



Evaluating Pre-trial Programs Using Interpretable Machine Learning Matching Algorithms for Causal Inference



Travis Seale-Carlisle*, Saksham Jain*, Courtney Lee, Caroline Levenson, Swathi Ramprasad, Brandon Garrett, Sudeepa Roy, Cynthia Rudin, Alexander Volfovsky

Background & Motivation



During the pretrial stage, which of the following options is best for *low-risk* individuals?

- Detained in jail
- Released to the community without supervision
- Released to the community with supervision

Pre-trial programs may provide:

- Access to education
- Housing assistance
- Employment assistance
- Healthcare

In exchange for these services, may require:

- Periodic check-ins with a case manager
- Drug testing and treatment
- Wearing an electronic monitor
- Taking part in therapy

They aim to **reduce new criminal activity**. But do they?

Some studies find that they do (e.g., Goldkamp & White, 2006; Lowenkamp & VanNostrand, 2013). Others find they do not (e.g., Cadigan & Lowenkamp, 2011; Robinson et al., 2011). Bechtel et al., 2017 says most studies are not rigorous (purely descriptive, not peer reviewed).

The Task

Evaluate whether the Criminal Justice Resource Center's program in Durham NC reduces future crime.

Treatment (687 people): participation in the program between 2016 – 2019.

Control (728 people): people recommended for the program but didn't take part (judge decides, positivity holds). Pre-trial release w/o supervision.

Outcome: new criminal charge within one year after case disposition.

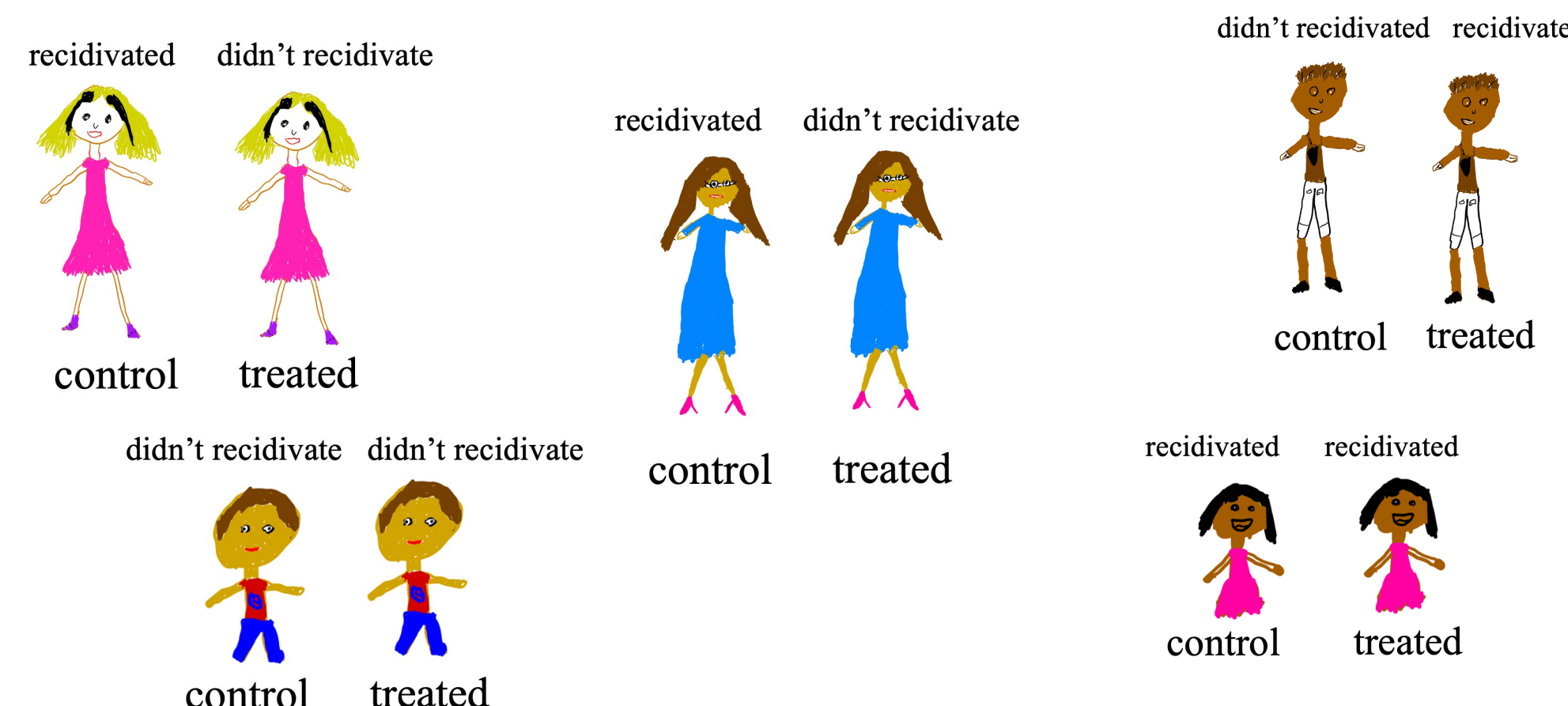
The Data

Criminal history, education level, age, gender, race, stable housing, risk score, etc. of individuals, along with release status, case outcomes, and arrests within one year of case disposition.

Interpretable Causal Analysis

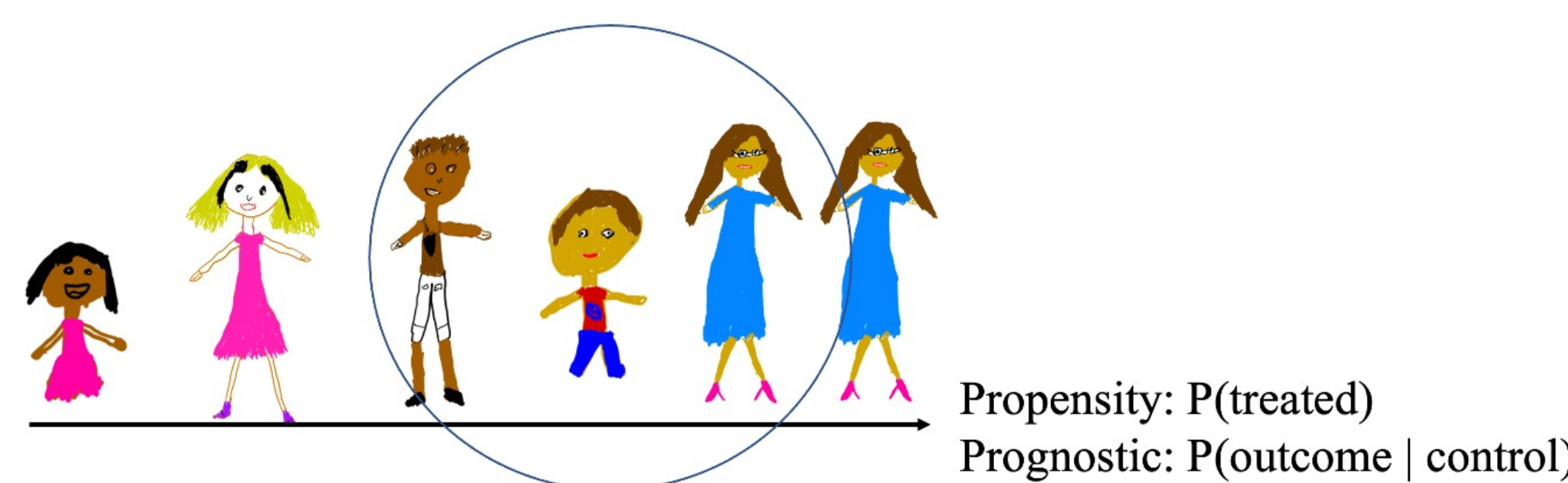
DAME-FLAME Algorithm (from Duke Almost Matching Exactly Lab)

- Provides interpretable matched groups that can be scrutinized by domain experts
- Learns to automatically identify important variables for matching (using ML on a training set)
- Yields accurate causal estimates



Propensity/Prognostic Score Matching

DAME-FLAME is easier to assess for trust than propensity score or prognostic score matching, which produce uninterpretable causal effect estimates: data that are far apart in *covariate space* are lumped together.



Other Common Causal Inference Methods, either:

- Require manual determination of bins for matching, e.g., Coarsened Exact Matching, or manual feature selection.
- Produce inaccurate causal estimates, e.g., multiple linear regression under misspecification.
- Uninterpretable (typical ML methods)

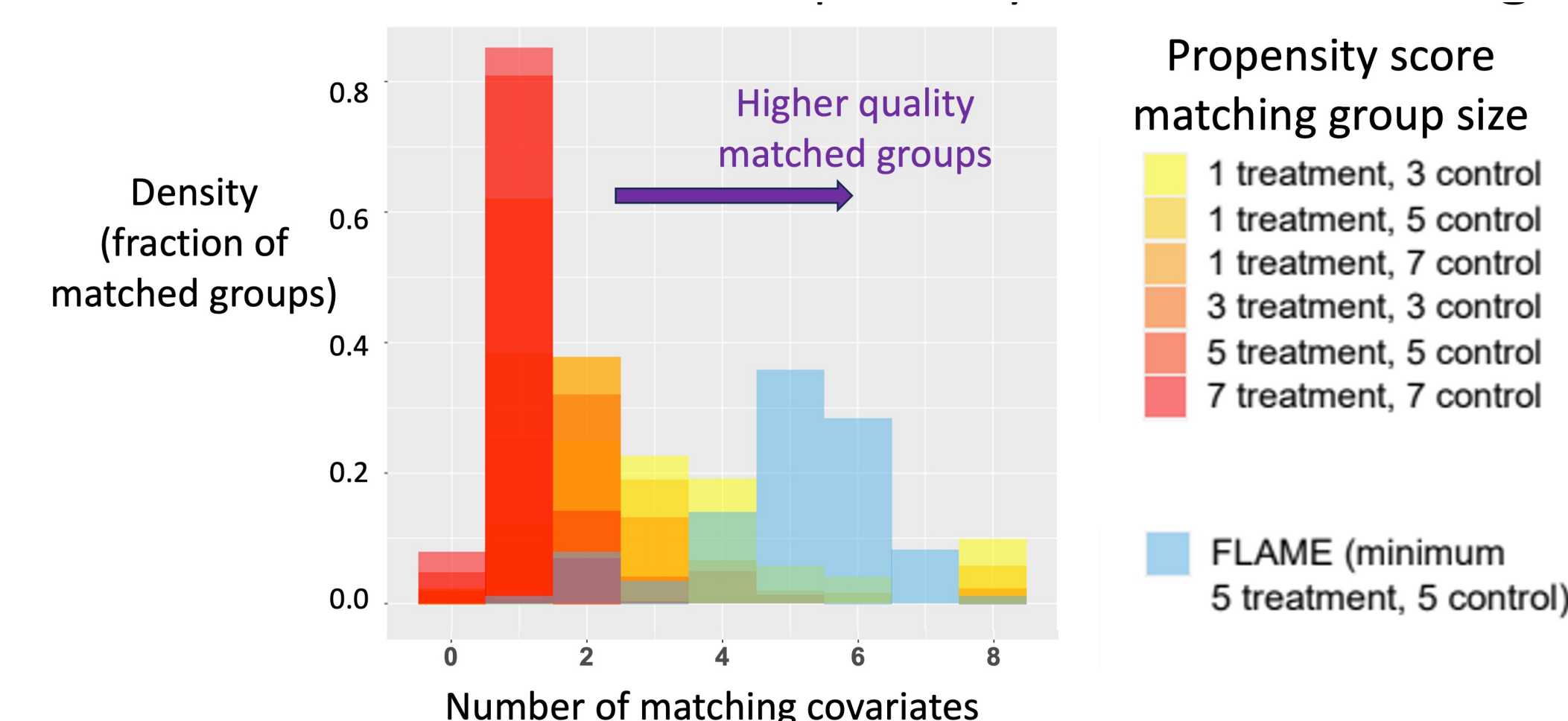
Results & Discussion

Propensity Score Matching - low quality matched groups

Unit	treated	GENDER	RACE	CLASS CHARGE	YEARS AT RESIDENCE	VETERAN	PRIOR ADULT CONVICTIONS	AGE	VPRAI SCORE
924	1	Male	Black	H or I	≥ 1 year	no	0 - 5	18 - 20	3
1397	0	Female	Black	2 or 3	none	no	0 - 5	27 - 29	2
904	0	Female	Black	H or I	< 1 year	no	0 - 5	18 - 20	2
242	0	Female	White	A1 or 1	≥ 1 year	no	0 - 5	35 - 39	0
275	0	Female	Black	2 or 3	< 1 year	no	6 - 10	40 - 44	7
1006	0	Male	Black	H or I	< 1 year	no	6 - 10	35 - 39	6

FLAME-DAME – high quality matched groups

Unit	treated	GENDER	RACE	CLASS CHARGE	YEARS AT RESIDENCE	VETERAN	PRIOR ADULT CONVICTIONS	AGE	VPRAI SCORE
924	1	Male	Black	H or I	≥ 1 year	no	0 - 5	18 - 20	3
25	0	Male	Black	H or I	≥ 1 year	no	0 - 5	27 - 29	3
170	0	Male	Black	H or I	≥ 1 year	no	0 - 5	24 - 27	3
182	0	Male	Black	H or I	≥ 1 year	no	0 - 5	24 - 27	3
289	1	Male	Black	H or I	≥ 1 year	no	0 - 5	45 - 49	3
322	1	Male	Black	H or I	≥ 1 year	no	0 - 5	21 - 23	3
373	0	Male	Black	H or I	≥ 1 year	no	0 - 5	27 - 29	3
408	1	Male	Black	H or I	≥ 1 year	no	0 - 5	24 - 27	3
411	1	Male	Black	H or I	≥ 1 year	no	0 - 5	21 - 23	3
471	0	Male	Black	H or I	≥ 1 year	no	0 - 5	35 - 39	3
512	0	Male	Black	H or I	≥ 1 year	no	0 - 5	30 - 34	3
591	1	Male	Black	H or I	≥ 1 year	no	0 - 5	27 - 29	3
744	0	Male	Black	H or I	≥ 1 year	no	0 - 5	24 - 27	3
748	0	Male	Black	H or I	≥ 1 year	no	0 - 5	21 - 23	3
749	0	Male	Black	H or I	≥ 1 year	no	0 - 5	27 - 29	3
914	1	Male	Black	H or I	≥ 1 year	no	0 - 5	21 - 23	3
985	1	Male	Black	H or I	≥ 1 year	no	0 - 5	21 - 23	3
1015	1	Male	Black	H or I	≥ 1 year	no	0 - 5	24 - 27	3
1039	1	Male	Black	H or I	≥ 1 year	no	0 - 5	21 - 23	3
1059	1	Male	Black	H or I	≥ 1 year	no	0 - 5	40 - 44	3
1130	0	Male	Black	H or I	≥ 1 year	no	0 - 5	24 - 27	3
1255	1	Male	Black	H or I	≥ 1 year	no	0 - 5	18 - 20	3
1323	1	Male	Black	H or I	≥ 1 year	no	0 - 5	24 - 27	3



Main Result

- DAME-FLAME estimated average treatment effect of .018, with 95% CI overlapping zero. Pretrial program had **no significant effect on new criminal charges**.
- This result replicates Bechtel et al. (2017).
- Pretrial programs may **need to consider alternative strategies to reduce new criminal activity**. (Do more studies though.)
- **DAME-FLAME is quite useful in evaluating pre-trial programs** and can be useful in other areas in criminology where random assignment is not possible, ethical, or feasible.

Acknowledgements: National Science Foundation grant IIS-2147061; Wilson Center for Science & Justice, Duke Law School