.34
		16	75.57	62.81	85.95	94.05	71.78	95.89	99.57	96.00	98.16	99.46	88.32
		20	77.99	76.66	87.41	92.02	74.72	96.72	96.62	92.22	96.01	96.21	85.06
		24	82.05	85.86	89.37	92.78	80.64	95.09	94.68	92.98	93.88	94.68	85.16
		30	86.40	92.00	91.80	94.90	84.20	94.60	94.40	93.10	95.40	95.60	88.40
	1	14	67.80	43.66	79.88	96.22	74.88	84.51	99.02	96.59	98.29	99.76	93.05
		16	76.22	58.67	85.67	95.89	73.89	93.22	98.89	94.33	97.44	98.44	91.11
		20	75.54	73.59	86.33	92.09	73.38	96.61	97.23	92.39	95.07	95.79	86.43
		24	82.56	84.27	88.81	91.83	79.84	95.26	95.67	89.11	93.65	94.35	84.88
		30	85.54	90.56	91.67	94.28	83.43	94.88	94.08	89.56	94.28	94.88	86.95

---

## Data Application

# Pyridine data





# Results for male rats

		Male rats				
		50 ppm	100 ppm	250 ppm	500 ppm	1000 ppm
0 ppm	PI (CI)	71.7 (36.7;91.7)	73.93 (27.7;95.5)	63.32 (26.2;89.4)	7.9 (1.5;32.6)	<0.01 (<0.01;<0.01)
	Unadj p	0.2142	0.3039	0.4946	0.0058	< 0.0001
	Adj p	0.4015	0.5362	0.7809	0.0175	<0.0001
50 ppm		-	58.82 (25.9;85.4)	44.64 (17.9;74.8)	8.62 (1.2;41.7)	<0.01 (<0.01;<0.01)
			0.6169	0.744	0.0229	<0.0001
			0.8591	0.8591	0.0572	<0.0001
100 ppm		-	-	38.83 (12.8;73.3)	1.01 (0.2;5.7)	<0.01 (<0.01;<0.01)
				0.5394	<0.0001	<0.0001
				0.8091	<0.0001	<0.0001
250 ppm		-	-	-	5.48 (0.79;29.7)	<0.01 (<0.01;<0.01)
					0.0054	<0.0001
					0.0175	<0.0001
500 ppm		-	-	-	-	<0.01 (<0.01;<0.01)
						<0.0001
						<0.0001

# Results for female rats

		Female rats				
		50 ppm	100 ppm	250 ppm	500 ppm	1000 ppm
0 ppm	PI(CI)	42.78 (14.9;76.2)	60.83 (4.6;98.04)	45.23 (15.6;78.7)	9.18 (0.7;59.5)	1.04 (0.1;7.4)
	Unadj p	0.6921	0.8019	0.8005	0.0924	<0.0001
	Adj p	0.8591	0.8591	0.8591	0.2131	0.0001
50 ppm	-	-	57.61 (18.8;88.9)	49.92 (19;80.9)	19.25 (3.5;61)	<0.01 (<0.01;<0.01)
	-	-	0.7311	0.9964	0.1336	<0.0001
	-	-	0.8591	0.9964	0.2672	<0.0001
100 ppm	-	-	-	61.16 (19.0;91.4)	50.76 (8.5;92)	3.56 (0.06;69.6)
	-	-	-	0.637	0.9802	0.1156
	-	-	-	0.8591	0.9964	0.2478
250 ppm	-	-	-	-	56.68 (14.2;91.2)	11.75 (2.5;40.6)
	-	-	-	-	0.7971	0.0163
	-	-	-	-	0.8591	0.0445
500 ppm	-	-	-	-	-	24.49 (1.6;86.5)
	-	-	-	-	-	0.4546
	-	-	-	-	-	0.7577

```
require(devtools)
install_github("JaspersStijn/SmallSamplePIMFinal", force=TRUE);

library("SmallSamplePIM")

fit = GEE_MH_fit(data=subset_to_be_fitted,
  response="weight_use",
  treatment = "Dose",
  control = c("CreatT5", "TPTS", "ALBT5"),
  correction = "Firth",
  link="logit")

> expit(fit$beta_est);fit$pval
[1] 0.7169745
[1] 0.2141575
```

	Dose	weight_use	CreatT5	TPTS	ALBT5
1	0	207.9	0.5	6.5	3.3
2	0	203.3	0.5	6.4	3.5
3	0	191.2	0.5	6.3	3.3
4	0	185.1	0.4	6.6	3.7
5	0	206.2	0.5	6.2	3.6
6	0	211.0	0.5	5.9	3.6
7	0	181.4	0.5	6.4	3.1
8	0	200.5	0.5	6.3	3.6
9	0	181.1	0.5	6.3	3.7
10	0	227.3	0.5	6.2	3.3
11	50	169.5	0.5	6.3	3.5
12	50	210.8	0.5	6.0	3.2
13	50	202.2	0.6	6.7	3.8
14	50	213.4	0.5	6.2	3.5
15	50	215.9	0.4	6.3	3.7
16	50	214.0	0.5	6.7	3.5
17	50	200.1	0.6	6.4	3.9
18	50	210.4	0.5	6.4	3.4
19	50	215.5	0.6	6.7	3.7
20	50	200.7	0.4	6.4	3.9

<https://github.com/JaspersStijn/SmallSamplePIMFinal/>

---

## Discussion

- ▶ PIM reformulated into GEE framework for better small sample performance
  - ▶ Especially MBN adjustment + Firth correction
  - ▶ Sample size as low as 7 per group
  - ▶ Only for comparing two groups

- ▶ PIM reformulated into GEE framework for better small sample performance
  - ▶ Especially MBN adjustment + Firth correction
  - ▶ Sample size as low as 7 per group
  - ▶ Only for comparing two groups
- ▶ GPC based on PI directly translates to other GPC effect sizes:
  - ▶ e.g.  $NTB = 2PI - 1$

- ▶ PIM reformulated into GEE framework for better small sample performance
  - ▶ Especially MBN adjustment + Firth correction
  - ▶ Sample size as low as 7 per group
  - ▶ Only for comparing two groups
- ▶ GPC based on PI directly translates to other GPC effect sizes:
  - ▶ e.g.  $NTB = 2PI - 1$
- ▶ Inference for other effect sizes based on delta method

- ▶ PIM reformulated into GEE framework for better small sample performance
  - ▶ Especially MBN adjustment + Firth correction
  - ▶ Sample size as low as 7 per group
  - ▶ Only for comparing two groups
- ▶ GPC based on PI directly translates to other GPC effect sizes:
  - ▶ e.g.  $NTB = 2PI - 1$
- ▶ Inference for other effect sizes based on delta method

# Any Questions?

stijn.jaspers@uhasselt.be



- Amorim, G., Thas, O., Vermeulen, K., Vansteelandt, S., and De Neve, J. (2015). Small sample inference for probabilistic index models. *Computational Statistics and Data Analysis*, 110(511):1276–1283.
- Fay, M. and Graubard, B. (2001). Small-sample adjustments for wald-type tests using sandwich estimators. *Biometrics*, 57(4):1198–1206.
- MacKinnon, J. (1985). Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of Econometrics*, 29(3):305–325.
- Miglioretti, D. and Heagerty, P. (2004). Marginal modeling of multilevel binary data with time varying covariates. *Biostatistics*, 5:381–398.
- Morel, J., Bokossa, M., and Neerchal, N. (2003). Small sample correction for the variance of gee estimators. *Biometrical Journal*, 45(4):395–409.
- Thas, O., De Neve, J., Clement, L., and Ottoy, J.-P. (2012). Probabilistic index models. *Journal of the Royal Statistical Society B*, 74(4):623–671.