2023 INTERNATIONAL CONGRESS OF ACTUARIES

BRIDGE TO TOMORROW

28 MAY – 1 JUNE 2023 · SYDNEY

Actuaries Institute Australia

IAA AAI International Actuarial Association Association Actuarielle Internationale

BRIDGE TO TOMORROW
2023 INTERNATIONAL
CONGRESS OF ACTUARIES
28 MAY – 1 JUNE 2023 · SYDNEY

Drivers of Mortality-
A Study using Artificial Intelligence and Machine Learning Techniques

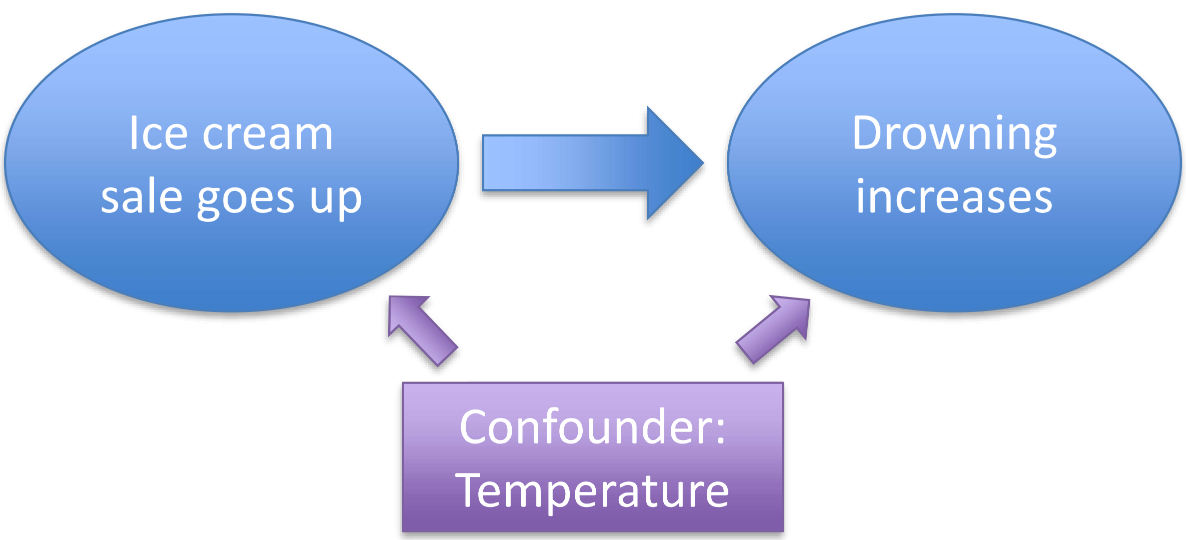
© Iris Deng, Pricing Actuary at Zurich Financial Services, Australia
Simon Kong, Senior Actuary at Allianz Retire+, Australia
Lin Gu, Research Scientist at RIKEN, Japan

Overview and Scope

Correlation vs Causality

- Traditional regression methods (e.g GLM) and machine learning tools uncover the correlation between a treatment (i.e. risk factor) and an outcome.
- However, there are other confounding factors that may influence both the probability of treatment and the outcome.
- Correlation can be misleading when the influence of those confounders are not appropriately controlled.

Correlation Can be Misleading



- When ice cream sale goes up, drowning increase.
- The two variables are positively correlated, but has no causal dependency.
- The missing confounder here is temperature.

Forecasting vs Causal Problems

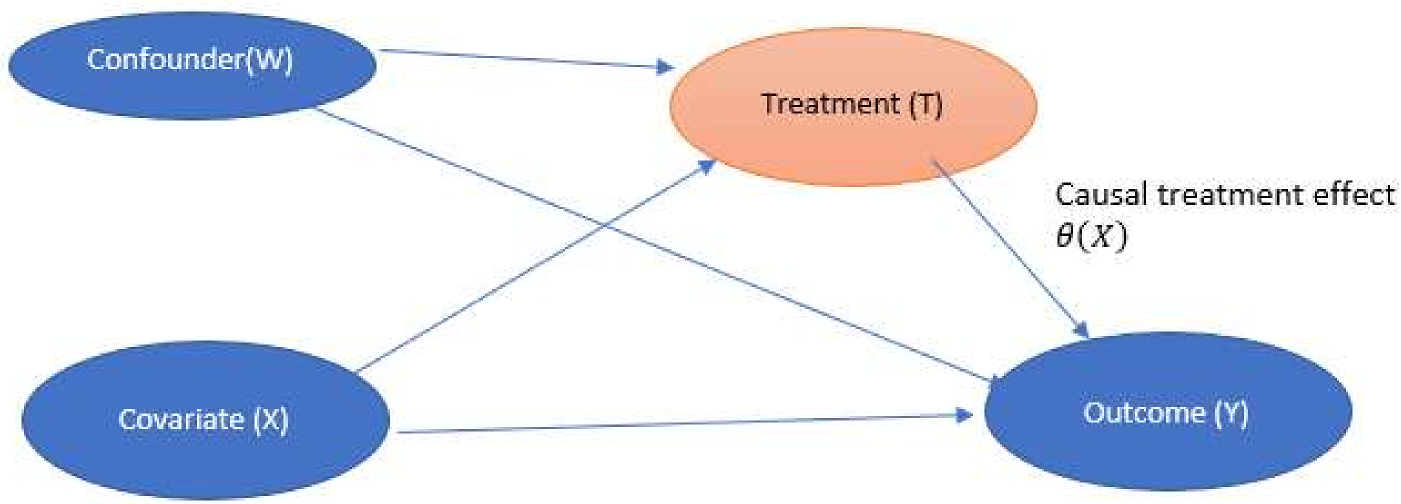
- **Forecasting:** if you want to project future claims cashflows, observed correlation is useful.

- **Causal:** if you want to estimate the effect of a treatment (i.e. risk factor) on the mortality rate, observed correlations are potentially misleading.

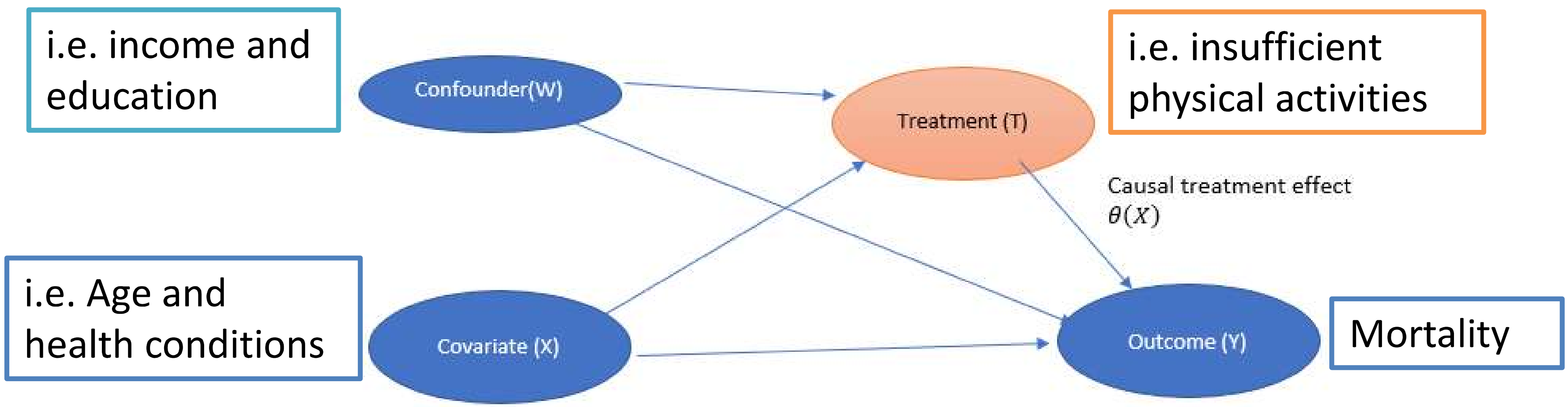
Solving the causal relationship can lead to more appropriate rating factors for underwriting and setting premium rates.

Causal inference – the basics

- Causal impact is defined as the magnitude by which the **outcome (Y)** is changed by a unit-level interventional change in **treatment (T)**.
- **Covariates (X)** and **confounders (W)** are factors that simultaneously influence the likelihood of the treatment (T) and the outcome (Y).
- The causal treatment effect of T on the outcome (Y) is defined as a function of X i.e. $\boldsymbol{\theta}(\boldsymbol{X})$.



Causal inference – example



Problem: What is the effect of insufficient physical activities on mortality i.e. $\theta(X)$?

Causal structure in this example: Probability of physical activities is dependent on both X and W. However, effectiveness of physical activities is only dependent on age and health conditions, and independent of income and level of qualification.

Our Study – Applying Causal Forest DML

Model Used in the Study:

- Double Machine Learning with Causal Forest as the second stage model.
- Random forest and Least Absolute Shrinkage & Selection Operator (LASSO) are the two first stage models considered in our study.

Data:

- Population data by Statistical Area Level 2 (SA2) from the Australian Bureau of Statistics (ABS) and the Australian Institute of Health and Welfare (AIHW).

Adopted Structure:

- **X:** age, gender, health risk factors and disease prevalence/incidence
- **T:** selected variables that are being assessed for causal effect;
- **W:** other influencing factors such as education, married status, occupation
- **Y:** mortality rates

Model Results and Comparison

Treatment Variable	CATE [1]	P-value
insufficient physical activity	3.80	<0.05
mental illness	11.95	<0.1
adequate fruit intake	-0.028	<0.1
smoker	4.45	<0.15
diabetes	0.067	<0.15
kidney disease	0.174	<0.2
obesity	0.037	<0.2
heart disease	0.03	>0.2
lung disease	4.2	>0.2

[1] *CATE of 1 represents an average increase in mortality of 1 per thousand from the treatment.*

	simple average	Random forest	LASSO	Causal Forest
Mean Square Error (MSE)	12.6	2.84	2.87	2.84

Insufficient Physical Activities vs Obesity:

- Insufficient physical activity is a more prominent risk factor for mortality compared to obesity

Mental Illness

- Prominent risk factor for mortality and has a large impact on mortality

Adequate Fruit Intake

- Negative relationship between adequate fruit intake and mortality; causal impact on mortality

Smoker

- Increase mortality, however not significant at 10% level

Potential Application

Rate Setting and Underwriting

- Focus on underlying causal effect instead of potentially misleading correlations
- Improve equity by selecting the appropriate risk factors for premium discrimination and underwriting
- Compared to traditional regressions methods, causal forest can give causal effects that are substantially different

Potential Application

Potential Intervention

- Enables insurers to make more informed decisions on the cost and benefit trade-off of potential interventions e.g. gym membership, education, reward programs to encourage certain behaviours

Applying causal framework can lead to more cost-efficient interventions

Further Research

- Performing the analysis on the insured data instead of population data.
- Consideration of additional risk factors.
- Segmentation of data for analysis by age group and gender.
- Use of other first stage models and alternative causal machine learning models

BRIDGE TO TOMORROW
2023 INTERNATIONAL CONGRESS OF ACTUARIES
28 MAY – 1 JUNE 2023 · SYDNEY

Q & A