

# Multinomial Extension of Propensity Score Trimming Methods: A Simulation Study

Kazuki Yoshida<sup>1,2</sup> Daniel H Solomon<sup>2</sup> Sebastien Haneuse<sup>1</sup> Seoyoung C Kim<sup>2</sup> Elisabetta Patorno<sup>2</sup>  
Sara K Tedeschi<sup>2</sup> Houchen Lyu<sup>2</sup> Sonia Hernandez-Diaz<sup>1</sup> Robert J Glynn<sup>1,2</sup>

<sup>1</sup> Harvard T.H. Chan School of Public Health, Boston, MA, USA; <sup>2</sup> Brigham and Women's Hospital, Boston, MA, USA

## Conflict of Interest

KY received support from Harvard T.H. Chan School of Public Health (partially supported by Pfizer, Takeda, Bayer and ASISA) and Honjo International Scholarship Foundation. DHS receives salary support from institutional research grants from Eli Lilly, Amgen, Pfizer, AstraZeneca, Genentech, and Corrona. He also receives royalties from UpToDate, and serves in unpaid roles in studies funded by Pfizer and Eli Lilly. RJJ received research support in the form of grants to his institution from AstraZeneca, Kowa, Novartis, and Pfizer. Other authors declared no COI.

## Background

- Crump et al. (2009), Stürmer et al. (2010) and Walker et al. (2013) have proposed propensity score (PS) trimming methods for binary exposures.
- Stürmer et al. (2010) demonstrated benefits of PS trimming in reducing unmeasured confounding in the presence of unmeasured confounders that were more prevalent in the tails of the PS distribution.
- We extended binary PS trimming methods to the general multinomial settings. We illustrated and examined their characteristics in the visually-tractable 3-group setting.

## Notations

Notation	Explanation
$i \in \{1, \dots, n\}$	Index for an individual
$I = \{1, \dots, n\}$	Index set for entire sample
$A_i \in \{0, \dots, J\}$	Multinomial treatment variable
$X_i$	Covariates
$e_{ji} = P[A_i = j   X_i]$	Propensity score for treatment $j$
$p_j = P[A_i = j]$	Treatment $j$ prevalence
$\pi_{ji} = \frac{p_j}{\sum_{k=0}^J p_k}$	Preference score for treatment $j$ ; prevalence-adjusted version of PS
$F_{e_{ji} A_i}^{-1}(x j)$	Treatment $j$ specific $100 \times x$ percentile value of corresponding PS $e_{ji}$ e.g., $F_{e_{ji} A_i}^{-1}(0.05 j)$ : 5-th percentile of $e_{ji}$ in treatment group $j$
$\alpha_{J,c}, \alpha_{J,s}, \alpha_{J,w}$	Trimming thresholds for $J + 1$ groups

## Definitions and Explanation

### Method Proposed multinomial definition

$$\begin{aligned} \text{Crump } I_{J,c} &= \left\{ i \in I : e_{ji} \geq \alpha_{J,c} \forall j \in \{0, \dots, J\} \right\} \\ \text{Stürmer } I_{J,s} &= \left\{ i \in I : e_{ji} \geq F_{e_{ji}|A_i}^{-1}(\alpha_{J,s}|j) \forall j \in \{0, \dots, J\} \right\} \\ \text{Walker } I_{J,w} &= \left\{ i \in I : \pi_{ji} \geq \alpha_{J,w} \forall j \in \{0, \dots, J\} \right\} \end{aligned}$$

- Multinomial Crump trimming retains subjects who have all PSs above the threshold  $\alpha_{J,c}$ .
- Multinomial Stürmer trimming is asymmetric in that the lower threshold for each PS is different unlike multinomial Crump trimming. The lower threshold is the  $100 \times \alpha_{J,s}$  percentile of each PS in the corresponding treatment group.
- Multinomial Walker trimming is similar to multinomial Crump trimming except the use of a preference score in place of PS and a different threshold value  $\alpha_{J,w}$ .
- We only need the lower threshold for each PS. Trimming the upper tail is implicit because individuals who have a very high PS for one treatment have very low PSs for the other treatments (see figure).
- These definitions reduce to the original definitions when there are only two groups.

## Provisional Thresholds

- We need thresholds ( $\alpha_{J,c}$ ,  $\alpha_{J,s}$ , and  $\alpha_{J,w}$ ) that change with the number of groups ( $J + 1$ ).
- We used the following values as provisional thresholds for illustration.

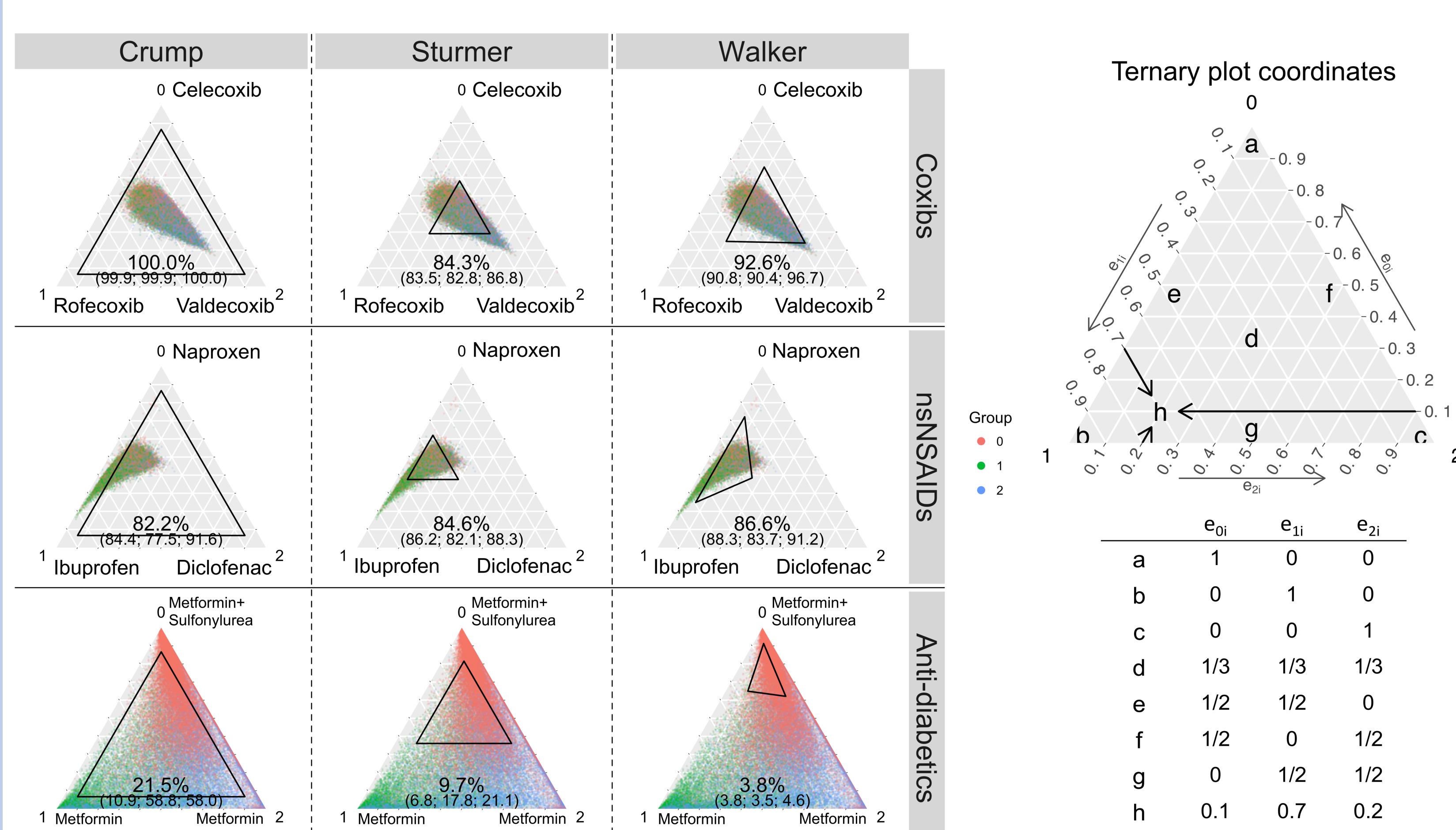
Groups	$J$	Crump ( $\alpha_{J,c}$ )	Stürmer ( $\alpha_{J,s}$ )	Walker ( $\alpha_{J,w}$ )
2	1	0.10	0.050	0.30
3	2	0.07	0.033	0.20
4	3	0.05	0.025	0.15
5	4	0.04	0.020	0.12
	$J + 1$	$\frac{1}{J+1}$	$\frac{1}{J+1}$	$\frac{1}{J+1}$

- Crump lower bounds are on the multinomial propensity score, Stürmer lower bounds are on multinomial propensity score quantile, and Walker lower bounds are on the multinomial preference score.

## Empirical Data Illustration

- We used three observational datasets for illustration (figure): Coxibs (very similar indications; similar group sizes), non-selective NSAIDs (very similar indications; small diclofenac group), and anti-diabetic medications (more different indications; much larger sulfonylurea group).
- These triangular panels are ternary scatter plots of individuals in three groups. Being close to a corner means a high propensity of being in that group. See the coordinate system explanation for what propensity scores correspond to points a through h.
- The inner triangles indicate the region of retained individuals after trimming at the provisional thresholds.

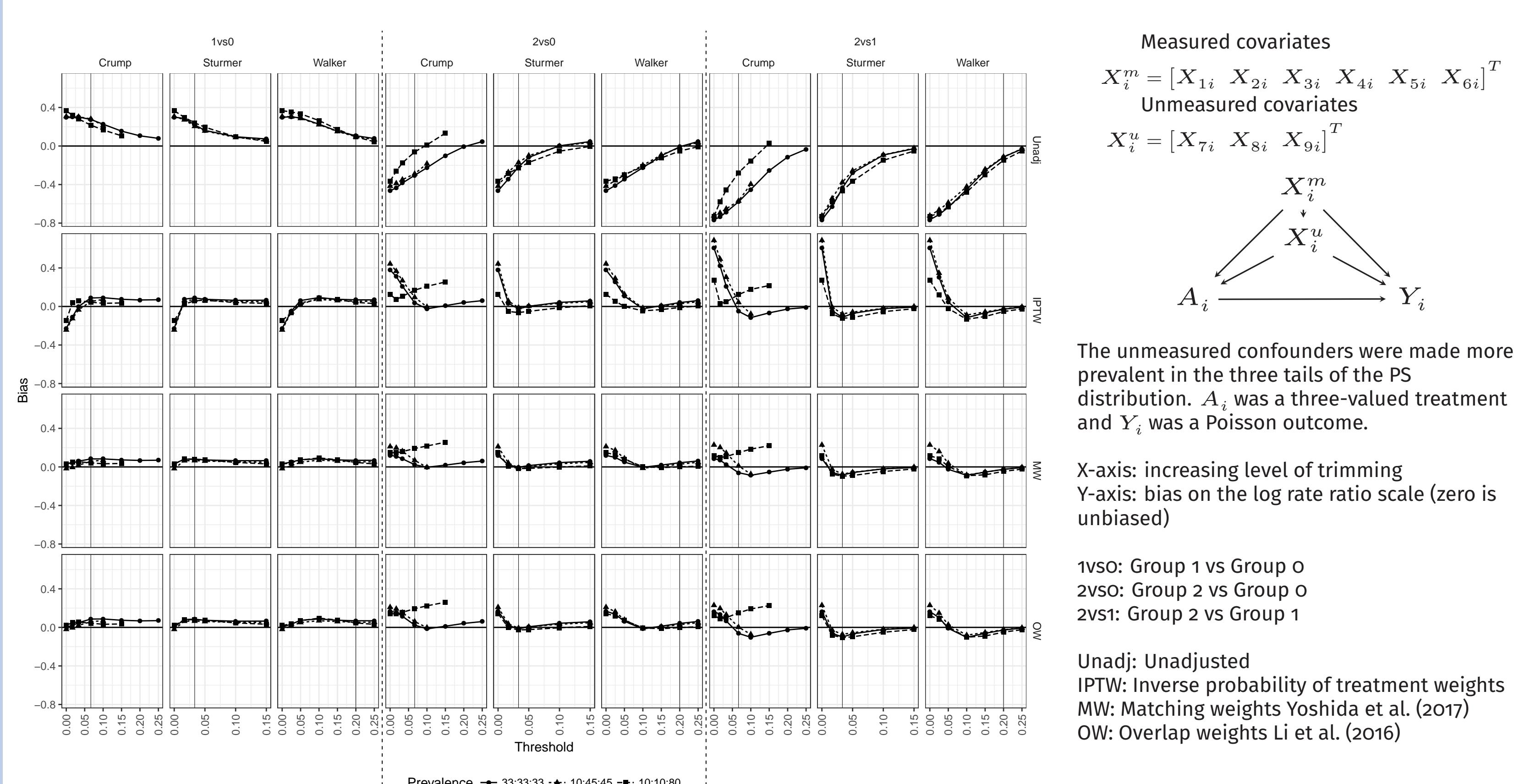
## Empirical Data Illustration (Continued)



- When the groups were similar in patient characteristics as expected in NSAIDs examples, most people were kept in the cohort after trimming.
- Stürmer's and Walker's methods adapted to the skewness in the distributions due to marginal prevalence of treatments.

## Simulation Study

- We conducted simulation to examine bias reduction by trimming in settings in which the tails of PS distributions had unmeasured confounders.



- Bias was reduced by all methods, but Stürmer's and Walker's methods reduced bias more successfully when group sizes were highly imbalanced.
- All methods reduced variance of the IPTW estimator, but not MW and OW estimators.

## Conclusions

- We proposed a multinomial extension of the existing two-group PS trimming methods.
- The extensions of Stürmer and Walker's PS trimming methods reduced bias in 3-group exposure settings even with highly imbalanced treatment frequencies (10:10:80).
- Examining how effect estimates vary at different trimming thresholds can be a useful sensitivity analysis although roles of treatment effect heterogeneity may also need consideration.

## Bibliography

- Crump, R. K., Hotz, V. J., Imbens, G. W., and Mitnik, O. A. (2009). Dealing with limited overlap in estimation of average treatment effects. *Biometrika*, 96(1):187–199.  
Li, F., Morgan, K. L., and Zaslavsky, A. M. (2016). Balancing Covariates via Propensity Score Weighting. *Journal of the American Statistical Association*, 110(511):1–11.  
Stürmer, T., Rothman, K. J., Avorn, J., and Glynn, R. J. (2010). Treatment effects in the presence of unmeasured confounding: Dealing with observations in the tails of the propensity score distribution—a simulation study. *Am. J. Epidemiol.*, 172(7):843–854.  
Walker, A. M., Patrick, A. R., Lauer, M. S., Hornbrook, M. C., Marin, M. G., Platt, R., Roger, V. L., Stang, P., and Schneeweiss, S. (2013). A tool for assessing the feasibility of comparative effectiveness research. *Comp Eff Res*, page 11.  
Yoshida, K., Hernandez-Diaz, S., Solomon, D. H., Jackson, J. W., Gagne, J. J., Glynn, R. J., and Franklin, J. M. (2017). Matching Weights to Simultaneously Compare Three Treatment Groups: Comparison to Three-way Matching. *Epidemiology*, 28(3):387–395.