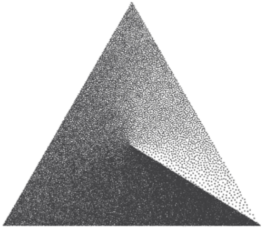
A scalable toolbox for exposing indirect discrimination in a insurance rates

Marie-Pier Côté (U. Laval)

Joint work with
Olivier Côté (U. Laval) **and**
Arthur Charpentier (UQÀM)



Insurance Data Science Conference

June 19th, 2025

UNIVERSITÉ LAVAL

Fairness and equity: no consensus

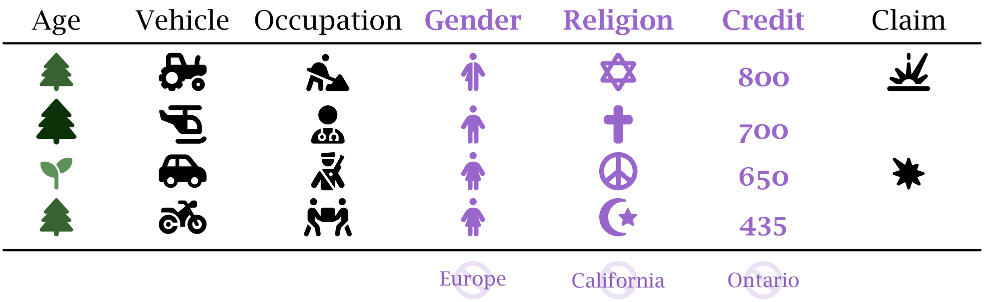
> The strategic goal of the Council of Europe in the field of anti-discrimination, diversity and inclusion is to ensure genuine equality and **full access to rights and opportunities for all members** of society.

2021 Report by the Secretary General of the Council of Europe entitled "State of democracy, human rights and the rule of law: A democratic renewal for Europe"

> The public release of these [Federal Equity Action] plans demonstrated immense public waste and **shameful discrimination.** That ends today. Americans deserve a government committed to **serving every person with equal dignity and respect** [...]

Executive order of The White House issued on January 20, 2025 entitled "Ending Radical And Wasteful Government DEI Programs And Preferencing"

Notation



Allowed variables
X

Prohibited variables
D
(Collected)

Response
Y

Collecting the sensitive variable

“ If you can’t measure it, you can’t manage it.”

– Peter Drucker (1909–2005)

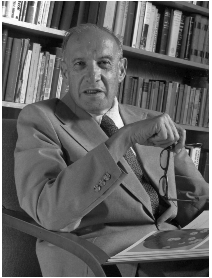


Image by Jeff McNeill under CC-BY-SA 2.0.

Objectives

- Present the three **dimensions** of fairness in actuarial pricing.
- Define pricing benchmarks in line with each dimension.
- Propose a simple metric to quantify proxy discrimination at the individual level.
- Quantify policyholder vulnerability to proxy effects via a **case study**.

This joint work with Olivier Côté and Arthur Charpentier is supported by a Canadian insurance company.

The dimensions of fairness in insurance pricing

The three dimensions of fairness

	Causality	Actuarial fairness	Solidarity
Ideal	Models reflect “true” risk effects	Premiums reflect risk	Risk-sharing across protected groups
Criterion	Proxy parity	Loss ratio parity	Premium parity
Benefit	Awareness of biases	Competitiveness	Protection for vulnerable groups
Mitigation of	Proxy effects	Cross-subsidies	Disparities
Affordability issues for	High-risk policyholders	Vulnerable high-risk policyholders	Non-vulnerable high-risk policyholders

Actuarial Fairness (Arrow, 1963)

A premium is actuarially fair if "it represents an unbiased estimate of the expected value of all future costs associated with the risk transfer" (Casualty Actuarial Society, 1988).

- Self-sustaining loss ratios (no cross-subsidies).
- Avoiding **non risk-based adjustments**.

Solidarity

We leverage the effect of allowed $\mathbf{X}$ on claim Y while aiming for solidarity on D:

equal premiums (in expectation or distribution) across protected groups.

This is referred to as **demographic parity of premiums** (Charpentier et al., 2023; Lindholm et al., 2024b; Charpentier, 2024).

Causality and proxy effects

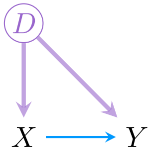
Avoiding proxy effects requires two actions:

- Exclude factors that do not determine risk,
- Limit effect of risk factors to their "**true**" risk relevance.

Even valid risk factors can suffer from proxy effects.

A variable's use – not the variable itself – determines its role as a proxy.

In fairness analysis with respect to D, causality seeks to identify the **effect of X on Y** without **proxy effects from D** (Lindholm et al., 2022; Côté et al., 2025a).



A spectrum of five fairness benchmarks

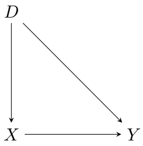
Five fair benchmarks

Premium	Best-estimate	Unaware	Aware	Hyperaware	Corrective
Notation	$\mu^B(\mathbf{x}, d)$	$\mu^U(\mathbf{x})$	$\mu^A(\mathbf{x})$	$\mu^H(\mathbf{x})$	$\mu^C(\mathbf{x}, d)$
Formula	$\mathbb{E}(Y \mid \mathbf{X} = \mathbf{x}, D = d)$	$\mathbb{E}(Y \mid \mathbf{X} = \mathbf{x})$	$\mathbb{E}_D\{\mu^B(\mathbf{x}, D)\}$	$\mathbb{E}\{\mu^C(\mathbf{x}, D) \mid \mathbf{X} = \mathbf{x}\}$	$\mathcal{T}^{d \to \star}\{\mu^B(\mathbf{x}, d)\}$
Direct discrimination	✔	✘	✘	✘	✔
Proxy discrimination	–	✔	✘	✔	–
Demographic disparities	✔	✔	✔	✘	✘
Pillar	**AF**	**AF**	**C**	**S**	**S**

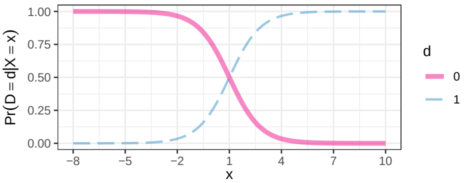
For the real data, we estimate the premium spectrum using:

- `lightGBM` (Ke et al., 2017) to learn conditional expectations,
- empirical marginals of D for population-level integration, and
- optimal transport mappings via `Equipy` (Fernandes Machado et al., 2025).

Ex. 1: setup

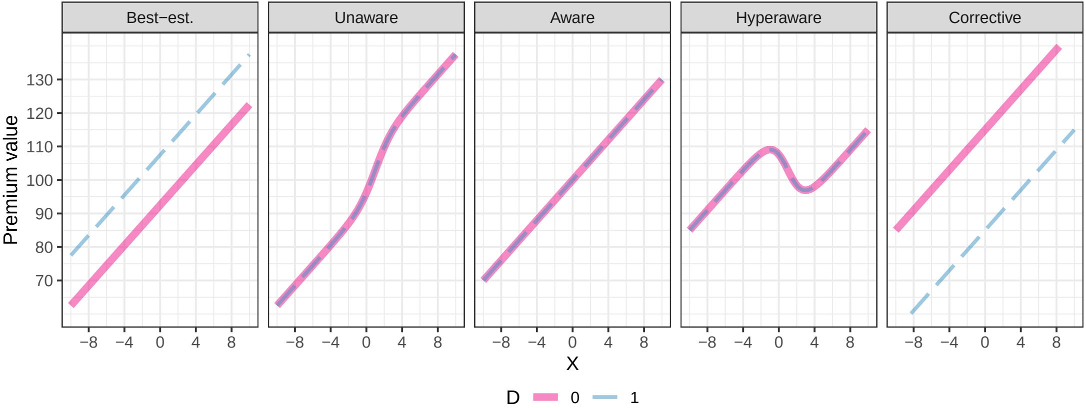


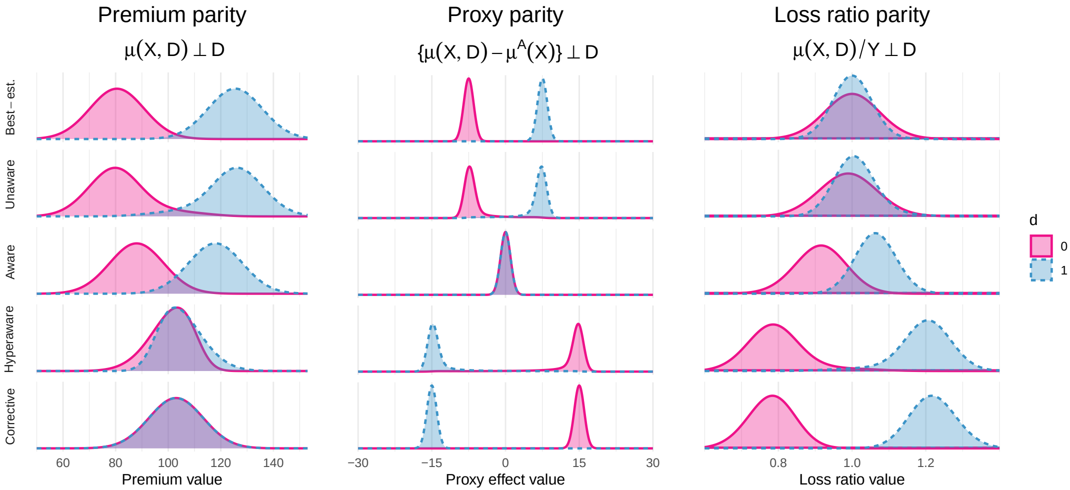
- Let $D \in \{0, 1\}$ be Bernoulli with $\Pr(D = 1) = 0.5$.
- The variables X and Y are Gaussian and the DAG is satisfied.



Ex. 1: Premiums in terms of x and d

Best-estimate Unaware Aware Hyperaware Corrective




Premium parity
μ(X, D) ⊥ D
Proxy parity
{μ(X, D) − μ^A(X)} ⊥ D
Loss ratio parity
μ(X, D)/Y ⊥ D
Best − est.
Unaware
Aware
Hyperaware
Corrective
60
80
100
120
140
Premium value
−30
−15
0
15
30
Proxy effect value
0.8
1.0
1.2
Loss ratio value
d
0
1

Case study: Québec car insurance

Vulnerability to proxies in Québec car insurance

Objective: Quantify **proxy effects** regarding credit score in material damage premiums for at-fault accidents (Chapter B2) in Québec (Canada).

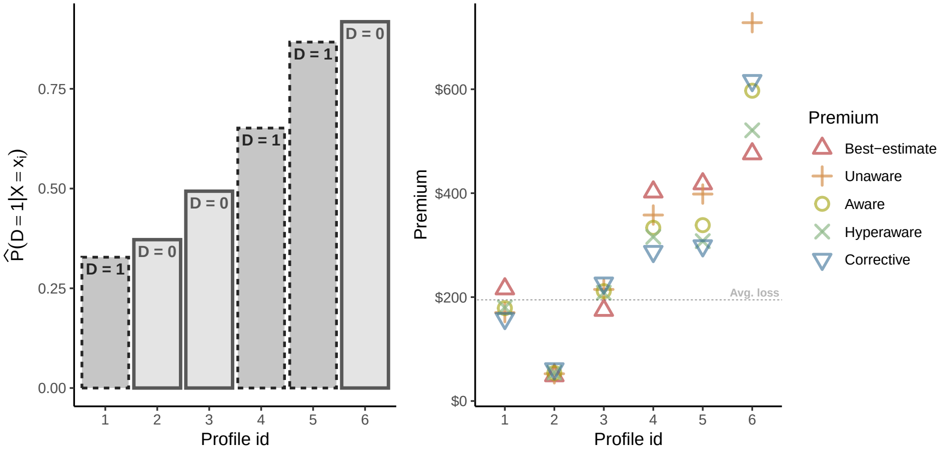
Data: $\approx 768{,}000$ insured vehicles in the province, from 2016–2017. Data obtained via partnership with an insurance company.

Note: Personal data anonymized; strict confidentiality measures applied.

Data overview

Notation	Concept	Domain	Notes
Y	Claim amount ($)	$\mathbb{R}^+$	$\overline{Y} \approx 200$, with 97% at 0
D	Low credit indicator	$\{0, 1\}$	1 indicates low credit, with $\overline{D} \approx 0.40$
$\mathbf{X}$	Policyholder info	Dim. 16	E.g., gender, driving experience, mileage, education, occupation
	Geographic info	Dim. 4	E.g., FSA and territorial risk score
	Vehicle info	Dim. 4	E.g., vehicle age, new purchase, vehicle risk score
	Policy info	Dim. 3	E.g., home insurance, endorsements

Fairness range for six profiles



Proxy vulnerability

The vulnerability of a segment of insureds $\mathbf{x}$ to proxy effect is

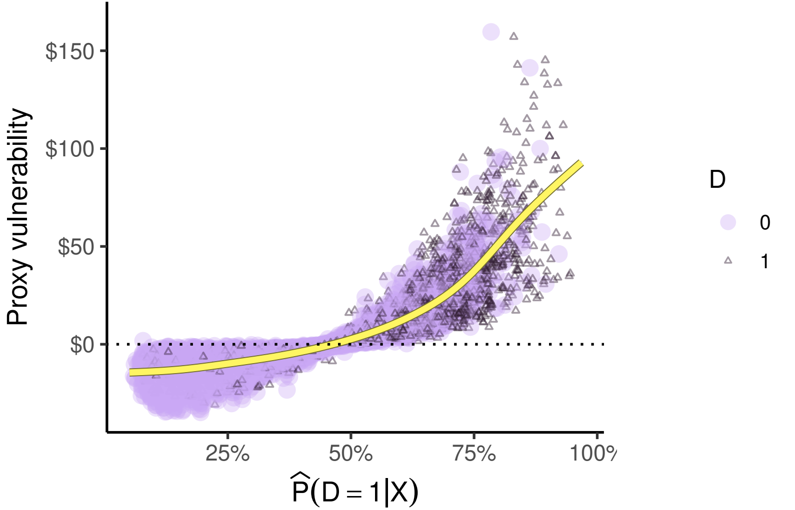
$$\Delta_{\text{proxy}}(\mathbf{x}) = \mu^U(\mathbf{x}) - \mu^A(\mathbf{x}),$$

which we call the **proxy vulnerability**.

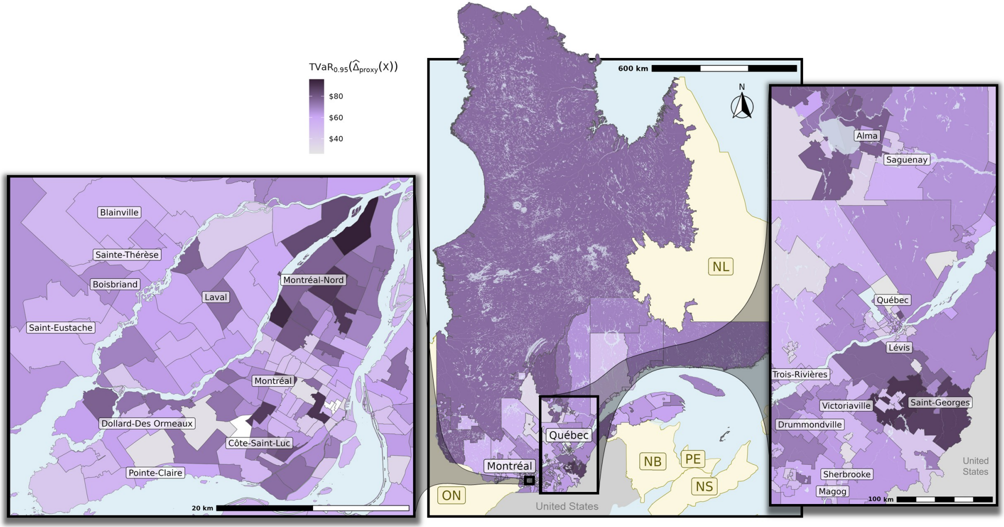
It is the premium difference between not collecting the sensitive variable and *controlling* for it.

Another definition of local proxy metric is proposed by Lindholm et al. (2024a).

Visualising the proxy vulnerability



Geographic distribution of the 95% TVaR of proxy vulnerability



Ingredients of proxy vulnerability

The potential for disparate treatment on D is the **risk spread**:

$$\Delta_{\text{risk}}(\mathbf{x}) = \sup_{d \in \mathcal{D}} \mu^B(\mathbf{x}, d) - \inf_{d \in \mathcal{D}} \mu^B(\mathbf{x}, d).$$

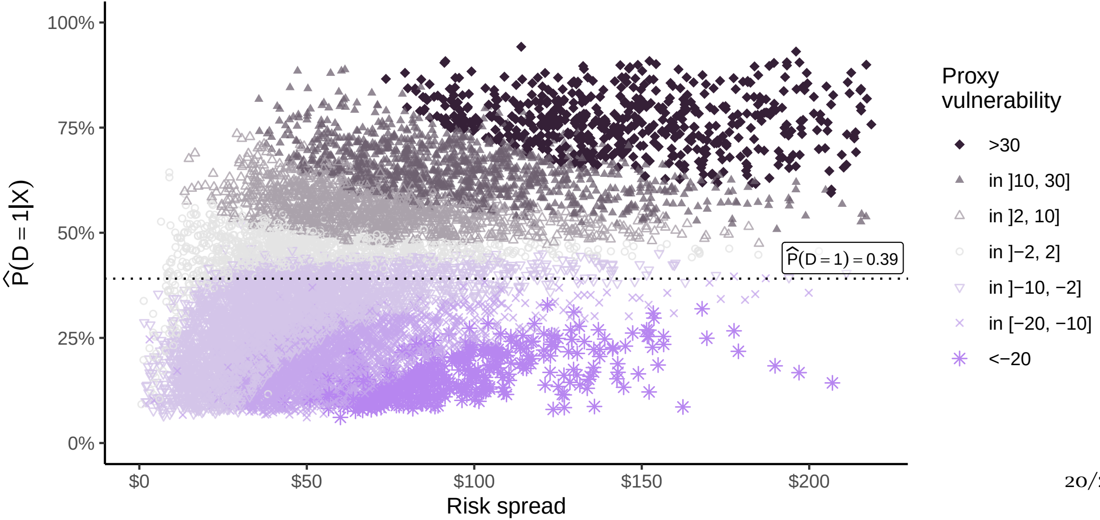
Proxy vulnerability arises from the interplay between

risk spread (potential direct discrimination on D)

and

propensity (ability to exploit it when using only $\mathbf{x}$).

Decomposing proxy vulnerability



Expanding the scope

Conclusion

- Our toolbox contains multiple other metrics derived from the spectrum.
- As data granularity increases, so does the potential for actuarial justification in perpetuating disparities.
- How to align fairness efforts in the market? (Côté et al., 2024)

Thank you 🙌

NSERC CRSNG

mitacs

References i

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review*, 53:941–973.

Casualty Actuarial Society (1988). Statement of Principles Regarding Property and Casualty Insurance Ratemaking. `https://www.casact.org/statement-principles-regarding-property-and-casualty-insurance-ratemaking`.

Charpentier, A. (2024). *Insurance, Biases, Discrimination and Fairness*. Springer, New York.

Charpentier, A., Flachaire, E., and Gallic, E. (2023). Optimal transport for counterfactual estimation: A method for causal inference. In *Optimal Transport Statistics for Economics and Related Topics*, pages 45–89. Springer.

Côté, M.-P., Côté, O., and Charpentier, A. (2024). Selection bias in insurance: why portfolio-specific fairness fails to extend market-wide. *Available at SSRN doi: 10.2139/ssrn.5018749*.

Côté, O., Côté, M.-P., and Charpentier, A. (2025a). A fair price to pay: Exploiting causal graphs for fairness in insurance. *Journal of Risk and Insurance*, 92:33–75.

Côté, O., Côté, M.-P., and Charpentier, A. (2025b). A scalable toolbox for exposing indirect discrimination in insurance rates. *(Submitted)*.

References ii

Fernandes Machado, A., Grondin, S., Ratz, P., Charpentier, A., and Hu, F. (2025). Equipy: Sequential fairness using optimal transport in python. *Available at arXiv doi: arxiv.org/abs/2503.09866*.

Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., and Liu, T.-Y. (2017). Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30.

Lindholm, M., Richman, R., Tsanakas, A., and Wüthrich, M. V. (2022). Discrimination-free insurance pricing. *ASTIN Bulletin*, 52(1):55–89.

Lindholm, M., Richman, R., Tsanakas, A., and Wüthrich, M. V. (2024a). Sensitivity-based measures of discrimination in insurance pricing. *Available at SSRN: ssrn.com/abstract=4897265*.

Lindholm, M., Richman, R., Tsanakas, A., and Wüthrich, M. V. (2024b). What is fair? Proxy discrimination vs. demographic disparities in insurance pricing. *Scandinavian Actuarial Journal*, 2024(9):935–970.