



Universität
Zürich ^{UZH}

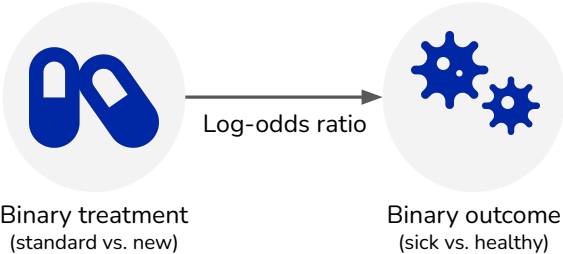
Epidemiology, Biostatistics and Prevention Institute
Susanne Dandl - NORDSTAT 2026

Nonparanormal Adjusted Marginal Inference

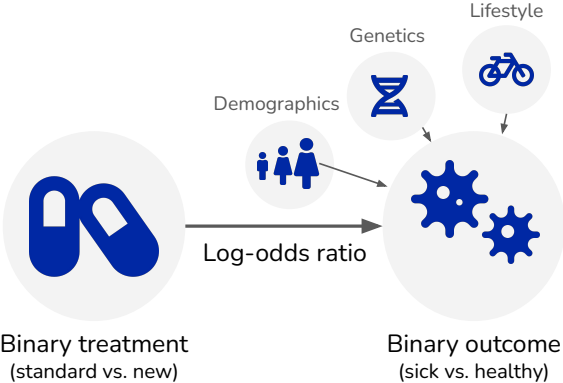
NORDSTAT 2026

Susanne Dandl, Torsten Hothorn

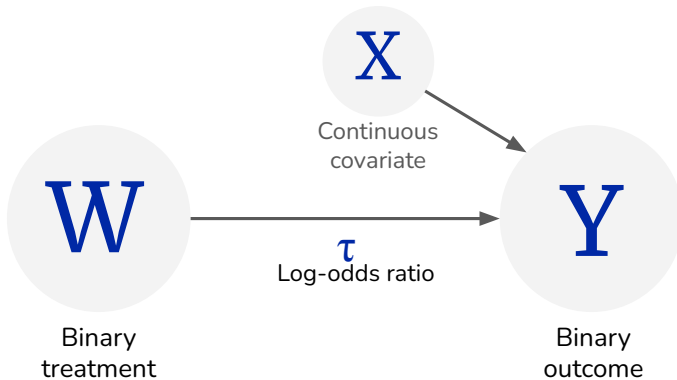
Motivation



Motivation



Motivation



Motivation

Marginal model (glm(w)) $\text{logit}(\mathbb{P}(Y = 1 \mid W = w)) = \alpha - \tau w$

Conditional model (glm(w, x)) $\text{logit}(\mathbb{P}(Y = 1 \mid W = w, X = x)) = \alpha_x - \tau_x w - \beta x$

Motivation

Marginal model (glm(w))

$$\text{logit}(\mathbb{P}(Y = 1 \mid W = w)) = \alpha - \tau w$$

Conditional model (glm(w, x))

$$\text{logit}(\mathbb{P}(Y = 1 \mid W = w, X = x)) = \alpha_x - \tau_x w - \beta x$$

prognostic effect

Motivation

Marginal model (glm(w))

$$\text{logit}(\mathbb{P}(Y = 1 \mid W = w)) = \alpha - \tau w$$

marginal effect

prognostic effect

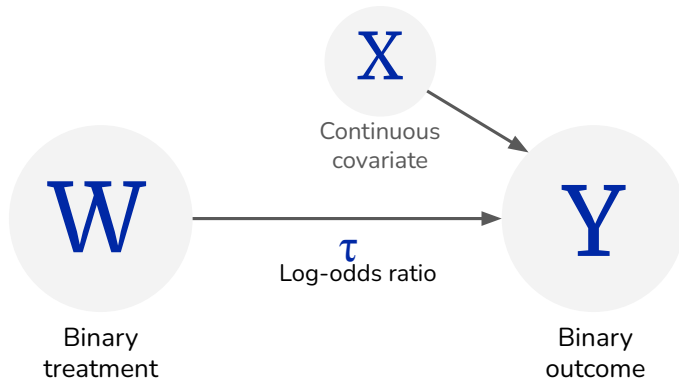
Conditional model (glm(w, x))

$$\text{logit}(\mathbb{P}(Y = 1 \mid W = w, X = x)) = \alpha_x - \tau_x w - \beta x$$

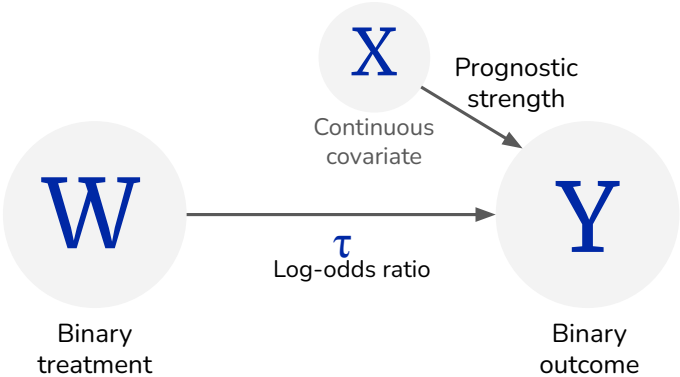
conditional effect

Do these models estimate the same effect,
such that $\tau = \tau_X$?

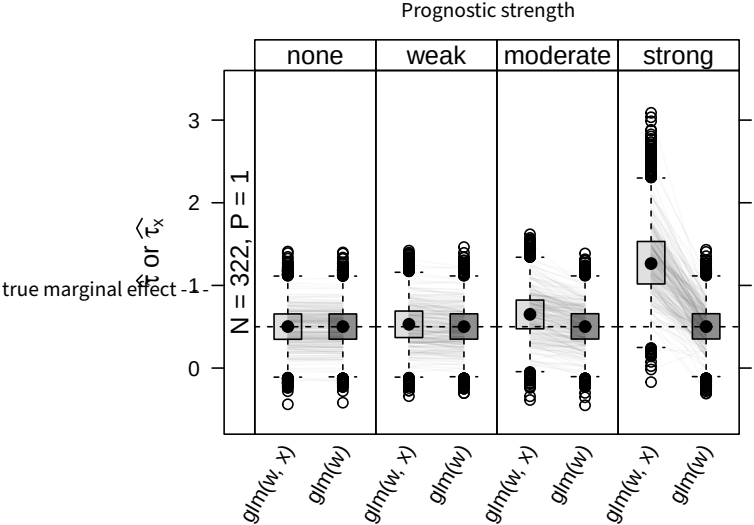
Simulation



Simulation



Simulation



Why $\tau \neq \tau_X$?

Why $\tau \neq \tau_X$?

Noncollapsibility!

Different viewpoints on reasons

Viewpoint: Error variance influences effect estimate [Karlson et al. (2012); McKelvey & Zavoina (1975)]

Viewpoint: Non-linear link/transformation function [Daniel et al. (2020)]

Viewpoint: Omitted variable bias [Hauk et al. (1991)]

⋮

Different viewpoints on reasons

Viewpoint: Error variance influences effect estimate [Karlson et al. (2012); McKelvey & Zavoina (1975)]

Viewpoint: Non-linear link/transformation function [Daniel et al. (2020)]

Viewpoint: Omitted variable bias [Hauk et al. (1991)]

⋮

⇒ Average treatment effect in a normal linear model is not affected!

Which one to prefer?

Marginal vs. conditional effect

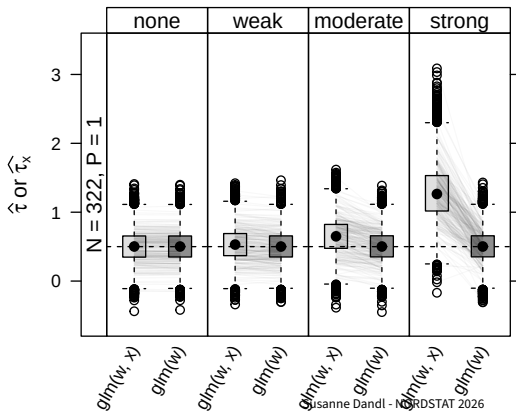
“Conditional and marginal odds ratios are like apples and oranges.” [Daniel et al. (2020)]

Marginal vs. conditional effect

“Conditional and marginal odds ratios are like apples and oranges.” [Daniel et al. (2020)]

Marginal effect

- + Effect estimate that is not influenced by other covariates
- + Comparability
[e.g., Zhang et al. (2008), Daniel et al. (2022)]
- + No fishing expedition
[Tsiatis et al. (2008)]



Marginal vs. conditional effect

“Conditional and marginal odds ratios are like apples and oranges.” [Daniel et al. (2020)]

Algorithm	Prognostic strength			
	none	weak	moderate	strong
glm(w,x)	0.589	0.622	0.711	0.938
glm(w)	0.588			

Table: Estimated power

Conditional effect

- + Improved inference, increased power
[e.g., Robinson & Jewell (1991), Ford et al. (1995)]
- + Correction for baseline imbalance
[e.g., Steyerberg et al. (2000)]

Our mission

“ We want to estimate a **marginal effect** of a treatment of interest whose interpretation is **not influenced by covariates** but benefits from **improved inference** by taking other covariates into account. ”

Strategies

1. Replace the noncollapsible effect measure by a collapsible one
[e.g., risk difference instead of odds ratio (Daniel et al. (2021))]
2. Directly adjust marginal log-odds ratios for prognostic covariates
[e.g., γ -standardisation (Tsiatis et al. (2008), Ye et al. (2024)), TMLE (van der Laan & Rubin (2006))]

Strategies

1. Replace the noncollapsible effect measure by a collapsible one

[e.g., risk difference instead of odds ratio (Daniel et al. (2021))]

2. **Directly adjust marginal log-odds ratios for prognostic covariates**

[e.g., Y-standardisation (Tsiatis et al. (2008), Ye et al. (2024)), TMLE (van der Laan & Rubin (2006))]

We propose to do that with
nonparanormal models!

Nonparanormal Adjusted Marginal Inference (NAMI)

Our method is based on **nonparanormal models** [Liu et al. (2009); Klein et al. (2020)]

1. **Marginal model** for the dependency between Y and W (independent of X !)
2. **Gaussian copula** for the dependency between $Y \mid W$ and X

Nonparanormal Adjusted Marginal Inference (NAMI)

Our method is based on **nonparanormal models** [Liu et al. (2009); Klein et al. (2020)]

1. **Marginal model** for the dependency between Y and W (independent of X !)

$$\mathbb{P}(Y = 1 \mid W = w) = \text{logit}^{-1}(\vartheta_1 - \tau w)$$

2. **Gaussian copula** for the dependency between $Y \mid W$ and X

Nonparanormal Adjusted Marginal Inference (NAMI)

Our method is based on **nonparanormal models** [Liu et al. (2009); Klein et al. (2020)]

1. **Marginal model** for the dependency between Y and W (independent of X !)

$$\mathbb{P}(Y = 1 \mid W = w) = \text{logit}^{-1}(\vartheta_1 - \tau w)$$

2. **Gaussian copula** for the dependency between $Y \mid W$ and X

$$\mathbb{P}(Y = 1 \mid W = w, X = x) = \Phi \left(\beta_1 h_X(x) + \beta_2 \Phi^{-1} \left(\text{logit}^{-1}(\vartheta_1 - \tau w) \right) \right)$$

Nonparanormal Adjusted Marginal Inference (NAMI)

Our method is based on **nonparanormal models** [Liu et al. (2009); Klein et al. (2020)]

1. **Marginal model** for the dependency between Y and W (independent of X !)

$$\mathbb{P}(Y = 1 \mid W = w) = \text{logit}^{-1}(\vartheta_1 - \tau w)$$

2. **Gaussian copula** for the dependency between $Y \mid W$ and X

$$\mathbb{P}(Y = 1 \mid W = w, X = x) = \Phi \left(\beta_1 h_X(x) + \beta_2 \Phi^{-1} \left(\text{logit}^{-1}(\vartheta_1 - \tau w) \right) \right)$$

* In the normal world, noncollapsibility is not a problem!

Nonparanormal Adjusted Marginal Inference (NAMI)

Our method is based on **nonparanormal models** [Liu et al. (2009); Klein et al. (2020)]

1. **Marginal model** for the dependency between Y and W (independent of X !)

$$\mathbb{P}(Y = 1 \mid W = w) = \text{logit}^{-1}(\vartheta_1 - \tau w)$$

2. **Gaussian copula** for the dependency between $Y \mid W$ and X

$$\mathbb{P}(Y = 1 \mid W = w, X = x) = \Phi \left(\beta_1 h_X(x) + \beta_2 \Phi^{-1} \left(\text{logit}^{-1}(\vartheta_1 - \tau w) \right) \right)$$

* Can handle diverse covariate types!

Nonparanormal Adjusted Marginal Inference (NAMI)

Our method is based on **nonparanormal models** [Liu et al. (2009); Klein et al. (2020)]

1. **Marginal model** for the dependency between Y and W (independent of X !)

$$\mathbb{P}(Y = 1 \mid W = w) = \text{logit}^{-1}(\vartheta_1 - \tau w)$$

2. **Gaussian copula** for the dependency between $Y \mid W$ and X

$$\mathbb{P}(Y = 1 \mid W = w, X = x) = \Phi \left(\beta_1 h_X(x) + \beta_2 \Phi^{-1} \left(\text{logit}^{-1}(\vartheta_1 - \tau w) \right) \right)$$

* We can rank covariates by their prognostic value!

Nonparanormal Adjusted Marginal Inference (NAMI)

Our method is based on **nonparanormal models** [Liu et al. (2009); Klein et al. (2020)]

1. **Marginal model** for the dependency between Y and W (independent of X !)

$$\mathbb{P}(Y = 1 \mid W = w) = \text{logit}^{-1}(\vartheta_1 - \tau w)$$

2. **Gaussian copula** for the dependency between $Y \mid W$ and X

$$\mathbb{P}(Y = 1 \mid W = w, X = x) = \Phi \left(\beta_1 h_X(x) + \beta_2 \Phi^{-1} \left(\text{logit}^{-1}(\vartheta_1 - \tau w) \right) \right)$$

* $R^2 := 1 - \beta_2^{-2}$ provides a coefficient of determination!

Nonparanormal Adjusted Marginal Inference (NAMI)

Our method is based on **nonparanormal models** [Liu et al. (2009); Klein et al. (2020)]

1. **Marginal model** for the dependency between Y and W (independent of X !)

$$\mathbb{P}(Y = 1 \mid W = w) = \text{logit}^{-1}(\vartheta_1 - \tau w)$$

2. **Gaussian copula** for the dependency between $Y \mid W$ and X

$$\mathbb{P}(Y = 1 \mid W = w, X = x) = \Phi \left(\beta_1 h_X(x) + \beta_2 \Phi^{-1} \left(\text{logit}^{-1}(\vartheta_1 - \tau w) \right) \right)$$

* $\vartheta_1, \tau, \beta_1, h_X$ and β_2 are jointly estimated by maximum likelihood! [Hothorn (2025)]

Can this be generalized to
other outcome types?

Generalizations

Our method is based on **nonparanormal models** [Liu et al. (2009); Klein et al. (2020), Hothorn (2025)]

1. **Marginal model** for the dependency between Y and W (independent of X !)

Binary: $\mathbb{P}(Y = 1 \mid W = w) = \text{logit}^{-1}(\vartheta_1 - \tau w)$, τ as log-odds ratio

Generalizations

Our method is based on **nonparanormal models** [Liu et al. (2009); Klein et al. (2020), Hothorn (2025)]

1. Marginal model for the dependency between Y and W (independent of X !)

Binary: $\mathbb{P}(Y = 1 \mid W = w) = \text{logit}^{-1}(\vartheta_1 - \tau w)$, τ as log-odds ratio

Normal: $\mathbb{P}(Y \leq y \mid W = w) = \Phi(\vartheta_1 + \vartheta_2 y - \tau w)$, τ as Cohen's d

Generalizations

Our method is based on **nonparanormal models** [Liu et al. (2009); Klein et al. (2020), Hothorn (2025)]

1. Marginal model for the dependency between Y and W (independent of X !)

Binary: $\mathbb{P}(Y = 1 \mid W = w) = \text{logit}^{-1}(\vartheta_1 - \tau w)$, τ as log-odds ratio

Normal: $\mathbb{P}(Y \leq y \mid W = w) = \Phi(\vartheta_1 + \vartheta_2 y - \tau w)$, τ as Cohen's d

⇒ What if Y does not follow a normal distribution?

Generalizations

Our method is based on **nonparanormal models** [Liu et al. (2009); Klein et al. (2020), Hothorn (2025)]

1. Marginal model for the dependency between Y and W (independent of X !)

Binary: $\mathbb{P}(Y = 1 \mid W = w) = \text{logit}^{-1}(\vartheta_1 - \tau w)$, τ as log-odds ratio

Normal: $\mathbb{P}(Y \leq y \mid W = w) = \Phi(\vartheta_1 + \vartheta_2 y - \tau w)$, τ as Cohen's d

⇒ What if Y does not follow a normal distribution?

Continuous: $\mathbb{P}(Y \leq y \mid W = w) = \Phi(h_Y(y) - \tau w)$, τ as mean difference on latent scale

Generalizations

Our method is based on **nonparanormal models** [Liu et al. (2009); Klein et al. (2020), Hothorn (2025)]

1. Marginal model for the dependency between Y and W (independent of X !)

Binary: $\mathbb{P}(Y = 1 \mid W = w) = \text{logit}^{-1}(\vartheta_1 - \tau w)$, τ as log-odds ratio

Normal: $\mathbb{P}(Y \leq y \mid W = w) = \Phi(\vartheta_1 + \vartheta_2 y - \tau w)$, τ as Cohen's d

⇒ What if Y does not follow a normal distribution?

Continuous: $\mathbb{P}(Y \leq y \mid W = w) = \Phi(h_Y(y) - \tau w)$, τ as mean difference on latent scale

Survival: $\mathbb{P}(Y \leq y \mid W = w) = \text{cloglog}^{-1}(h_Y(y) - \tau w)$, τ as log-hazard ratio

General: $\mathbb{P}(Y \leq y \mid W = w) = G(h_Y(y) - \tau w)$

Generalizations

Our method is based on **nonparanormal models** [Liu et al. (2009); Klein et al. (2020), Hothorn (2025)]

1. Marginal model for the dependency between Y and W (independent of X !)

Binary: $\mathbb{P}(Y = 1 \mid W = w) = \text{logit}^{-1}(\vartheta_1 - \tau w)$, τ as log-odds ratio

Normal: $\mathbb{P}(Y \leq y \mid W = w) = \Phi(\vartheta_1 + \vartheta_2 y - \tau w)$, τ as Cohen's d

⇒ What if Y does not follow a normal distribution?

Continuous: $\mathbb{P}(Y \leq y \mid W = w) = \Phi(h_Y(y) - \tau w)$, τ as mean difference on latent scale

Survival: $\mathbb{P}(Y \leq y \mid W = w) = \text{cloglog}^{-1}(h_Y(y) - \tau w)$, τ as log-hazard ratio

General: $\mathbb{P}(Y \leq y \mid W = w) = G(h_Y(y) - \tau w)$

2. Gaussian copula for the dependency between $Y \mid W$ and X

$$\mathbb{P}(Y \leq y \mid W = w, X = x) = \Phi(\beta_1 h_X(x) + \beta_2 \Phi^{-1}(G(h_Y(y) - \tau w)))$$

Does it fulfill our mission?

Our mission

“ We want to estimate a **marginal effect** of a treatment of interest whose interpretation is **not influenced by covariates** but benefits from **improved inference** by taking other covariates into account. ”

Marginal Effect Not Influenced by Covariates

$$\begin{aligned}\int \mathbb{P}(Y \leq y \mid W = w, X = x) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} &= \dots \\ &= \mathbb{P}(Y \leq y \mid W = w)\end{aligned}$$

Marginal Effect Not Influenced by Covariates

$$\begin{aligned}\int \mathbb{P}(Y \leq y \mid W = w, X = x) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} &= \mathbb{E}_{\mathbf{X}}(\mathbb{P}(Y \leq y \mid W = w, X = x)) \\ &= \mathbb{E}_{\mathbf{X}}\left(\Phi\left(\beta_1 h_X(x) + \beta_2 \Phi^{-1}\left(G(h_Y(y) - \tau w)\right)\right)\right) \\ &= \dots \\ &= G(h_Y(y) - \tau w) \\ &= \mathbb{P}(Y \leq y \mid W = w)\end{aligned}$$

Improved Inference

Given: Normally distributed $Y \mid W = w$ and X with correlation $\rho = -\lambda/\sqrt{1+\lambda^2}$

Unadjusted standard error of Cohen's d

$$\text{SE}(\tau) = \sqrt{\frac{2}{N} \left(\frac{\tau^2}{4} + 2 \right)}.$$

Adjusted standard error of Cohen's d

$$\text{SE}(\tau, \lambda) = \sqrt{\frac{2}{N} \left(\frac{(1 + \lambda^2) \tau^2 + 8}{4(1 + \lambda^2)} \right)}.$$

Improved Inference

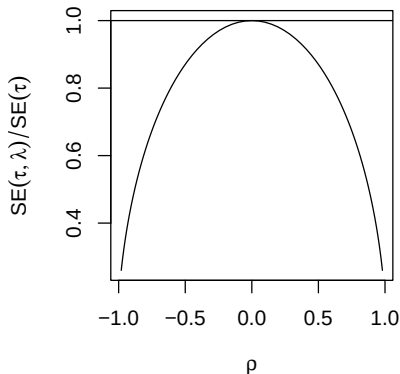
Given: Normally distributed $Y \mid W = w$ and X with correlation $\rho = -\lambda/\sqrt{1+\lambda^2}$

Unadjusted standard error of Cohen's d

$$SE(\tau) = \sqrt{\frac{2}{N} \left(\frac{\tau^2}{4} + 2 \right)}.$$

Adjusted standard error of Cohen's d

$$SE(\tau, \lambda) = \sqrt{\frac{2}{N} \left(\frac{(1 + \lambda^2) \tau^2 + 8}{4(1 + \lambda^2)} \right)}.$$



Improved Inference

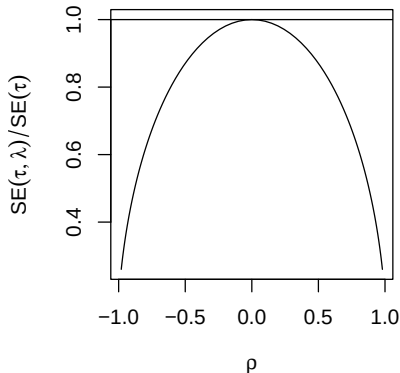
Given: Normally distributed $Y \mid W = w$ and X with correlation $\rho = -\lambda/\sqrt{1+\lambda^2}$

Unadjusted standard error of Cohen's d

$$SE(\tau) = \sqrt{\frac{2}{N} \left(\frac{\tau^2}{4} + 2 \right)}.$$

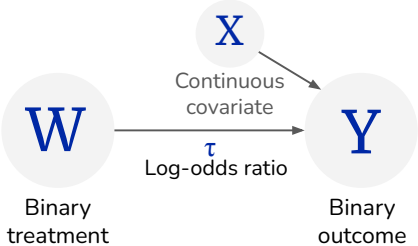
Adjusted standard error of Cohen's d

$$SE(\tau, \lambda) = \sqrt{\frac{2}{N} \left(\frac{(1 + \lambda^2) \tau^2 + 8}{4(1 + \lambda^2)} \right)}.$$

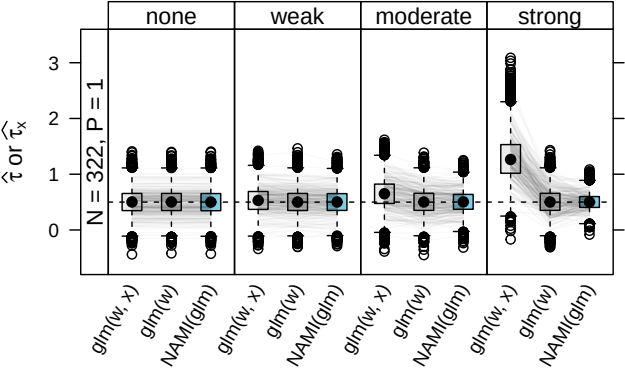


No closed forms are offered for the logistic regression model/odds ratio

Simulation 1 - Binary outcome, log-odds ratio



10000 repetitions
Sample size N for 60% power in unadjusted analysis



Simulation 1 - Binary outcome, log-odds ratio

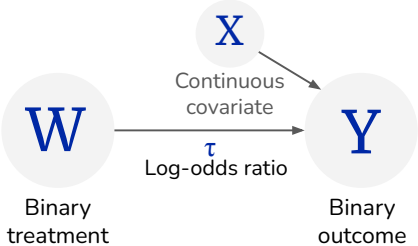
DGP	Algorithm	P	Prognostic strength			
			none	weak	moderate	strong
binary	glm(w)	P = 1	0.588			
	glm(w, x)	P = 1	0.589	0.622	0.711	0.938
	NAMI(glm)	P = 1	0.589	0.625	0.724	0.944

Table: Estimated power

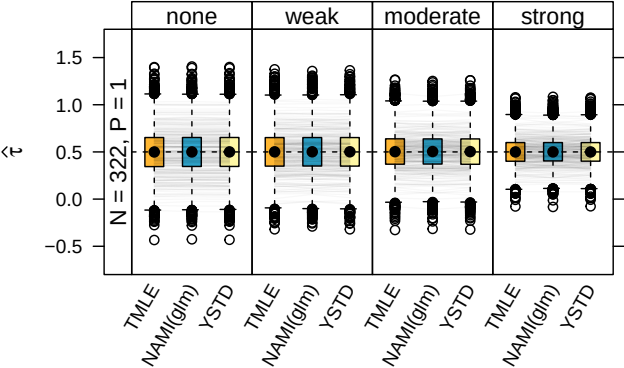
DGP	Algorithm	P	Prognostic strength			
			none	weak	moderate	strong
binary	glm(w)	P = 1	0.049			
	glm(w, x)	P = 1	0.049	0.051	0.050	0.055
	NAMI(glm)	P = 1	0.049	0.053	0.050	0.058

Table: Empirical size

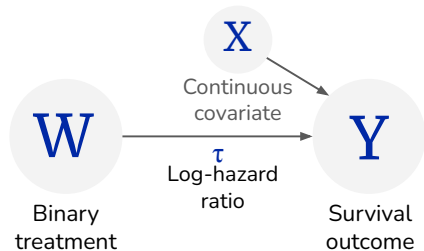
Simulation 1 - Binary outcome, log-odds ratio



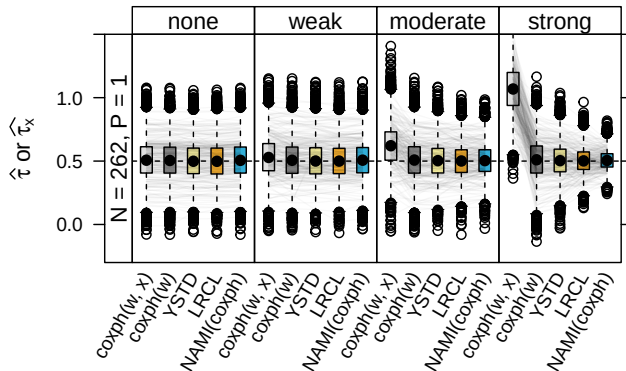
10000 repetitions
Sample size N for 60% power in unadjusted analysis



Simulation 2 - Right-censored survival outcome, log-hazard ratio



10000 repetitions
Sample size N for 60% power in unadjusted analysis



Conclusion

Summary of contributions

- + Marginal effect estimate unaffected by covariates
- + Reduced standard errors $\Rightarrow$ improved inference
- + Works for diverse outcome and covariate types
- + Rankings according to prognostic strength
- + Implementation in `tram` R package available [Hothorn (2020)]

Dandl, S., Hothorn, T.
(2026). Biometrics.
Volume 82, Issue 2,
doi:10.1093/biomtc/ujag065

Conclusion

Summary of contributions

- + Marginal effect estimate unaffected by covariates
- + Reduced standard errors $\Rightarrow$ improved inference
- + Works for diverse outcome and covariate types
- + Rankings according to prognostic strength
- + Implementation in `tram` R package available [Hothorn (2020)]

Limitations

- Empirical size distortions for large P
- Assumes monotonicity of prognostic effect
- Assumes a constant, non-varying treatment effect [Wuethrich & Hothorn (2026)]
- Lack of theoretical assessment of robustness properties

Dandl, S., Hothorn, T.
(2026). Biometrics.
Volume 82, Issue 2,
doi:10.1093/biomtc/ujag065

Misspecification

M1: Survival outcome with a misspecified marginal model [Aalen & Røysland (2015)]

M2: Binary outcome with a misspecified Copula model (quadratic prognostic effect)

Method	Survival M1	Binary M2
MI	0.526	0.440
LTM	0.981	0.436
NAMI	0.904	0.436

Table: Power

Method	Survival M1	Binary M2
MI	0.055	0.052
LTM	0.051	0.051
NAMI	0.036	0.052

Table: Empirical size

Misspecification

M1: Survival outcome with a misspecified marginal model [Aalen & Røysland (2015)]

M2: Binary outcome with a misspecified Copula model (quadratic prognostic effect)

Method	Survival M1	Binary M2
MI	0.526	0.440
LTM	0.981	0.436
NAMI	0.904	0.436
LRCL	0.660	
YSTD		0.436
TMLE		0.429
TMLEXGB		0.429

Table: Power

Method	Survival M1	Binary M2
MI	0.055	0.052
LTM	0.051	0.051
NAMI	0.036	0.052
LRCL	0.052	
YSTD		0.052
TMLE		0.050
TMLEXGB		0.047

Table: Empirical size

Literature I

Karlson, K. B., Holm, A., Breen, R. (2012). Comparing regression coefficients between same-sample nested models using logit and probit: A new method. *Sociological Methodology*, 42(1), 286-313. doi: 10.1177/0081175012444861.

McKelvey, R. D., Zavoina, W. (1975). A statistical model for the analysis of ordinal level dependent variables. *The Journal of Mathematical Sociology*, 4(1), 103–120. doi: 10.1080/0022250X.1975.9989847.

Daniel, R., Zhang, J., Farewell, D. (2021). Making apples from oranges: Comparing noncollapsible effect estimators and their standard errors after adjustment for different covariate sets. *Biometrical Journal*, 63: 528–557. doi: 10.1002/bimj.201900297

Hauck, W. W., Neuhaus, J. M., Kalbfleisch, J. D., Anderson, S. (1991). A consequence of omitted covariates when estimating odds ratios. *Journal of clinical epidemiology*, 44(1), 77–81. doi: 10.1016/0895-4356(91)90203-l

Zhang, M., Tsiatis, A. A., & Davidian, M. (2008). Improving efficiency of inferences in randomized clinical trials using auxiliary covariates. *Biometrics*, 64(3), 707–715. doi: 10.1111/j.1541-0420.2007.00976.x.

Literature II

Tsiatis, A. A., Davidian, M., Zhang, M., & Lu, X. (2008). Covariate adjustment for two-sample treatment comparisons in randomized clinical trials: a principled yet flexible approach. *Statistics in medicine*, 27(23), 4658–4677. doi: 10.1002/sim.3113.

Robinson, L. D., Jewell, N. P. (1991). Some surprising results about covariate adjustment in logistic regression models. *International Statistical Review/Revue Internationale de Statistique*, 59(2), 227–240. doi: 10.2307/1403444

Ford, I., Norrie, J., Ahmadi, S. (1995). Model inconsistency, illustrated by the Cox proportional hazards model. *Statistics in Medicine*, 14(8), 735–746. doi: 10.1002/sim.4780140804.

Steyerberg, E. W., Bossuyt, P.M., Lee, K. L. (2000). Clinical trials in acute myocardial infarction: should we adjust for baseline characteristics? *American Heart Journal*. 139(5):745-51. doi: 10.1016/s0002-8703(00)90001-2.

Van Lancker, K., Bretz, F., Dukes, O (2024). Covariate adjustment in randomized controlled trials: General concepts and practical considerations. *Clinical Trials*. 21(4):399-411. doi: 10.1177/17407745241251568

Literature III

van der Laan, M. J. and Rubin, D. (2006). Targeted maximum likelihood learning. The International Journal of Biostatistics, vol. 2, no. 1. doi: 10.2202/1557-4679.1043.

Ye, T., Shao, J., Yi, Y. (2024). Covariate-adjusted log-rank test: Guaranteed efficiency gain and universal applicability. Biometrika, 111(2), 691–705. doi: 10.1093/biomet/asad045.

Liu, H., Lafferty, J., Wasserman, L. (2009). The Nonparanormal: Semiparametric estimation of high dimensional undirected graphs. Journal of Machine Learning Research, 10, 2295 - 2328.

Klein N., Hothorn, T., Barbanti, L. & Kneib, T. (2022). Multivariate conditional transformation models. Scandinavian Journal of Statistics, 49: 116–142. doi: 10.1111/sjos.12501.

Hothorn, T. (2025). On nonparanormal likelihoods. arXiv preprint arXiv:2408.17346. doi : 10.48550/arXiv.2408.17346.

Hothorn, T. (2020). Most likely transformations: The mlt package. Journal of Statistical Software, 92(1), 1–68. doi: 10.18637/jss.v092.i01.

Literature IV

Wuethrich, L., Hothorn, T. (2026). Joint estimation of marginal and heterogeneous treatment effects. arXiv preprint arXiv:2605.23691.

Aalen, O. O., Cook, R. J., Røysland, K. (2015). Does Cox analysis of a randomized survival study yield a causal treatment effect? Lifetime Data Analysis 21, 579–593.