



Local Causal Discovery for Structural Evidence of Direct Discrimination

Presenter: Jacqueline Maasch

maasch@cs.cornell.edu | arXiv:2405.14848

Jacqueline Maasch¹, Kyra Gan¹, Violet Chen², Agni Orfanoudaki³, Nil-Jana Akpinar^{4*}, Fei Wang⁵

¹Cornell Tech, ²Stevens Institute of Technology, ³University of Oxford,
⁴Amazon AWS, ⁵Weill Cornell Medicine (*Work done outside of Amazon)



AAAI-25 / IAAI-25 / EAAI-25

FEBRUARY 25 – MARCH 4, 2025 | PHILADELPHIA, USA



Presentation overview

1 Background

1. What is causal fairness analysis?
2. How can we detect direct discrimination?
3. A new causal discovery method for practical use.
4. Real-world fairness analysis on clinical data.



Fairness with respect to protected attributes

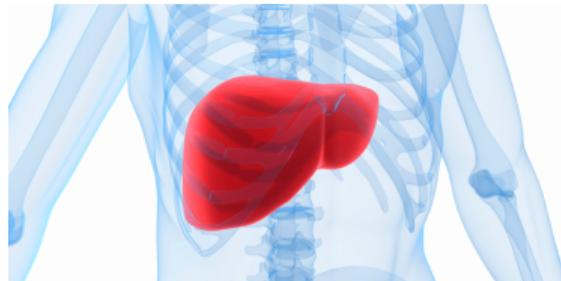
1 Background

- Fairness is essential in policy design and algorithmic decision-making.
- Under the law, **mechanism matters**:
 1. Direct discrimination.
 2. Indirect unfairness.
 3. Spurious unfairness (common cause).
- **Problem**: Statistical associations cannot disentangle mechanisms.
- **Solution**: Causal inference can (*with prior knowledge*).

Case study: Is liver transplant allocation fair?

1 Background

Liver transplantation is a critical therapeutic for acute liver failure.



Sex-based disparities have been observed.^{1,2}

Case study: Is liver transplant allocation fair?

1 Background

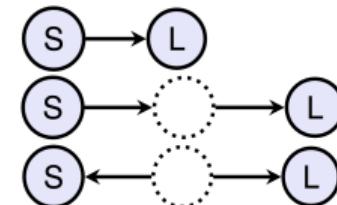
Fairness query: Are sex-based disparities in liver allocation due to direct discrimination?

Graphical query: Is patient sex (S) a causal parent of liver allocation (L)?

Sex is a **parent** (direct cause)

Sex is an **ancestor** (indirect cause)

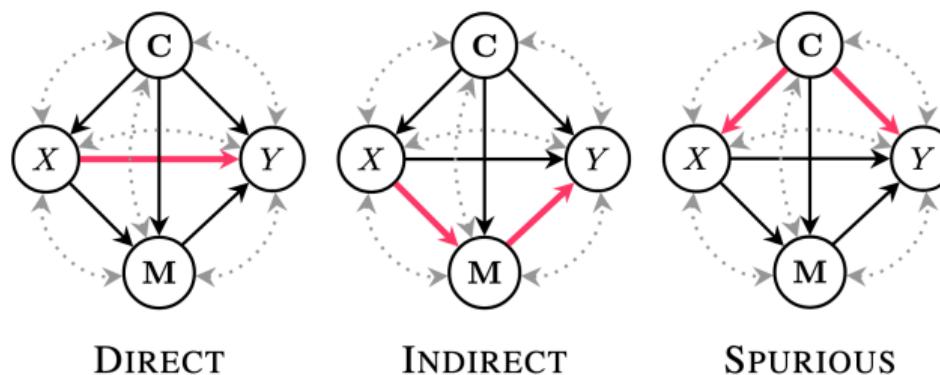
Common cause (spurious)



Causal Fairness Analysis (CFA)

1 Background

Theoretical framework for disentangling mechanisms of unfairness using **structural causal models** (SCMs) and **graphical modeling**.



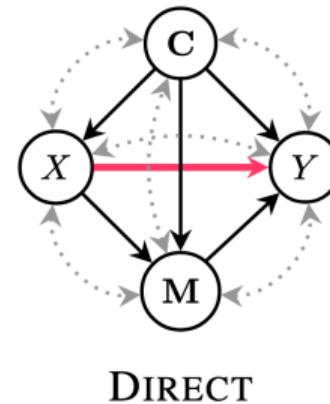
Graphical signatures of direct discrimination

1 Background

1. Structural direct criterion (SDC).³

2. Direct effect estimation.

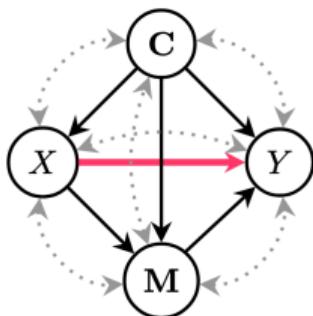
- Controlled direct effect.⁴
- Natural direct effect.⁴
- Counterfactual direct effect.⁵



Structural direct criterion (SDC)

1 Background

Detecting direct discrimination == causal parent discovery.³



DIRECT

$$SDC = \begin{cases} 1 & \text{if } X \text{ is a parent of } Y, \\ 0 & \text{if } X \text{ is not a parent of } Y. \end{cases}$$



Weighted Controlled Direct Effect (WCDE)

1 Background

- **WCDE**: Expected change in outcome as the exposure changes, adjusting for mediators \mathbf{M} (blocking indirect effects).
- Given exposure X , outcome Y , mediators $\mathbf{M}' \in pa(Y)$, and covariates \mathbf{S} that control for confounding of both $\{X, Y\}$ and $\{\mathbf{M}', Y\}$ ⁶,

$$WCDE = \sum_{\mathbf{m}'} (\mathbb{E}[Y \mid do(x, \mathbf{m}')] - \mathbb{E}[Y \mid do(x^*, \mathbf{m}')]) P(\mathbf{m}') \quad (1)$$

$$= \sum_{\mathbf{m}'} \sum_{\mathbf{s}} (\mathbb{E}[Y \mid x, \mathbf{s}, \mathbf{m}'] - \mathbb{E}[Y \mid x^*, \mathbf{s}, \mathbf{m}']) P(\mathbf{s}) P(\mathbf{m}'). \quad (2)$$

WCDE is nonzero if and only if X is a parent of Y .



What if the graph is unknown?

1 Background

We can learn it from observational data via causal discovery.

- Prior methods pose limitations:
 - High sample and time complexity.
 - Disagreement with expert knowledge.^{7,8}
 - Conflicting fairness conclusions.⁹
- **What if we tailor discovery to CFA for direct discrimination?**



LD3: Contributions

2 Local Discovery for Direct Discrimination (LD3)

- **Efficient parent discovery.**
 - Linear no. of conditional independence tests w.r.t. total input variables.
 - On real-world data, baselines required $11\text{--}1021\times$ more tests and $46\text{--}5870\times$ more time compared to LD3.
- **Addresses both indicators of direct discrimination.**
 1. **SDC.**³ Results from LD3 directly evaluate the SDC.
 2. **WCDE.**⁴ LD3 returns a valid adjustment set for the WCDE.
- **Real-world fairness analysis.**
 - LD3 recovered known relations more effectively than baselines.



Learn labels, not the global graph

2 Local Discovery for Direct Discrimination (LD3)

- **Goal:** Learn the relationship of each variable to the protected attribute and outcome to identify parents of outcome.
- **Any variable can take on exactly one of eight causal roles** (labels) w.r.t. a cause-effect pair of interest, as shown in Maasch et al. (UAI'24).¹⁰
- **Local discovery:** We learn these labels, and abstract away the rest.



Learn labels, not the global graph

2 Local Discovery for Direct Discrimination (LD3)

EXHAUSTIVE, DISJOINT CAUSAL PARTITIONS W.R.T. $\{X, Y\}$

- Z_1 Confounders and their proxies.
- Z_2 Colliders and their proxies.
- Z_3 Mediators and their proxies.
- Z_4 Non-descendants of Y where $Z_4 \perp\!\!\!\perp X$ and $Z_4 \not\perp\!\!\!\perp X|Y$.
- Z_5 Instruments and their proxies.
- Z_6 Descendants of Y . All active paths with X are mediated by Y .
- Z_7 Descendants of X . All active paths with Y are mediated by X .
- Z_8 Nodes that share no active paths with X nor Y .

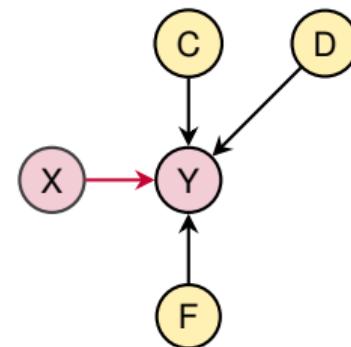
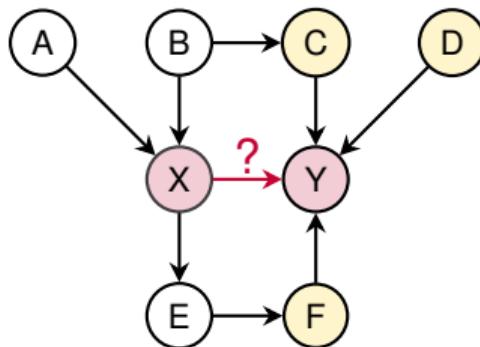
Learning causal partition labels is more computationally efficient than learning the full graph.

Table from Maasch et al. (UAI'24).¹⁰

Learn labels, not the global graph

2 Local Discovery for Direct Discrimination (LD3)

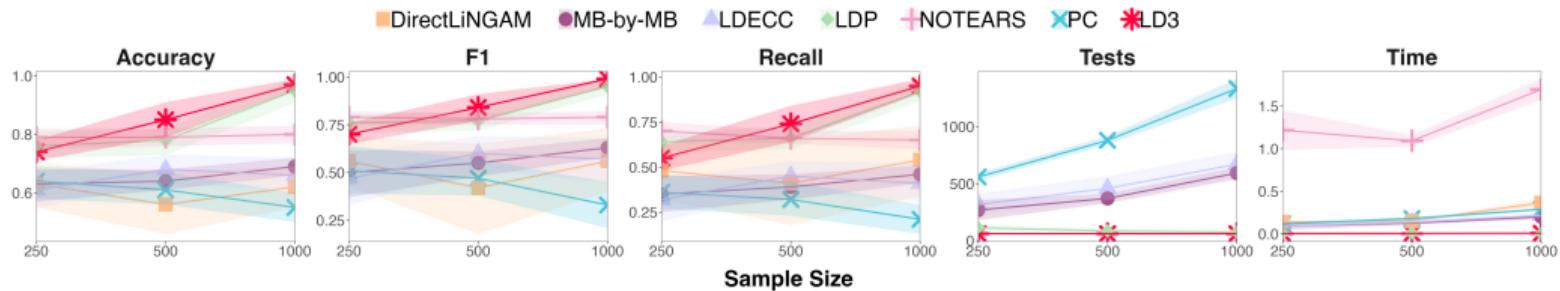
Does the red edge exist? Finding other parents of Y can tell us.



(A) Unknown graph of input data. **(B)** WCDE adjustment set returned by LD3.

LD3: faster, fewer tests, better parent recall

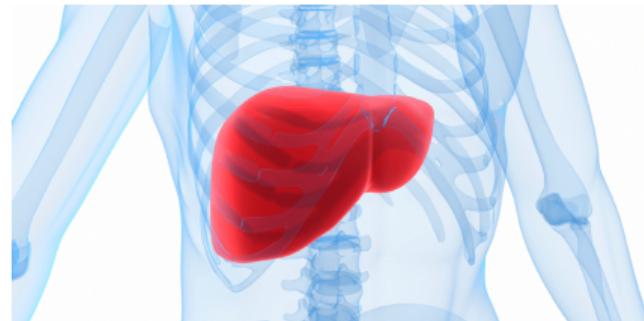
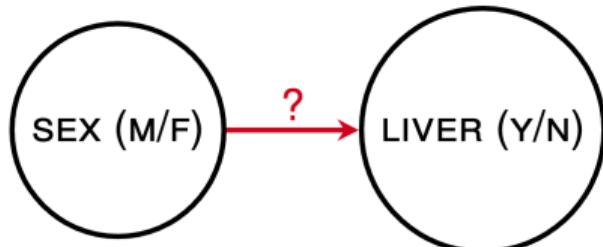
2 Local Discovery for Direct Discrimination (LD3)



Benchmark: Linear-Gaussian model of grape production¹¹ from bnlearn.¹²

Is liver allocation fair?

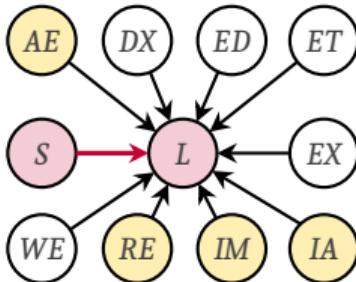
3 Real-world causal fairness analysis: Is liver allocation fair?



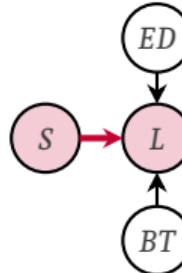
- **Fairness query:** Are sex-based disparities due to direct discrimination?
- **Graphical query:** Is patient sex a causal parent of liver allocation?
- **Data:** Ntl. Standard Transplant Analysis and Research (STAR), 2017-2019.¹³
- **Sample size:** $n = 21,101$ (36% female).

LD3 detects known relations

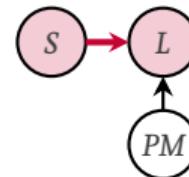
3 Real-world causal fairness analysis: Is liver allocation fair?



LD3
(local discovery)



PC
(global discovery)



LDECC
(local discovery)

SDC = 1, WCDE $\neq 0$: All methods detect sex (S) as a parent of liver allocation (L).¹

Known parents: MELD score (IM), age (IA), region (RE), active exception case (AE).

¹Same independence test, same significance level.



Thank you! Any questions?

`maasch@cs.cornell.edu`

arXiv:2405.14848



`jmaasch.github.io`



References

- [1] O. O. Oloruntoba et al. "Gender-based disparities in access to and outcomes of liver transplantation". In: *World journal of hepatology* 7.3 (2015), p. 460.
- [2] A. M. Allen et al. "Reduced access to liver transplantation in women: role of height, MELD exception scores, and renal function underestimation". In: *Transplantation* 102.10 (2018), pp. 1710–1716.
- [3] D. Plečko et al. "Causal Fairness Analysis: A Causal Toolkit for Fair Machine Learning". In: *Foundations and Trends® in Machine Learning* 17.3 (2024), pp. 304–589.
- [4] J. Pearl. "Direct and Indirect Effects". In: *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*. 2001.
- [5] J. Zhang et al. "Fairness in Decision-Making —The Causal Explanation Formula". In: *Proceedings of the AAAI Conference on Artificial Intelligence* 32.1 (2018). DOI: 10.1609/aaai.v32i1.11564.
- [6] J. Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2000.
- [7] X. Shen et al. "Challenges and Opportunities with Causal Discovery Algorithms: Application to Alzheimer's Pathophysiology". In: *Scientific Reports* 10.2975 (2020).
- [8] A. H. Petersen et al. "Constructing Causal Life-Course Models: Comparative Study of Data-Driven and Theory-Driven Approaches". In: *American Journal of Epidemiology* 192.11 (2023), pp. 1917–1927.



References

- [9] R. Binkytė et al. "Causal discovery for fairness". In: *Workshop on Algorithmic Fairness through the Lens of Causality and Privacy*. PMLR. 2023, pp. 7–22.
- [10] J. Maasch et al. "Local Discovery by Partitioning: Polynomial-Time Causal Discovery Around Exposure-Outcome Pairs". In: *Proceedings of the 40th Conference on Uncertainty in Artificial Intelligence*. 2024. doi: <https://doi.org/10.48550/arXiv.2310.17816>.
- [11] A. Magrini et al. "A conditional linear Gaussian network to assess the impact of several agronomic settings on the quality of Tuscan Sangiovese grapes". In: *Biometrical Letters* 54.1 (2017), pp. 25–42.
- [12] M. Scutari. *Learning Bayesian Networks with the bnlearn R Package*. en. arXiv:0908.3817 [stat]. 2010.
- [13] OPTN. *Data Request Instructions*. Accessed: 2024-05-20. 2024.



Assumptions

- 1. Y has no descendants in the observed variable set.** This is satisfied when Y is a terminal variable in the temporal ordering (e.g., outcome is death, or a policy or algorithmic decision made at a known time point).
- 2. All parents of Y are observed.** Latent variables that are not parents of Y are permissible. Thus, this is a milder condition than assuming causal sufficiency.

Pseudocode

Algorithm 1: *LD3*

Input: Exposure X , outcome Y , variable set \mathbf{Z} , CI test of choice, significance level α .

Output: Adjustment set \mathbf{A}_{DE} , SDC results.

Assumptions: Sufficient conditions A1 and A2.

```

1:  $\mathbf{Z}' \leftarrow \mathbf{Z}$ 
2: for  $\forall Z \in \mathbf{Z}'$  do
3:   if  $Z \perp\!\!\!\perp X \wedge Z \perp\!\!\!\perp Y$  then  $Z \in \widehat{\mathbf{Z}}_8$ 
4:   if  $Z \not\perp\!\!\!\perp Y \wedge Z \perp\!\!\!\perp Y|X$  then  $Z \in \widehat{\mathbf{Z}}_{5,7}$ 
5:   if  $Z \perp\!\!\!\perp X \wedge Z \not\perp\!\!\!\perp X|Y$  then  $Z \in \widehat{\mathbf{Z}}_4$ 
6:  $\mathbf{Z}' \leftarrow \mathbf{Z}' \setminus \widehat{\mathbf{Z}}_8 \cup \widehat{\mathbf{Z}}_{5,7} \cup \widehat{\mathbf{Z}}_4$ 
7: for  $\forall Z \in \mathbf{Z}'$  do
8:   if  $Z \not\perp\!\!\!\perp Y|X \cup \widehat{\mathbf{Z}}_4 \cup \{\mathbf{Z}' \setminus Z\}$ 
     then  $Z \in \widehat{\mathbf{Z}}_{1 \in pa(Y)} \cup \widehat{\mathbf{Z}}_{3 \in pa(Y)}$ 
9: for  $\forall \widehat{Z}_4 \in \widehat{\mathbf{Z}}_4$  do
10:  if  $\widehat{Z}_4 \not\perp\!\!\!\perp Y|X \cup \widehat{\mathbf{Z}}_{1 \in pa(Y)} \cup \widehat{\mathbf{Z}}_{3 \in pa(Y)} \cup \{\widehat{\mathbf{Z}}_4 \setminus \widehat{Z}_4\}$ 
     then  $\widehat{Z}_4 \in \widehat{\mathbf{Z}}_{4 \in pa(Y)}$ 
11:  $\mathbf{A}_{DE} \leftarrow \widehat{\mathbf{Z}}_{1 \in pa(Y)} \cup \widehat{\mathbf{Z}}_{3 \in pa(Y)} \cup \widehat{\mathbf{Z}}_{4 \in pa(Y)}$ 
12: if  $X \perp\!\!\!\perp Y|\widehat{\mathbf{Z}}_{1 \in pa(Y)} \cup \widehat{\mathbf{Z}}_{3 \in pa(Y)}$  then  $SDC \leftarrow 0$ 
13: else  $SDC \leftarrow 1$ 
14: return  $\mathbf{A}_{DE}, SDC$ 

```

Graphical criterion for identifying the WCDE

3.1 A Graphical Criterion for the Weighted CDE

Under conditions where the WCDE is identifiable by Equation 5 (Remark 1), we propose the following criterion.

Definition 8 (Graphical criterion for identifying the WCDE). Under the causal partition taxonomy defined in Maasch et al. (2024), we define the set \mathbf{A}_{DE} that contains all parents of the outcome:

$$\mathbf{A}_{DE} := \mathbf{Z}_{1 \in pa(Y)} \cup \mathbf{Z}_{3 \in pa(Y)} \cup \mathbf{Z}_{4 \in pa(Y)}. \quad (6)$$

Theorem 3 (\mathbf{A}_{DE} is a VAS for the WCDE). \mathbf{A}_{DE} is a valid adjustment set for the WCDE (Equation 5), satisfying the identification conditions in Definition 5.



Variance inflation from unnecessary adjustment

n	A_{DE}		ALL Z	
	Mean	Variance	Mean	Variance
100	6.98	0.83	6.97	7.79
1000	7.17	0.86	7.28	8.27
10000	6.91	1.16	7.25	9.04

n	ALL Z		TRUE A_{DE}		PRED A_{DE}		
	Mean	Variance	Mean	Variance	Mean	Variance	$A_{DE} F1$
500	0.239	0.052	0.347	0.004	0.344	0.004	0.99 [0.98,1.0]
1000	-0.011	0.038	0.35	0.003	0.349	0.003	0.99 [0.98,1.0]
10000	0.151	0.013	0.345	0.000	0.344	0.000	0.99 [0.98,1.0]

Adjusting for the set A_{DE} returned by LD3 prevents unnecessary adjustment. Adjusting for all input variables Z can inflate estimator variance, as illustrated for linear-Gaussian (top) and quadratic-categorical (bottom) data generating processes where $|Z| = 17$ (best values in bold).



STAR liver data

1. **Sex (exposure, protected attribute):** Recipient sex.
2. **Liver allocation (outcome):** Did the candidate receive a liver transplant?
3. **Recipient blood type:** Recipient blood group at registration.
4. **Initial age:** Age in years at time of listing.
5. **Ethnicity:** Recipient ethnicity category.
6. **Hispanic/Latino:** Is the recipient Hispanic/Latino?
7. **Education:** Recipient highest educational level at registration.
8. **Initial MELD:** Initial waiting list MELD/PELD lab score.
9. **Active exception case:** Was this an active exception case?
10. **Exception type:** Type of exception relative to hepatocellular carcinoma (HCC).
11. **Diagnosis:** Primary diagnosis at time of listing.
12. **Initial status:** Initial waiting list status code.
13. **Number of previous transplants:** Number of prior transplants that the recipient received.
14. **Weight:** Recipient weight (kg) at registration.
15. **Height:** Recipient height at registration.
16. **BMI:** Recipient body mass index (BMI) at listing.
17. **Payment method:** Recipient primary projected payment type at registration.
18. **Region:** Waitlist UNOS/OPTN region where recipient was listed or transplanted.

STAR liver data

UNOS POLICY (2017-2019)				
	Female (n = 7679)	Male (n = 13422)	p-value	Test
<i>Active exception case</i>	0.36 (0.73)	0.48 (0.83)	7.241e-28	t-test
<i>Diagnosis 1 (PSC: Primary Sclerosing Cholangitis)</i>	0.03 (0.18)	0.04 (0.2)	0.037	χ^2
<i>Diagnosis 6 (AHF: acute hepatic failure)</i>	0.06 (0.23)	0.02 (0.15)	0.004	χ^2
<i>Diagnosis 7 (Cancer)</i>	0.09 (0.28)	0.16 (0.37)	0.010	χ^2
<i>Height</i>	161.9 (7.46)	175.91 (8.51)	0.000	t-test
<i>Initial MELD</i>	20.5 (10.21)	18.83 (9.87)	1.588e-31	t-test
<i>Payment method</i>	0.53 (0.5)	0.54 (0.5)	0.012	χ^2
<i>Recipient age</i>	54.46 (12.42)	56.03 (10.74)	4.091e-22	t-test
<i>Weight</i>	75.97 (18.53)	90.63 (19.59)	0.000	t-test
UNOS POLICY (2020-2022)				
	Female (n = 8574)	Male (n = 14233)	p-value	Test
<i>Active exception case</i>	0.39 (0.58)	0.43 (0.63)	3.629e-07	t-test
<i>Ethnicity 9 (Multiracial, non-hispanic)</i>	0.01 (0.08)	0.0 (0.07)	0.022	Fisher's exact
<i>Exception type 1 (Unknown)</i>	0.29 (0.45)	0.28 (0.45)	0.022	χ^2
<i>Height</i>	161.96 (7.8)	176.36 (8.25)	0.000	t-test
<i>Initial MELD</i>	22.13 (10.52)	20.99 (10.48)	1.862e-15	t-test
<i>Initial status</i>	0.04 (0.2)	0.02 (0.12)	2.858e-31	t-test
<i>Recipient age</i>	53.83 (12.75)	54.74 (11.64)	3.059e-08	t-test
<i>Weight</i>	75.76 (18.71)	91.1 (20.32)	0.000	t-test

Table D.6: Mean values (standard deviations) for features with statistically significant differences between males and females ($\alpha = 0.05$). Summary statistics for all features are available on GitHub (<https://anonymous.4open.science/r/LD3-4440>).

COMPAS results

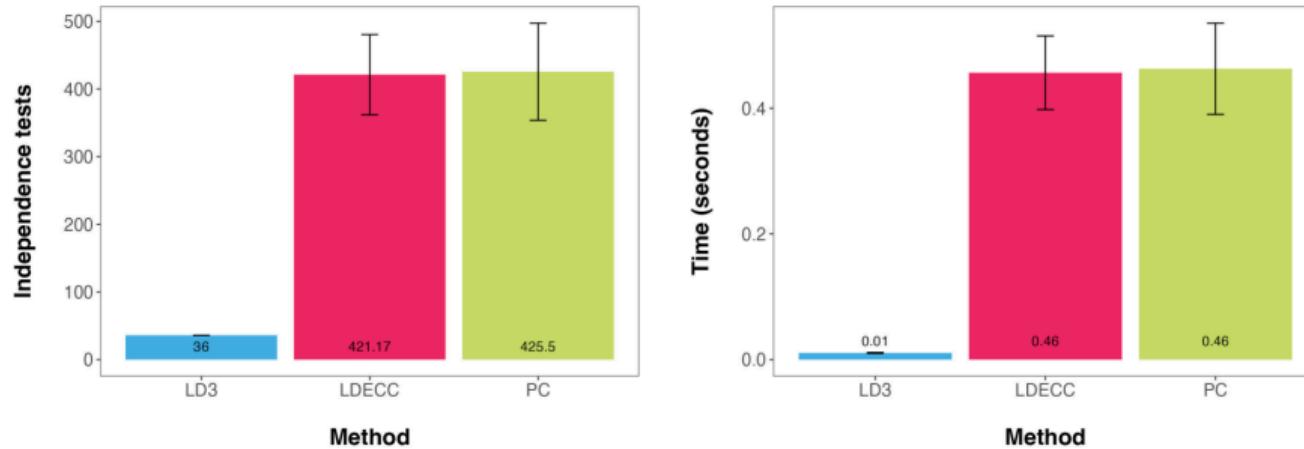


Figure D.2: Average number of independence tests performed and average time (seconds) per method on COMPAS experiments.