Statistical Considerations in Defining Estimands in Randomised Trials of Complex Disease Processes

Richard Cook

Joint Work with Alexandra Bühler and Jerry Lawless

Department of Statistics and Actuarial Science
University of Waterloo

2023 PHUSE Meeting

Roche, Mississauga, June 8, 2023

Clinical Trials for Complex Disease Processes

In trials involving *"complex outcomes"* patients are at risk of several types of events which are subject to censoring.

Examples

Palliative trials where the aim is to reduce morbidity when risk of death is appreciable Hortobagyi et al. (1996)

Trials of *respiratory disease* (e.g. COPD) where patients can have recurrent exacerbations and die Caverley et al. (2012)

Cardiovascular trials where there is risk of MI, stroke, hospitalization, CV death and non-CV death Solomon et al. (2004)

Effect of Radiosurgery Alone vs Radiosurgery With Whole Brain Radiation Therapy on Cognitive Function in Patients With 1 to 3 Brain Metastases

A Randomized Clinical Trial

Paul D. Brown, MD; Kurt Jaeckle, MD; Karla V. Ballman, PhD; Elana Farace, PhD; Jane H. Cerhan, PhD; S. Keith Anderson, MS; Xiomara W. Carrero, BS; Fred G. Barker II, MD; Richard Deming, MD; Stuart H. Burri, MD; Cynthia Ménard, MD; Caroline Chung, MD; Volker W. Stieber, MD; Bruce E. Pollock, MD; Evanthia Galanis, MD; Jan C. Buckner, MD; Anthony L. Asher, MD

IMPORTANCE Whole brain radiotherapy (WBRT) significantly improves tumor control in the brain after stereotactic radiosurgery (SRS), yet because of its association with cognitive decline, its role in the treatment of patients with brain metastases remains controversial.

OBJECTIVE To determine whether there is less cognitive deterioration at 3 months after SRS alone vs SRS plus WBRT.

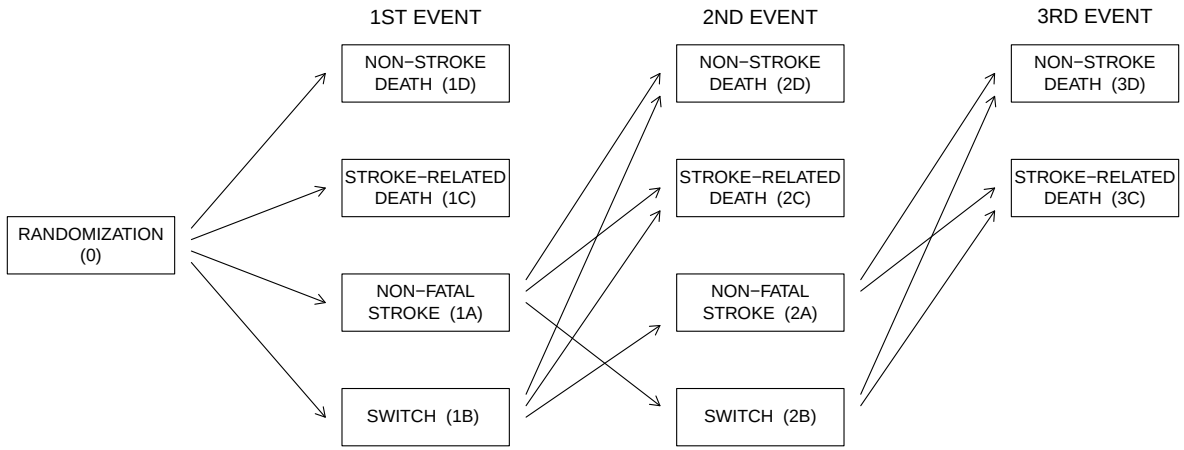
Brown PD, Jaeckle K, Ballman KV, et al. (2016). Effect of radiosurgery alone vs. radiosurgery with whole brain radiation therapy on cognitive function in patients with 1 to 3 brain metastases: a randomized clinical trial. *Journal of the American Medical Association*, **316**(4): 401–409.

Carotid endarterectomy vs. medical care in stroke prevention

Barnett et al. (1998) report on a multicenter clinical trial designed to evaluate the effect of carotid endarterectomy vs. medical therapy.

Endpoints include

- any ipsilateral stroke
- any stroke
- stroke or stroke-related death
- any stroke or death



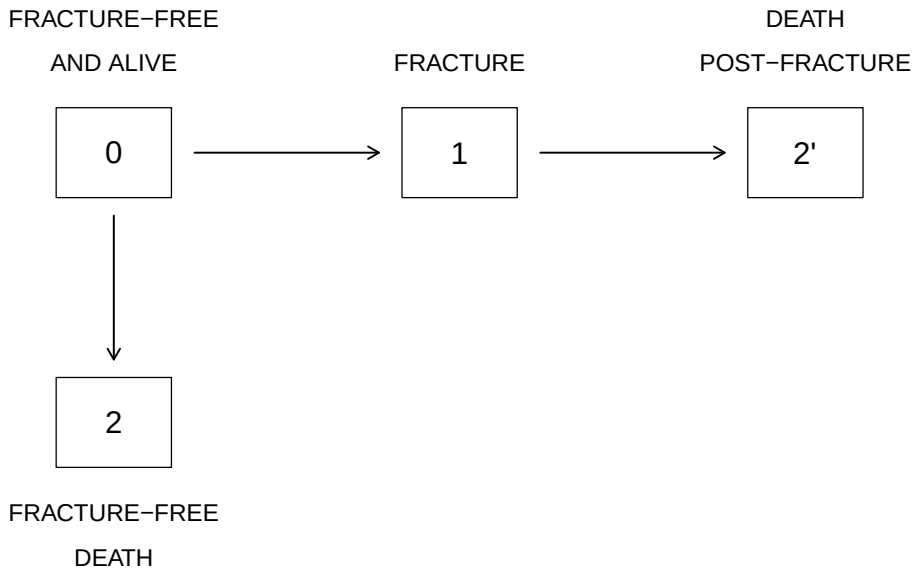
Skeletal complications and cancer metastatic to bone

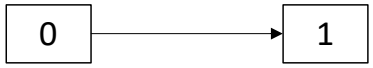
Hortobagyi et al. (1996) report on an international multicenter randomized trial involving patients with stage IV breast cancer having at least two bone lesions.

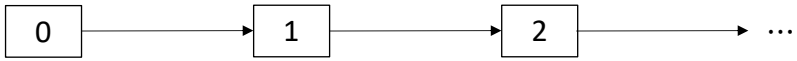
Patients were randomized within strata defined by ECOG status to receive

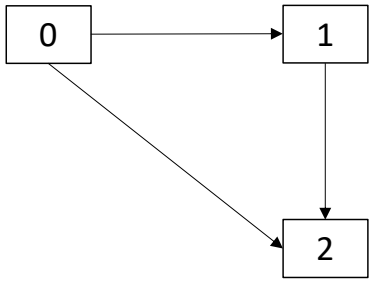
- a 90 mg infusion of pamidronate every four weeks ($n = 182$),
- a placebo control ($n = 189$).

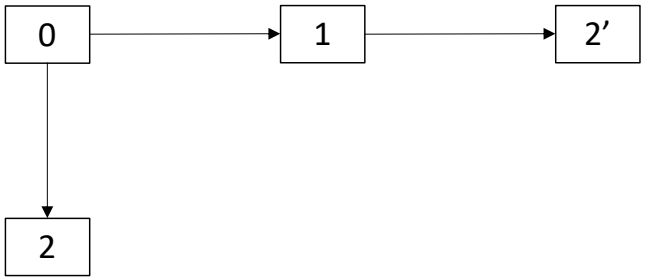
An aim is to reduce incidence of fractures

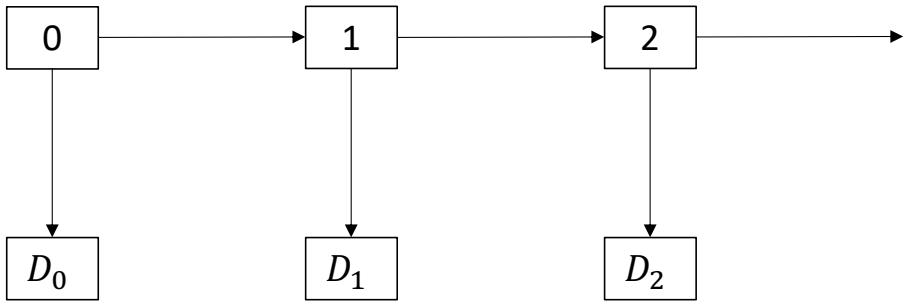












$\mathcal{S} = \{0, 1, \dots\}$ is state space with absorbing states $\mathcal{A}$

$\{Z(s), 0 < s\}$ is process

Intensity Functions

- $X = 1$ for treated and $X = 0$ for control
- $\{V(s), 0 < s\}$ is a *marker process*
- $\mathcal{H}(t) = \{Z(s), V(s), 0 < s < t, X\}$

Intensity function for a $k \longrightarrow l$ transition — Andersen et al. (1993)

$$\lim_{\Delta t \downarrow 0} \frac{P(Z(t + \Delta t^-) = l \mid Z(t^-) = k, \mathcal{H}(t))}{\Delta t} = \lambda_{kl}(t \mid \mathcal{H}(t)) \qquad k, l \in \mathcal{S}$$

Remarks

Intensity-based models play a critical role in understanding causal effects of life history processes

Aalen, Røysland and Gran (2012)

Causal effects are manifest over time so *attempts to understand causal mechanisms* should be based on intensity functions

Remarks

Intensity-based models play a critical role in understanding causal effects of life history processes

Aalen, Røysland and Gran (2012)

Causal effects are manifest over time so *attempts to understand causal mechanisms* should be based on intensity functions

Conditioning on the life history induces "collider bias"

Hernán (2010)

Aalen et al. (2015)

Clinical trials are not *primarily* designed to enhance understanding of causal mechanisms but rather to *test and estimate effects on marginal process features* and *facilitate regulatory decision making*

Bühler et al. (2023a)

Guiding Principles for Defining Estimands — Bühler et al. (2023a)

1. An estimand should target a *marginal process feature* with clear *scientific relevance*.

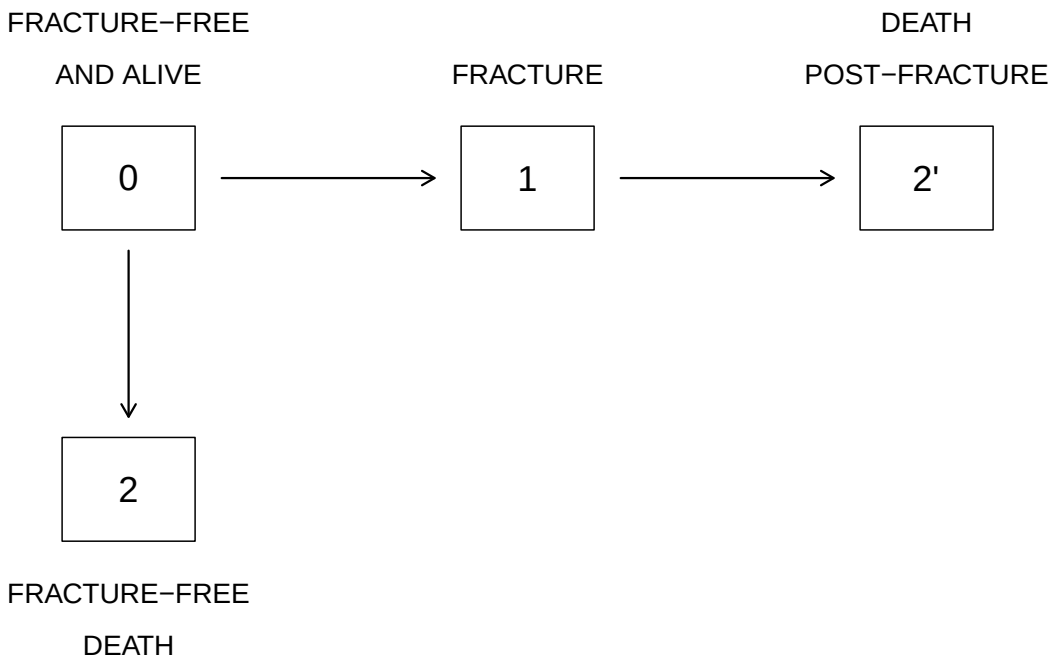
2. Features and estimands should be *interpretable* in the *"real world"* rather than in any hypothetical world.

   - Target of inference should be *an element of the observable process!*

3. Estimands should not be sensitive to *uncheckable assumptions* and consistent with *observed data and scientific background*

   - Models on which an estimand is based should be assessed using available data.

Aim is to evaluate the effect of bisphosphonates in reducing risk of fractures



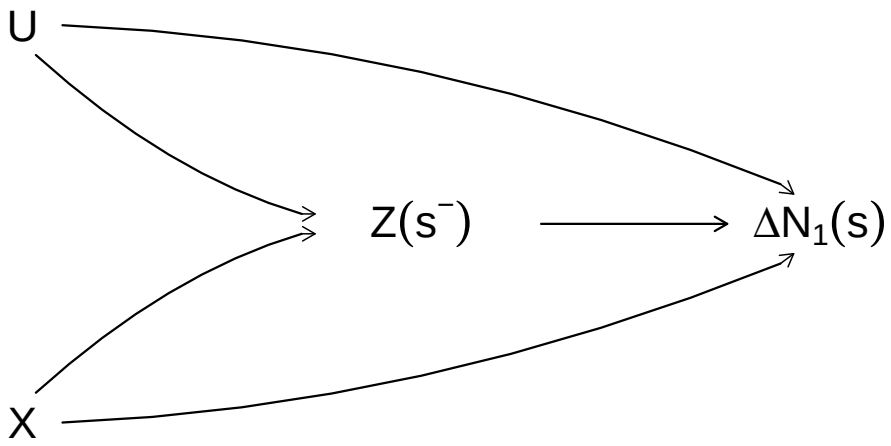
T_k is entry time to state k

$$N_k(s) = I(T_k \leq s), \;\; \Delta N_k(s) = N_k(s + \Delta s^-) - N_k(s^-)$$

Competing Risks

Cause-specific Cox regression

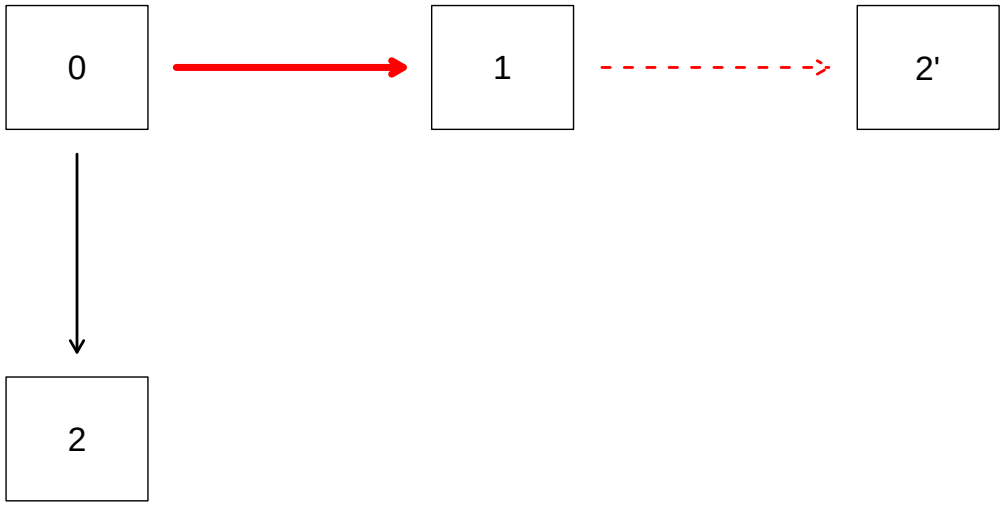
$$\lambda_1(s \mid X) = \lim_{\Delta s \downarrow 0} \frac{P(\Delta N_1(s) = 1 \mid Z(s^-) = 0,\ X)}{\Delta s} = \lambda_{10}(s)\, e^{\gamma_1 X}$$



$$U \not\perp X \mid Z(s^-) = 0$$

"Collider bias" from conditioning on $Z(s^-) = 0$

Marginal Models based on State Occupancy



Cumulative Incidence Function

Let $F_k(t) = P(T_k < t \mid Z(0) = 0) = P(N_k(t) = 1)$

General transformation models — Fine and Gray (1999)

$$g(F_1(t)) = \alpha_1(t) + x\,\beta_1 \tag{1}$$

This *fails to distinguish states* — Putter et al. (2020)

0 – alive and event-free 2 – event-free death

Limiting values under intensity-based processes

Suppose $0 - k$ intensities are Bühler et al. (2023b)

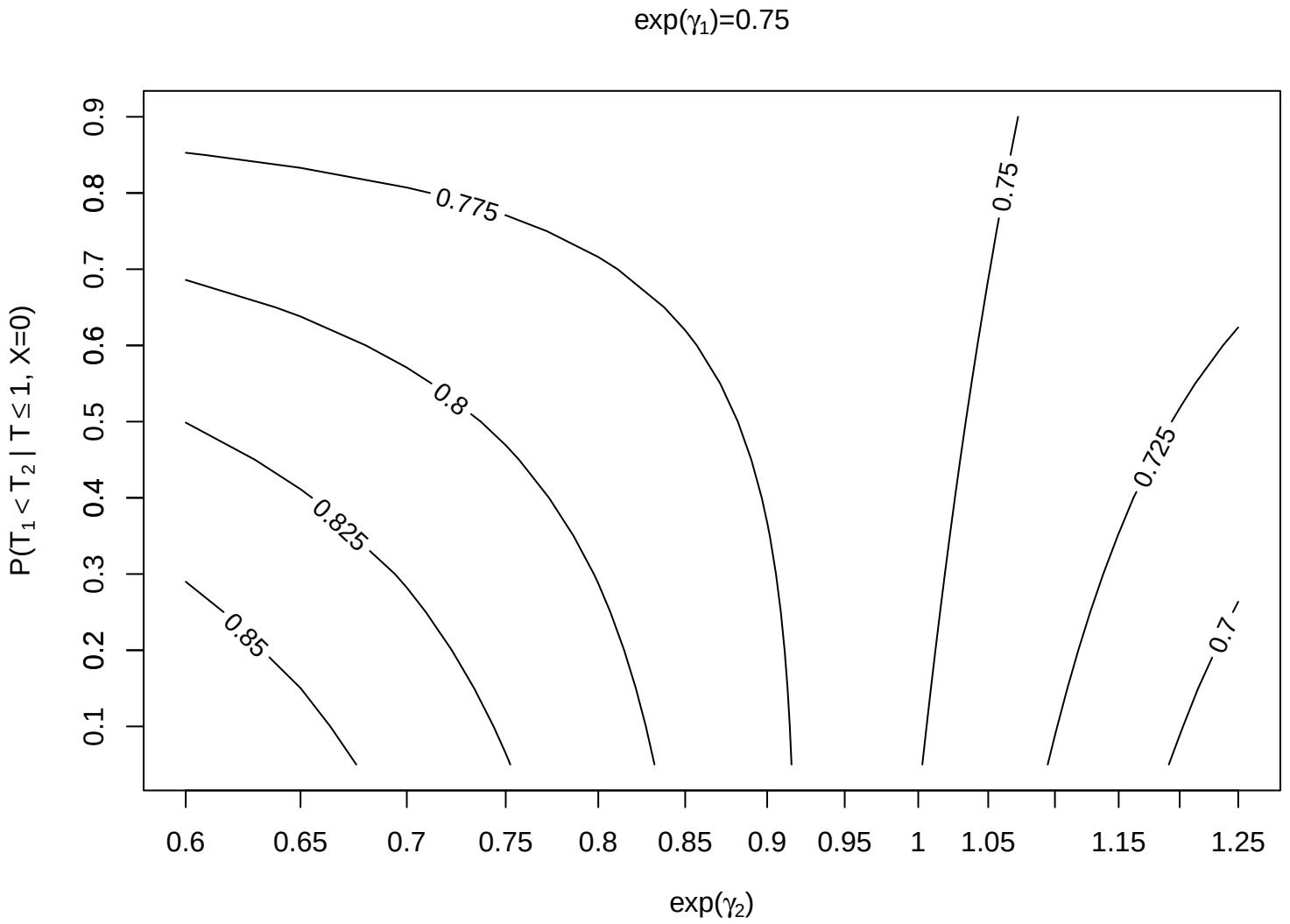
$$\lambda_{0k}(t \mid X) = \lambda_k \exp(\gamma_k X), \; k = 1, 2 \; .$$

Set $\exp(\gamma_1) = 0.75$

For given $\gamma = (\gamma_1, \gamma_2)'$, determine λ_k so that

- $P(T \leqslant 1 \mid X = 0) = 0.6$
- $P(T_1 < T_2 \mid T \leqslant 1, X = 0) \in (0, 1)$

Limiting Value of Estimator from Fine-Gray Model



Remarks

Analyses based on any *single* marginal feature are inadequate

Supplementary analyses are needed for a full understanding — Scharfstein (2019)

- based on other *marginal features*
- using intensity-based or partially conditional models

Aalen-Johansen estimates of state occupancy probabilities are robust to violations of the Markov assumption — Aalen et al (2001)

Utility-based analyses have appeal in some settings — Cook et al. (2003)

- restricted mean survival time (RMST) — McCaw et al. (2019)
- quality adjusted life years (QALY) — Gelber et al. (1989)

II. Models Incorporating Intercurrent Events Bühler et al. (2023a)

Intercurrent events are defined as

"events occurring after treatment initiation that affect either the interpretation or the existence of the measurements associated with the clinical question of interest" ICH E9(R1)

Type 1 Intercurrent Event: *precludes observation or occurrence* of event of interest

- Loss to follow-up
- Death

Type 2 Intercurrent Event: *changes the interpretation* of event of interest

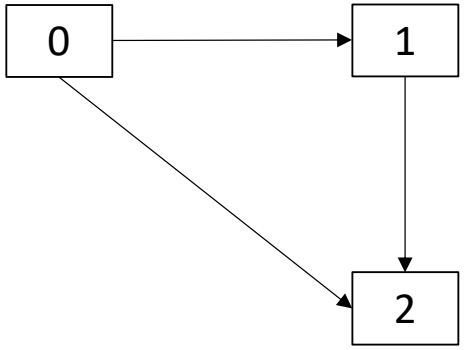
- Rescue treatment
- crossover or treatment discontinuation

Strategies for dealing with Intercurrent Events

- Treatment Policy Approach
- Composite Endpoint Approach
- While on Treatment Approach
- Hypothetical Strategies
- Principal Stratification

Rubin (2006)

We consider the meaning and limitations of these approaches through formulation of *joint multistate models for the disease and intercurrent event processes*



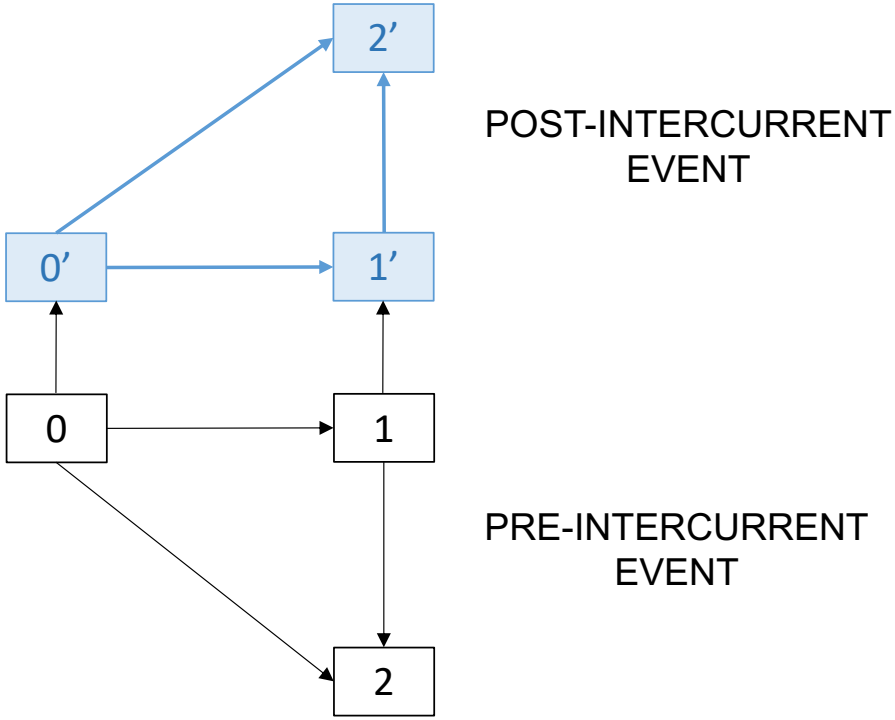
A. Illness-Death Process

$\{Z(s), 0 < s\}$

$\mathcal{H}(t) = \{Z(s), 0 < s < t, X\}$

$\lambda_{kl}(t \mid \mathcal{H}(t))$

$$q_{kl}(t \mid X) = \lim_{\Delta t \downarrow 0} \frac{P(Z(t + \Delta t^-) = l \mid Z(t^-) = k, X)}{\Delta t}$$



B. Illness-Death-IE Joint Process

$\{Z^\circ(s), 0 < s\}$

$\mathcal{H}^\circ(t) = \{Z^\circ(s), 0 < s < t, X\}$

$\lambda^\circ_{kl}(t \mid \mathcal{H}^\circ(t))$

$$q^\circ_{kl}(t \mid X) = \lim_{\Delta t \downarrow 0} \frac{P(Z^\circ(t + \Delta t^-) = l \mid Z^\circ(t^-) = k, X)}{\Delta t}$$

While on Treatment Strategy

A Type I Intercurrent Event: Censoring

Cook and Lawless (2019)

$$Y_c(t) = I(t \leq C)$$

$$Y_k(t) = I(Z(t^-) = k)$$

$$\frac{Y(t)Y_c(t)}{P(C > \min(T, t) \mid \mathcal{H}^\circ(t))} \{dN_1(t) - d\Gamma(t \mid X)\}$$

IPCW renders $C \perp \mathcal{H}(\cdot) \mid X$

While on Treatment Strategy

A Type II Intercurrent Event: Rescue treatment

$Y_E(t) = I(t \leq E)$ where E is time of *introduction of rescue treatment*.

$$\frac{Y_E(t)Y_k(t)}{P(E \geq \min(T,t) \mid \mathcal{H}^\circ(t))}\{dN_1(t) - d\Gamma(t \mid X)\}$$

IPW renders $E \perp \mathcal{H}(\cdot) \mid X$

This hypothetical scenario *has little bearing on clinical care*.

Principal Stratification

- Rubin (2006) introduces the idea of the *survivor average causal effect* (SACE)
- *Principal strata are latent* so identifiability and estimation of the SACE requires additional modeling assumptions Egleston et al. (2007)

Questions

Is the *interpretation of the SACE aligned with scientific aims?*

> *If we cannot translate our causal question into a target trial, the question is not well-defined* Hernán (ISCB, 2020)

Should we *re-evaluate the specification of study objectives in presence of mortality?*

Remarks on Counterfactuals

> *"Many counterfactual analyses are based, explicitly or implicitly, on an attitude that I term fatalism. This considers the various potential responses $Y_i(u)$, when treatment i is applied to unit u, as predetermined attributes of unit u, waiting only to be uncovered by suitable experimentation."*
>
> Dawid (JASA, 2000a)

In their discussion of his paper, Robins and Greenland (2000) and Rubin (2006) *reaffirm the need for counterfactuals*, however in Dawid's rejoinder . . .,

> *"... my own attitude is that, as there is no difficulty in determining an empirically meaningful probability structure for the observable (Y, Z) given treatment – even though this is defined over an unusual space, where Z automatically takes the value "undefined" whenever $Y = 0$ ... The real problem is how to define a sensible utility measure on this outcome space."*
>
> Dawid (2000b)

Remarks

- Inverse probability of censoring weighting deals with unrepresentativeness due to selective attrition

- When censoring at some intercurrent events, use of IPCW for the *while on treatment strategy* creates a pseudo-sample not representative of any real-world setting

- Intention-to-treat and the *treatment policy strategy* is a preferred way of integrating the IE into the response process. Scharfstein (2019)

  – *generalizability and transportability of study findings* are sensitive to patient mix and standard of care

  – this a key issue in *multinational clinical trials*

References

Hortobagyi, G. N., Theriault, R. L., Porter, L., Blayney, D., Lipton, A., Sinoff, C., Wheeler, H., Simeone, J. F., Seaman, J., and Knight, R. D. (1996). Efficacy of pamidronate in reducing skeletal complications in patients with breast cancer and lytic bone metastases. *New England Journal of Medicine*, **335**(34): 1785–1791.

Andersen, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*. Springer-Verlag, New York, NY.

Cook, R. J. and Lawless, J. F. (2018). Multistate Models for the Analysis of Life History Data. *CRC Press*.

Hernán, M. A. (2010). The hazards of hazard ratios. *Epidemiology*, **21**(1): 13–15.

Bühler, A., Cook, R. J., and Lawless, J. F. (2023a). Multistate models as a framework for estimand specification in clinical trials of complex processes. *Statistics in Medicine*.

Bühler, A., Cook, R. J., and Lawless, J. F. (2023b). Generalized Linear Models for Marginal Features in Complex Life History Processes. *Manuscript*.

Cook, R. J., Lawless, J. F., and Lee, K.-A. (2003). Cumulative processes related to event histories. *SORT*, **27**(1): 13–30.

Aalen O. O., Røysland K, Gran JM (2012a). Causality, mediation and time: a dynamic viewpoint. *Journal of the Royal Statistical Society: Series A*, **175**(4): 831–861.

Hortobagyi GN, Theriault RL, Porter L et al. (1996). Efficacy of Pamidronate in reducing skeletal complications in patients with breast cancer and lytic bone metastases. *New England Journal of Medicine*, **335**: 1875 – 1792.

Putter H, Schumacher M, van Houwelingen HC (2020). On the relation between the cause-specific hazard and the subdistribution rate for competing risks data: The Fine-Gray model revisited. *Biometrical Journal*, **62**: 790–807.

McCaw ZR, Yin G and Wei LJ (2019). Using the restricted mean survival time difference as an alternative to the hazard ratio for analyzing clinical cardiovascular studies. *Circulation*, **140** (17), 1366–1368.

Gelber RD, Gelman RS and Goldhirsch A (1989). A quality-of-life-oriented endpoint for comparing therapies. *Biometrics* **45**: 781–795.

Glasziou PP, Simes RJ and Gelber RD (1990). Quality adjusted survival analysis. *Statistics in Medicine* **9**: 1259–1276.

Cook RJ, Lawless JF and Lee K-A (2003). Cumulative processes related to event histories. *SORT*, **27**(1): 13–30.

Andersen, Per Kragh, and Richard D. Gill. (1982). Cox's regression model for counting processes: a large sample study. *The annals of statistics*, 1100-1120.

Ich, C. H. M. P. (2019). E9 (R1) Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials. In *Proceedings of the international conference on harmonisation of technical requirements for registration of pharmaceuticals for human use*.

Lin, D. Y., Wei, L. J., Yang, I., and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **62**(4): 711-730.

Fine, J. P., and Gray, R. J. (1999). A proportional hazards model for the subdistribution of a competing risk. *Journal of the American statistical association*, **94**(446): 496-509.

Scheike, T. H., Zhang, M. J., and Gerds, T. A. (2008). Predicting cumulative incidence probability by direct binomial regression. *Biometrika*, **95**(1): 205-220.

Aalen, O. O., Borgan, Ø., and Fekjær, H. (2001). Covariate adjustment of event histories estimated from Markov chains: the additive approach. *Biometrics*, **57**(4): 993-1001.

Scharfstein, D. O. (2019). A constructive critique of the draft ICH E9 addendum. *Clinical Trials*, **16**(4): 375-380.

Lawless, J. F., and Cook, R. J. (2019). A new perspective on loss to follow-up in failure time and life history studies. *Statistics in Medicine*, **38**(23): 4583-4610.

Lawless, J.F. (1995). The analysis of recurrent events for multiple subjects. *Applied Statistics*, **44**: 487-498.