Causal Inference 2: Model-based estimation of causal contrasts

Tuesday, 6 June, 2023

Esa Läärä

Statistical Practice in Epidemiology with **R**
2 to 7 June, 2023
University of Tartu, Estonia

Outline

- Causal questions
- Associational contrasts of risks and hazards
- Causal estimands: contrasts of counterfactual quantities
- Outcome regression models
- Standardization or g-formula
- Exposure modelling, propensity scores and weighting
- Double robust estimators and machine learning algorithms

Some literature

- Austin & Stuart (2015) *Stat Med* 34(28):3661-3679.
- Funk et al. (2011) *Am J Epidemiol* 173(7):761767
- Hernan & Robins (2020). *Causal Inference: What if?*. CRC Press.
- Luque Fernandez et al. (2018) *Stat Med* 2018;37(16):2530-2546
- Sjölander (2016) *Eur J Epidemiol* 31:563-574
- Smith et al. (2022) *Stat Med* 2022;41(2):407-432.
- Zhou et al. (2022) `PSweight` vignette.

Causal question in PECOT format & Example

P **Population**: 2900 women with breast cancer (Rotterdam study)

E **Exposure**: Hormonal treatment (HT)

C **Comparator**: Placebo, no HT

O **Outcome**: Recurrence or death

T **Time frame**: 10 y from surgery to outcome

Causal questions of interest – comparisons of counterfactuals:

- What is the 10-year risk π^1 of the outcome, if everybody in P were exposed to HT, as compared with π^0, the risk if nobody were exposed?
- What is the 10-year risk π_1^1 of the outcome, among those in P, who are factually exposed to HT, as compared with the risk π_1^0, if they were not exposed?

Risks by factual exposure and their associational contrasts

- Let T = time to outcome event, $Y(t) = \mathbf{1}_{\{T \leq t\}}$ = binary indicator (1/0) for the outcome to occur by time t, and X be a risk factor,
- Let $\pi_x(t) = E[Y(t)|X = x]$ = risk of the outcome to occur by time t in the subset of the target population factually exposed to level $X = x$:

$$\pi_x(t) = P\{Y(t) = 1 \mid X = x\} = P\{T \leq t \mid X = x\}.$$

- For simplicity, let X be dichotomous & suppress dependence on time, too. Assume also absence of competing events.
- Common **associational contrasts** of risks between exposure groups:
  - **Risk difference** $\tau = \pi_1 - \pi_0 = E(Y|X = 1) - E(Y|X = 0)$,
  - **Risk ratio** $\phi = \pi_1/\pi_0$,
  - **Risk odds ratio** $\psi = \dfrac{\omega_1}{\omega_0} = \dfrac{\pi_1/(1-\pi_1)}{\pi_0/(1-\pi_0)}$.

Hazards by factual exposure and their associational contrasts

- Let the hazard of outcome Y at t among those exposed to level $X = x$ be

$$\lambda_x(t) = \lambda(t \mid X = x) = \lim_{h \to 0} P\{ Y(t+h) = 1 \mid Y(t) = 0, X = x\}/h.$$

- Common associational contrasts:
  - **Hazard difference** $\delta(t) = \lambda_1(t) - \lambda_0(t)$,
  - **Hazard ratio** $\rho(t) = \lambda_1(t)/\lambda_0(t)$.
    This is often assumed constant ρ – as in **Cox regression**.
- Other quantities of interest: contrasts of **restricted mean survival time**

$$\textbf{RMST}(\tau) = \int_0^\tau [1 - \pi_x(t)]\, dt, \;\; x = 0, 1.$$

- So far all these associational contrasts are **marginal** or **unconditional**, meaning that they are not conditioned on or stratified by any covariate Z.

Conditional associational contrasts

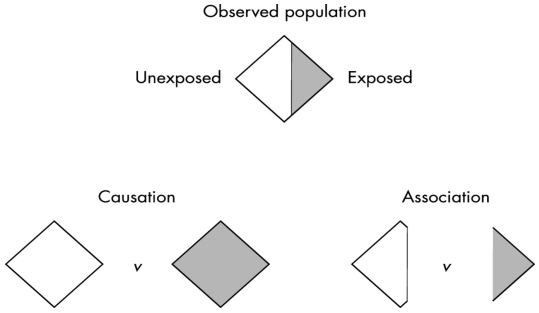
- Let Z be a covariate (can be multivariable) and

$$\pi_{xz} = P\{Y = 1 \mid X = x, Z = z\} = E(Y|X = x, Z = z)$$

be the risk of outcome by t in a population group where $X = x$ and $Z = z$, $x = 0, 1$.

- Let θ be any type of marginal associational contrast, and θ_z be the analogous **conditional associational contrast** between exposed and unexposed among those with $Z = z$.
- For instance, $\tau_z = \pi_{1z} - \pi_{0z}$ is the risk difference conditional on $Z = z$, i.e. z-**specific** risk difference.
- Similarly $\phi_z = \pi_{1z}/\pi_{0z}$ and $\psi_z = \pi_{1z}(1 - \pi_{1z})/[\pi_{0z}(1 - \pi_{0z})]$ are the z-specific risk ratio and odds ratio, respectively.
- Conditional contrasts of hazards: $\delta_z = \lambda_{1z} - \lambda_{0z}$ and $\rho_z = \lambda_{1z}/\lambda_{0z}$, where $\lambda_{xz} = \lambda_{xz}(t) = \lambda(t|X = x, Z = z)$.

Associational and causal contrasts



- **Associational**: Contrast of risks between the **subsets** of the population determined by the subjects' **factual** exposure value.
- **Causal**: Contrast of risks in the **entire population** under the alternative **potential** or **counterfactual** exposure values;
  see Hernan (2004), Hernan & Robins (2006), H&R (2020)

Causal estimands: contrasts of counterfactual risks

- Let $T^{X=x}$, or in short T^x, be time to event, and
  $Y^{X=x}(t) = Y^x(t) = \mathbf{1}_{\{T^x \leq t\}}$ indicate (1/0) event to occur by t,
  if exposure X were – **counterfactually** – forced to value x in the whole target population.
- The **counterfactual** risk for exposure value $X = x$
$$\pi^x(t) = P\{Y^{X=x}(t) = 1\} = E[Y^{X=x}(t)] = P\{T^x \leq t\}$$
- **Marginal causal contrasts** of risk, suppressing dependence on time.
  - risk difference (RD) $\tau^c = \pi^1 - \pi^0 = P\{Y^{X=1} = 1\} - P\{Y^{X=0} = 1\}$,
  - risk ratio (RR) $\phi^c = \pi^1/\pi^0$,
  - risk odds ratio (OR) $\psi^c = [\pi^1/(1-\pi^1)]/[\pi^0/(1-\pi^0)]$,
- Alternative notation: Judea Pearl's (2010) **do-operator**
$$P\{Y = 1|\text{do}(X = x)\} = P\{Y^{X=x} = 1\}.$$

Identifiability of causal contrasts of risks

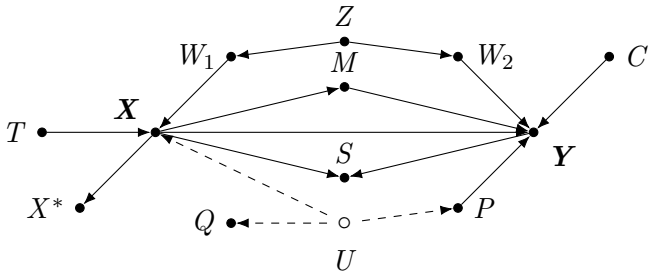
- If in the pertinent causal diagram, there are **open non-causal paths** between X and Y – e.g. due to unmeasured confounders U – the causal contrasts of interest are not identified $\Leftrightarrow$ **residual confounding**.
- Let Z' be a set of observed covariates that are **non-descendants** of X. If $Z \subset Z'$ (discrete) were sufficient to **block** all non-causal paths, then counterfactual risks are identified by **standardization** – or **g-formula**:

$$\begin{aligned}\pi^x &= E(Y^{X=x}) = E_Z[E_Y(Y|X=x,Z)] \\ &= \sum_z P\{Y=1 \mid X=x, Z=z\}P\{Z=z\}.\end{aligned}$$

- Causal contrasts τ^c, ϕ^c, ψ^c are obtained from π^1 and π^0 thus derived.
- If X is randomized, then $X \perp\!\!\!\perp Z \cup U$), and it holds simply

$$\pi^x = P\{Y^{X=x}=1\} = P\{Y=1 \mid X=x\} = \pi_x, \ \forall x, \text{ and e.g.}$$
$$\tau^c = \pi^1 - \pi^0 = \pi_1 - \pi_0 = \tau.$$

Example: Identifying causal contrast from DAG



- **Causal paths** $X \to Y$ and $X \to M \to Y$: Don't block!
- **Non-causal paths** between X and Y: Block!
  – If already blocked, don't open (e.g. by conditioning on $\boxed{S}$).
- **Backdoor paths** $X \leftarrow W_1 \leftarrow Z \to W_2 \to Y$ and $X \leftarrow U \to P \to Y$: Block with minimal effort. – **Sufficient sets**: P plus one from $\{Z, W_1, W_2\}$. – If P unobserved, substitute by Q, proxy of U.
- No need to adjust for T. – Adjusting for C can improve precision.

Conditional causal contrasts

- Let Z be a covariate (can be vector-valued).
  Conditional causal effect of X given $Z = z$ is defined by counterfactual z-conditional risks $\pi_z^x = P\{Y^{X=x} = 1 \mid Z = z\}$.
- These have their own identifiability conditions.
- For instance, $Z =$ sex, or any other variable that divides the target population to interesting subsets or strata.
- **Conditional** or z**-specific causal contrasts** of risks are, for instance

$$\tau_z^c = \pi_z^1 - \pi_z^0 = P\{Y^{X=1} = 1 \mid Z = z\} - P\{Y^{X=0} = 1 \mid Z = z\},$$
$$\phi_z^c = \pi_z^1 / \pi_z^0 = P\{Y^{X=1} = 1 \mid Z = z\} / P\{Y^{X=0} = 1 \mid Z = z\}$$

- If τ_z^c has the same value for all z, RD is **homogenous**. Otherwise it is **heterogenous** or **modified** by Z. – Similarly defined for RR and OR.

Causal contrasts in factual exposure groups

- Causal risk difference **among exposed** is defined
$$\tau_1^c = P\{Y^{X=1} = 1 \mid X = 1\} - P\{Y^{X=0} = 1 \mid X = 1\},$$
also known as **average treatment effect among treated** (ATT).
– The contrast **among unexposed** (ATU) is analogously defined.
- The effect often heterogenous, and groups noncomparable.
- If Z is a sufficient set, g-formulas for these contrasts are
$$\text{ATT} = \pi_1 - \sum \pi_{0z} P\{Z = z|X = 1\} = \text{“observed} - \text{expected”},$$
$$\text{ATU} = \sum \pi_{1z} P\{Z = z|X = 0\} - \pi_0 = \text{“expected} - \text{observed”}.$$
- Different standard populations for ATT, ATU, and for marginal contrast ATE, i.e. **average treatment effect in the whole population**:
$$\text{ATE} = \tau^c = \pi^{X=1} - \pi^{X=0} = \sum_z \pi_{1z} P\{Z = z\} - \sum_z \pi_{0z} P\{Z = z\}.$$

Example: Causal contrasts in exposure groups

- Suppose a population in which 45 % are truly exposed.
- Let a binary Z be sufficient to remove confounding, and
$$P\{X = 1|Z = 1\} = 0.75 \text{ and } P\{X = 1|Z = 0\} = 0.25.$$
- Let also factual risks $\pi_{xz} = P\{Y = 1|X = x, Z = z\}$ by X and Z be as shown in the cells of the table below $(x, z = 0, 1)$

$\Rightarrow$ Values for marginal risks, π_1, π_0, crude contrast $\tau = \pi_1 - \pi_0$, and conditional contrasts $\tau_z = \pi_{1z} - \pi_{0z}$ are shown in margins:

	$Z = 1$	$Z = 0$	π_x
$X = 1$	0.50	0.20	$\pi_1 = 0.40$
$X = 0$	0.25	0.10	$\pi_0 = 0.13$
Contrasts	$\tau_1 = 0.25$	$\tau_0 = 0.10$	$\tau = 0.27$

Example: Causal contrasts in exposure groups (cont'd)

$\Rightarrow$ Observed risks (bolded), counterfactual risks and their causal contrasts are

	True exposure group		
	$x = 1$	$x = 0$	Marginal
Risk $\pi_x^{X=1}$ if exposed	**0.40**	0.26	$\pi^{X=1} = 0.32$
Risk $\pi_x^{X=0}$ if unexposed	0.20	**0.13**	$\pi^{X=0} = 0.16$
Contrasts τ_x^c	ATT $= 0.20$	ATU $= 0.13$	$\tau^c =$ ATE $= 0.16$

- Here, the causal risk difference is bigger among exposed. Thus, being exposurd seems to be a modifier of the effect of exposure on this scale!
- Interestingly, the causal risk ratio is homogenous.

NB Popular design for straightforward estimation of ATT: **matched cohort study**.

Causal contrasts of hazards

- The counterfactual hazard, when everybody were exposed to $X = x$:

$$\lambda^x(t) = \lim_{h \to 0} \frac{1}{h} P\{t < T^x \leq t + h \mid T^x > t\} = \lim_{h \to 0} \frac{1}{h} \frac{P\{Y^x(t+h) = 1\}}{P\{Y^x(t) = 0\}}$$

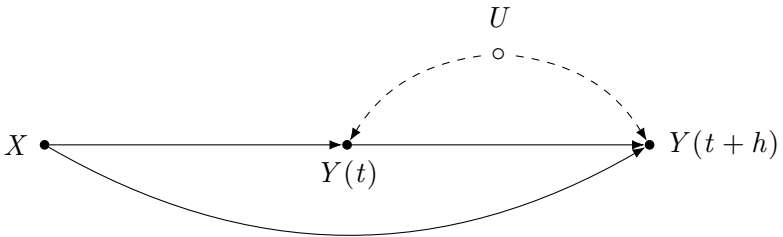
- Marginal causal contrasts of hazards:
  - hazard difference (HD) $\delta^c(t) = \lambda^1(t) - \lambda^0(t)$,
  - hazard ratio (HR)

$$\rho^c(t) = \frac{\lambda^1(t)}{\lambda^0(t)} = \frac{\lim P\{Y^1(t+h) = 1 \mid Y^1(t) = 0\}/h}{\lim P\{Y^0(t+h) = 1 \mid Y^0(t) = 0\}/h}$$

- Identified by corresponding associational contrasts, if X is randomized.
- However, if X has any effect, conditions $Y^1(t) = 0$ and $Y^0(t) = 0$ imply that when $t > 0$, the counterfactual populations at risk will be different.
- Even if exposure groups were comparable at $t = 0$, after that they are not.

Hazard of hazard ratios

- The hazard at any time $t > 0$ is affected by known and unknown causes of the outcome $\Rightarrow$ individual **frailty** U varies in the population.



- $Y(t)$ is a **collider** on the path from X to $Y(t+h)$ via U.
- Conditioning on $Y(t) = 0$ opens this non-causal path, thus inducing bias when evaluating the causal effect of X by $\rho^c(t)$ for $t > 0$.

See e.g. Aalen et al. (2015)

Hazard of hazard ratios (cont'd)

- If exposure to $X = 1$ increases the risk vs. $X = 0$, among the more frail subjects, those exposed tend to get the outcome earlier than the unexposed. Later on, the remaining exposed are less frail than the unexposed.

$\Rightarrow$ Interpretation of hazard ratios in causal terms is problematic, even if $\rho(t)$ were perfectly constant over time (like in Cox regression).

- This problem is less pronounced with rare outcomes than common ones. Yet, causal contrasts of risks are generally preferable to those of hazards
- Moreover, due to frailty, $\rho(t)$ may be far from constant. It can be quite high early on, but then decrease and even dive below 1 at some t^*. Then, it may be hazardous to conclude that the exposure is preventive for $t > t^*$.
- Thus, letting $\rho(t)$ to vary by t in modelling is often adviced.

See e.g. Hernan (2010), Stensrud et al. (2019), Martinussen (2021)

Example: WHI Trial on MHT and CHD (Manson et al. 2003)

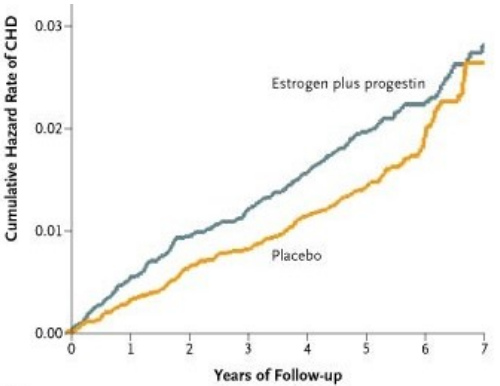
Table 2. Estrogen plus Progestin and the Risk of CHD, According to Year of Follow-up.*

Year of Follow-up	CHD		Hazard Ratio for CHD (95% CI)
	Estrogen-plus-Progestin Group	Placebo Group	
	no. of cases (annualized percentage)		
1	42 (0.50)	23 (0.29)	1.81 (1.09–3.01)
2	38 (0.45)	28 (0.35)	1.34 (0.82–2.18)
3	19 (0.23)	15 (0.19)	1.27 (0.64–2.50)
4	32 (0.39)	25 (0.32)	1.25 (0.74–2.12)
5	29 (0.41)	19 (0.28)	1.45 (0.81–2.59)
≥6	28 (0.37)	37 (0.56)	0.70 (0.42–1.14)

* CHD includes acute myocardial infarction (MI) necessitating hospitalization, silent myocardial infarction as determined by serial electrocardiography, and death due to CHD. There were nine silent myocardial infarctions (four in the estrogen-plus-progestin group and five in the placebo group). Hazard ratios are stratified according to age, presence or absence of a previous coronary event, and randomly assigned diet-modification group and are adjusted for previous coronary-artery bypass grafting or percutaneous transluminal coro-

- Women, 50-79 y, MHT: $N_1 = 8506$, placebo: $N_0 = 8102$
- Followed-up for max 8.6 y, mean 5.6 y.
- Cases & rates/10^4 y $D_1 = 188$, $I_1 = 39$, $D_0 = 147$, $I_0 = 33$.
- Crude IR $= 1.20$, adjusted 1.24 (1.00–1.54)

Example: WHI Trial (cont'd)



No. at Risk								
Estrogen plus progestin	8506	8375	8281	8196	7971	5794	3062	1339
Placebo	8102	8007	7920	7835	7636	5481	2725	988

- Curves of cumulative hazard approximate the development of cumulative risks $\pi_x(t)$ over time.
- In early years, the curve of MHT runs on top, reflecting higher hazard in that period.
- By 6-7 year, cumulative risks appear to have reached same level.

From hazards to causal contrasts of risk

- With censored data and especially when the exposure is not randomized, causal contrasts of risk require a valid model for hazards as function of time, exposure X, and relevant covariates Z.
- Suppose Z blocks all non-causal paths. Then **counterfactual conditional hazards** $\lambda^x(t|Z=z)$ are identified by observable hazards $\lambda_x(t|Z=z)$:

$$\lambda_x(t|Z=z) = \lim_{h\to 0} P\{\, Y(t+h)=1 |\, Y(t)=0, X=x, Z=z\}/h$$

- When Z is continuous or vector-valued, a reasonable model for $\lambda_x(t|Z=z)$ can often be built on e.g. Cox regression or its time-dep extensions.
- Discrete-time hazards $P\{\, Y(t_{k+1})=1 |\, Y(t_k)=0, X=x, Z=z\}$ and binary models on these also used (e.g. by Hernan & Robins).

From hazards to causal contrasts of risk (cont'd)

- When no competing events exist, counterfactual z-conditional risks $\pi^x(t|Z=z)$ are identified from factual z-conditional hazards and risks:
$$\pi^x(t|Z=z) = \pi_x(t|Z=z) = 1 - \exp\left\{-\int_0^t \lambda_x(u|Z=z)du\right\}.$$
- Counterfactual marginal risks are then obtained using the g-formula. If Z is discrete-valued, the formula is
$$\pi^x(t) = \sum_z \pi_x(t|Z=z)P\{Z=z\}.$$
- Thus, marginal risks are again weighted averages of the conditional ones, and use of g-formula corresponds to **direct standardization**.

NB. For those more mathematically oriented: If Z contains discrete and/or continuous variables, the sum is substituted by a **Stieltjes-integral**, and the g-formula is technically expressed as $\pi^x(t) = E_Z[\pi_x(t|Z)] = \int_z \pi_x(t|Z=z)dF(z)$, where $F(z)$ is the joint distribution function of Z in the population.

Causal modelling: Outcome regression

Modelling how expected values, risks, hazards, etc. depend on exposure X and covariates Z (modifiers, and/or confounders). – Common elements:

- Each subject i $(i = 1, \ldots, N)$ has an own **profile**, i.e. vector (x_i, z_i^{T}) of values of X and covariates Z.
- In the spirit of **generalized linear models**, let vector $(\alpha, \beta, \gamma^{\mathsf{T}})$ contain regression coefficients, and specify the **linear predictor** – assuming so far no **interactions**, nor **effect modifications**

$$\eta_i = \alpha + \beta x_i + \gamma^{\mathsf{T}} z_i$$

- **Product terms** can be added for interactions and modifications if needed, and **splines** may be used for continuous covariates.
- Further model specification depends on the type of outcome variable, causal contrasts of interest, and importance and choice of time scale(s).

Binary regression and classical interpretations of coefficients

- Basic model for risks π in fixed risk periods with complete follow-up without censoring and competing events:

$$g\{\pi(x_i)\} = \alpha + \beta x_i + \gamma^{\mathsf{T}} z_i, \quad i = 1, \ldots, N.$$

- Link $g(\cdot)$ and causal interpretation of β, assuming the validity of model (including homogeneity or non-modification of the contrast in question) and that Z blocks all backdoor paths:
  - id $\Rightarrow \beta =$ risk difference (RD) τ^c for $X = 1$ vs. $X = 0$, adjusted for Z
  - log $\Rightarrow \beta =$ log of risk ratio (RR) ϕ^c – " –
  - logit $\Rightarrow \beta =$ log of conditional risk odds ratio (OR), ψ_z^c, – " –
    NB. This is different from marginal OR due to **non-collapsibility**.

- Random component: Binomial family – Fitting: some GLM program.

- Issues with id & log in keeping predicted $\widehat{\pi}(\cdot)$ between 0 and 1.

Binary regression and causal contrasts by g-formula

- Assuming that Z is sufficient to block non-causal paths, a logistic model is fitted, which may even contain product terms allowing modification

$$\text{logit}(\pi_i) = \log[\pi_i/(1-\pi_i)] = \alpha + \beta x_i + \gamma^{\mathsf{T}} z_i + \delta^{\mathsf{T}}(x_i z_i), \quad i = 1, \ldots, n.$$

- For each individual i, predicted risks are computed for both $X =$ and $X = 0$, keeping $Z = z_i$ as it is

$$\widetilde{\pi}_i^{X_i=x} = \text{expit}\{\widehat{\alpha} + \widehat{\beta} x + \widehat{\gamma}^{\mathsf{T}} z_i + \widehat{\delta}^{\mathsf{T}}(x z_i)\}, \quad x = 0, 1.$$

- Marginal counterfactual risks for $x = 1, 0$ are estimated applying **g-formula**:

$$\widehat{\pi}^{X=x} = \widehat{E}_Z[E(Y|X = x, Z)] = \frac{1}{n}\sum \widetilde{E}(Y_i|X_i = x, Z = z_i) = \frac{1}{n}\sum \widetilde{\pi}_i^{X_i=x}$$

as the data provide a non-parametric estimate of the joint distribution of Z

- Estimators of **marginal causal contrasts** of risks are now, e.g.

$$\widehat{\tau}^c = \widehat{\pi}^{X=1} - \widehat{\pi}^{X=0}, \qquad \widehat{\psi}^c = [\widehat{\pi}^1/(1-\widehat{\pi}^1)]/[\widehat{\pi}^0/(1-\widehat{\pi}^0)]$$

Exposure modelling, propensity scores and weighting

Let X be a binary exposure variable. Assume again that Z is a sufficient set

- **Exposure model** predicting individual X_i:s by confounders is fitted

$$\text{logit}[P\{X_i = 1|Z = z_i\}] = \alpha^* + z_i^{\mathsf{T}}\gamma^*, \quad i = 1, \ldots, N.$$

- **Propensity scores** PS_i, or fitted probabilities of exposure:

$$\text{PS}_i = \widehat{P}\{X_i = 1|Z = z_i\} = \text{expit}(\widehat{\alpha}^* + z_i^{\mathsf{T}}\widehat{\gamma}^*).$$

- Individual **weights** $W_i = w(\text{PS}_i, X_i)$ are computed (see next slide).
- Counterfactual risks are estimated as weighted averages of the outcome in the two exposure groups

$$\widehat{\pi}^{X=x} = \frac{\sum_{i=1}^n \mathbf{1}_{\{X_i=x\}} W_i Y_i}{\sum_{i=1}^n \mathbf{1}_{\{X_i=x\}} W_i} = \frac{\sum_{X_i=x} W_i Y_i}{\sum_{X_i=x} W_i}, \quad x = 0, 1$$

- From these, marginal causal contrasts are estimated as before.

Exposure modelling, propensity scores and weighting (cont'd)

- **Inverse probability weights** (IPW) are used to estimate marginal causal contrasts (like ATE) in the whole target population. They are based on inverses of the fitted probabilities of belonging to the realized exposure group:

$$W_i = w(\mathsf{PS}_i, X_i) = \frac{\mathbf{1}_{\{X_i=1\}}}{\mathsf{PS}_i} + \frac{\mathbf{1}_{\{X_i=0\}}}{1-\mathsf{PS}_i}, \quad i = 1, \ldots, n.$$

- If the interest is on causal contrasts among the treated (like ATT), the **treated weights** are used: $W_i = 1$ for $X_i = 1$, and $W_i = \mathsf{PS}_i/(1-\mathsf{PS}_i)$ for $X_i = 0$.
- Other: **overlap weights**, **matching weights**, **entropy weights**.
- The goodness-of-fit of the exposure model needs to be assessed. For that purpose, various measures of **covariate balance** are developed.

Double robust (DR) estimators and machine learning methods

- The validity of either g-formula or PS-based weighting depends on, how accurately the outcome model or exposure model is specified.
- **Double robust** (DR) estimation of causal contrasts: Combination of g-formula and IPW. – Alternatives
  - **Augmented IPW** (AIPW); see Jonsson Funk et al. (2011),
  - **Targeted maximum likelihood estimation** (TMLE); see Schuler & Rose (2015), Luque-Fernandez et al. (2018)

  Validity of a DR estimator requires that either the exposure model or the outcome model is correctly specified.

- Algorithms developed for **supervised learning** increase flexibility in modelling both outcome and exposure (see Bi et al. 2019, Blakely et al. 2020).

Causal contrasts from censored time-to-event data

- Various methods to estimate counterfactual risks $\pi^{X=x}(t)$ and their contrasts (see Denz et al. 2023) – For instance

(a) Fit a Cox model $\lambda(t|x_i, z_i) = \lambda_0(t)\exp(\beta x_i + \gamma^T z_i)$, take estimates of coefficients and baseline cumulative hazard $\widehat{\Lambda}_0(t)$ from which:

$$\widehat{\pi}_i^{X=x}(t) = 1 - \exp\{-\widehat{\Lambda}_0(t)\exp(\widehat{\beta}x + \widehat{\gamma}^{\mathsf{T}} z_i)\}.$$

Counterfactuals $\pi^{X=x}(t)$ and contrasts are then estimated by g-formula.

(b) Get weights W_i from an exposure model, fit Cox with "intercept only" and X as a `strata()` variable and W_i:s as `weights`, and estimate $\widehat{\pi}^{X=x}(t)$ using `survfit()`, etc.

- Other: IPW Kaplan-Meier, use of pseudo-values, DR methods, ...
- Competing event settings: additional complexities in defining and analysing causal contrasts (see Rudolph et al. (2020), Young et al. (2020)).

Conclusion

- Careful specification of causal question and estimands needed.
- Selection of confounders: efficient blocking of backdoor paths.
- Basic estimation methods: outcome regression & g-formula, exposure modelling & PS-weighting, double robust estimators.
- Sufficiently flexible models desirable to reduce misspecification bias.
- Statistical inference: robust covariance matrix & delta method, bootstrapping, efficient influence curve, etc. – ignored here.
- We also limited to time-fixed exposure (binary) and confounders.
- Extensions are available to deal with polytomous exposure as well as time-varying exposure and confounding.