On the role of
prognostic factors and effect modifiers
in structural causal models

Rianne M. Schouten

TU/e EINDHOVEN UNIVERSITY OF TECHNOLOGY

We provide insights into the behavior of two types of non-confounding covariates.

- **Prognostic factors:** variables that influence the outcome, but not the treatment effect.
- **Effect modifiers:** variables that influence the treatment effect.

Neither of them influence the treatment assignment, they are not confounders.

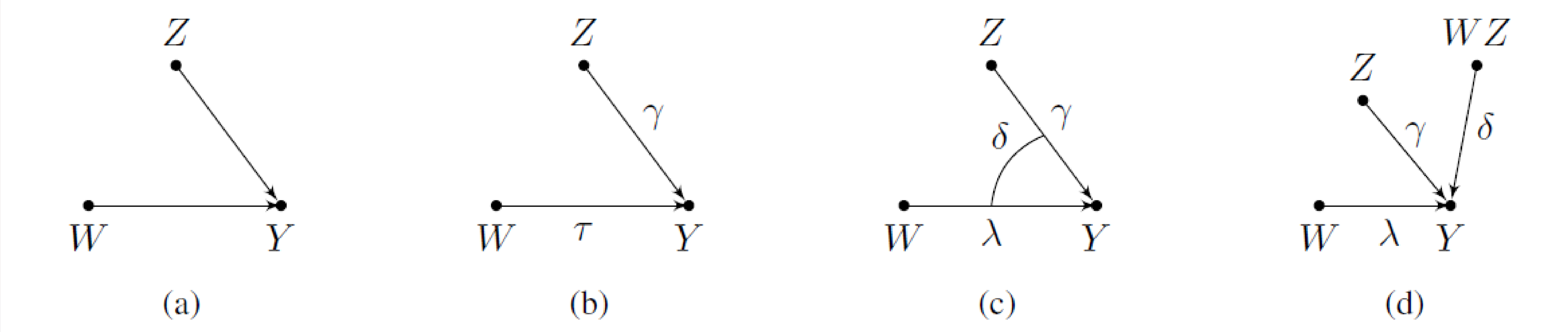


Figure 1: Causal diagrams with three variables: outcome Y, treatment assignment W and covariate Z (a) no structural restrictions; Z can be a prognostic factor, effect modifier or both (b-d) the causal relations are assumed to be linear (b) Z is a prognostic factor (c) Z is an effect modifier, diagram cf. [29] (d) Z is an effect modifier, a variation of the diagram cf. [22].

Controlled Experiment 1

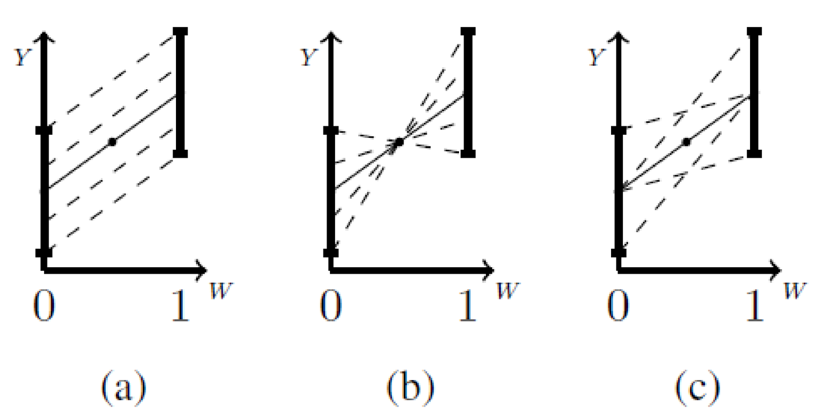
1. Simulate data: ignore fundamental problem of CI.

Control group $y_i^t = \mathrm{N}(5, 2)$ Treatment group $y_i^t = \mathrm{N}(7.5, 2)$

2. Re-order values to create three types of ITE distributions

$$y_i^t = \beta_{10} w_i^t + \beta_{01} z_i + \beta_{11} w_i^t z_i$$

3. Evaluate the variance components of a between-subjects and within-subjects ANOVA



Prognostic factor:	Effect modifier:	Both:
$\beta_{10} = 1$	$\beta_{10} = 1$	$\beta_{10} = 1$
$\beta_{01} \neq 0$	$\beta_{01} = 0$	$\beta_{01} \neq 0$
$\beta_{11} = 0$	$\beta_{11} \neq 0$	$\beta_{11} \neq 0$

Results

	bANOVA	wANOVA (a)	wANOVA (b)	wANOVA (c)
SS_{tot}	1046	1046	1046	706
SS_{group}	219	219	219	219
SS_{ind}	-	814	11	238
SS_{error}	827	13	816	238

Figure 3: Synthetic data results of Experiment 1 in Section 3. The table gives the Sum of Squares (SS) for a bANOVA and a wANOVA for three possible ITE distributions as visualized in Figure 2. More information on bANOVA and wANOVA can be found in Appendix A.

Controlled Experiment 2

1. Simulate data: ignore fundamental problem of CI.

$w_i^t = t$ random walk: (b,a,e,b,c,d,c,a,...)

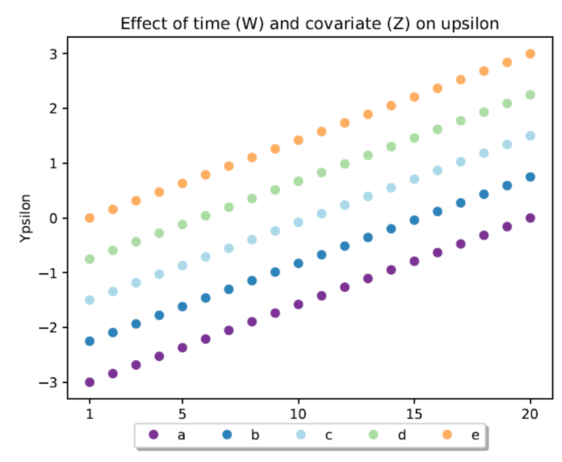
$$v_i^t = \alpha w_i^t + \xi z_i^t + \lambda w_i^t z_i^t$$
$$y_i^t = \mathrm{N}(10 + v_i^t, 0.1)$$

2. Use a local search algorithm to discover groups of individuals with exceptional values (requires aggregation of time-varying values z_i^t)

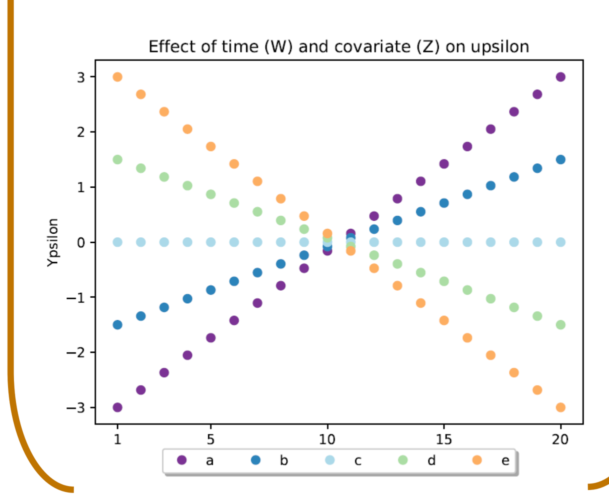
3. Evaluate the top-1 description

Ztype

Prognostic factor: $\alpha = 1, \xi = 1, \lambda = 0$



Effect modifier: $\alpha = 0, \xi = 0, \lambda = 1$



×

Aggfunc

Perfect scenario uses ground-truth.

$$f_{sum} = \sum v_i^t$$
$$f_{inc} = v_i^T - v_i^t$$

Imperfect scenario summarizes the event-sequences by counting the frequency and the average index location of every state value a, b, etc..

$f_{freq_a} \in [1,20]$ $f_{idx_a} \in [1,20]$

$f_{freq_b} \in [1,20]$ $f_{idx_b} \in [1,20]$

etc. etc.

×

Eval

Quantification with φ_μ selects individuals with high average outcome values: ${}^1/_n\, {}^1/_T \sum\sum y_i^t$

Quantification with φ_θ selects individuals with high average Increase from $t = 1$ to $t = T$: ${}^1/_n \sum y_i^T - y_i^1$

Results

Table 1: Description of the most exceptional subgroup, discovered with φ_μ (exceptionally high average outcome) and φ_θ (exceptionally high increase in outcome). Aggregation of event-sequences to single values per individual is done with perfect, ground-truth knowledge and imperfect knowledge. Covariate Z acts as a prognostic factor or effect modifier.

		Aggfunc	
Eval	*Ztype*	imperfect	perfect
φ_μ	prognostic	$f_{\text{freq}_a} \leq 3 \wedge f_{\text{freq}_e} \geq 4 \wedge f_{\text{freq}_b} \leq 5$	$f_{\text{sum}} \in [5, 22]$
	effect modification	$f_{\text{idx}_a} \geq 11 \wedge f_{\text{idx}_e} \leq 10$	$f_{\text{sum}} \in [-134, -25]$
φ_θ	prognostic	$f_{\text{idx}_e} \geq 10.5 \wedge f_{\text{idx}_a} \leq 9.5$	$f_{\text{incr}} = 3$
	effect modification	$f_{\text{freq}_e} \leq 3 \wedge f_{\text{freq}_a} \geq 4 \wedge f_{\text{idx}_e} \leq 14$	$f_{\text{incr}} \in [-38, -29]$

Conclusions

1. bANOVA without additional covariates assumes a worst-case scenario for underlying ITE distributions. Including covariates Z to control for prognostic and effect modification behavior reduces left-over variance and improves precision in estimating (from (b) to (c) to (a)).

2. The quality of individual-level representations of lower-level measurements influences whether or not (our assumptions about) the nature of covariate Z effects higher-level inference making.