Missing data and composite endpoints

Efficient estimation of the distribution of time to composite endpoint when one of the endpoints is incompletely observed

Rhian Daniel[1] and Butch Tsiatis[2]

[1]Centre for Statistical Methodology and Medical Statistics Department, London School of Hygiene and Tropical Medicine
[2]Department of Statistics, North Carolina State University

CSM Seminar, 25th January 2013

LONDON SCHOOL of HYGIENE &TROPICAL MEDICINE

This talk is based on:

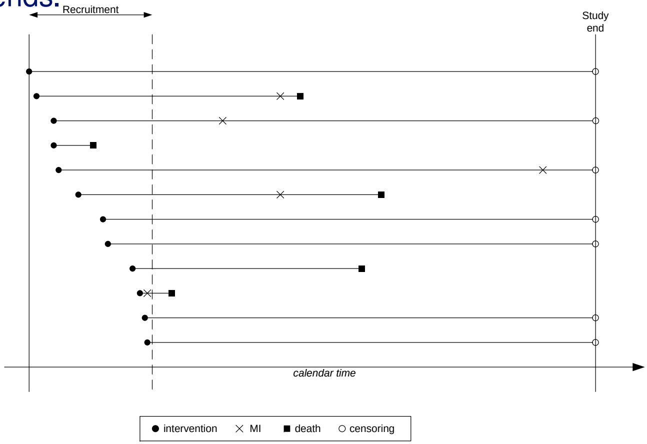
- Daniel RM, Tsiatis AA
  Efficient estimation of the distribution of time to composite endpoint when some endpoints are only partially observed.
  Lifetime Data Analysis, under revision.

Work carried out while I visited Prof Butch Tsiatis in Raleigh NC, January–April 2012.

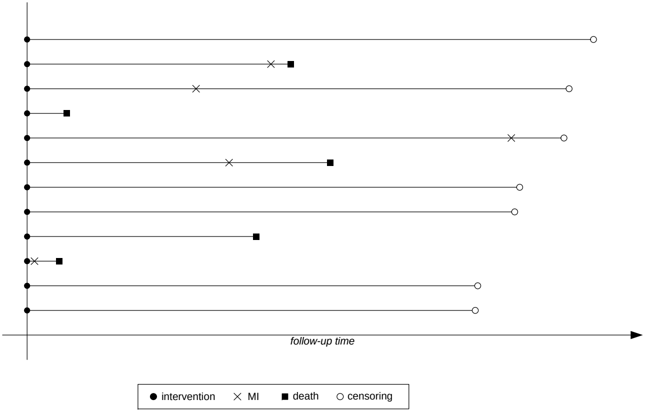
Funded by MRC Career Development Award in Biostatistics G1002283.

Setting (1)

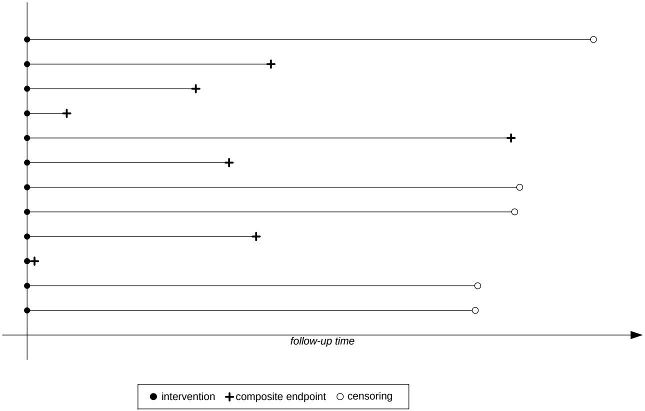
Consider a study in which patients are recruited (and given an intervention) and then followed up until MI or death, or until the study ends.



Typically we would analyse such a study on the follow-up timescale.

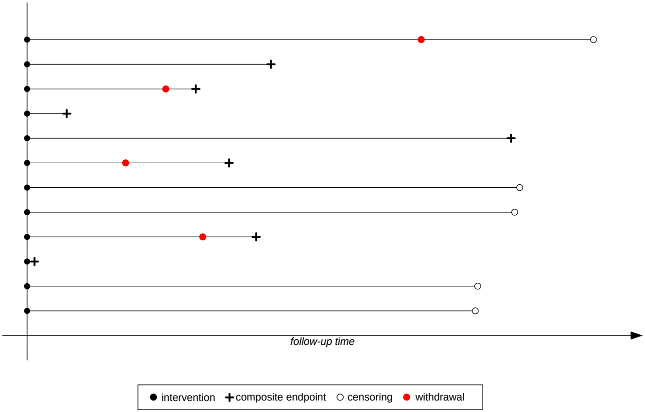


And often we look only at time to **composite endpoint**: MI *or* death.

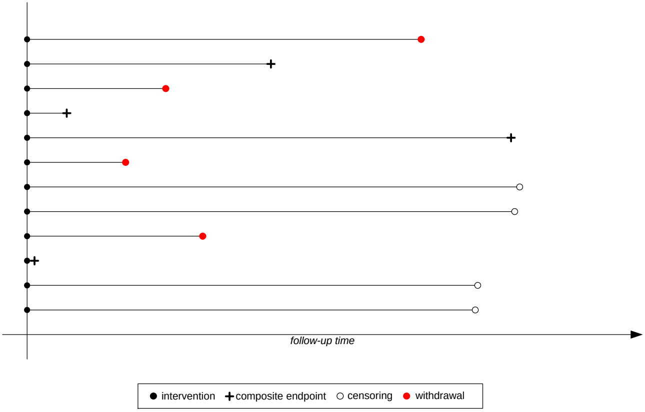


Setting (4)

Suppose some subjects withdraw before the end of the study.

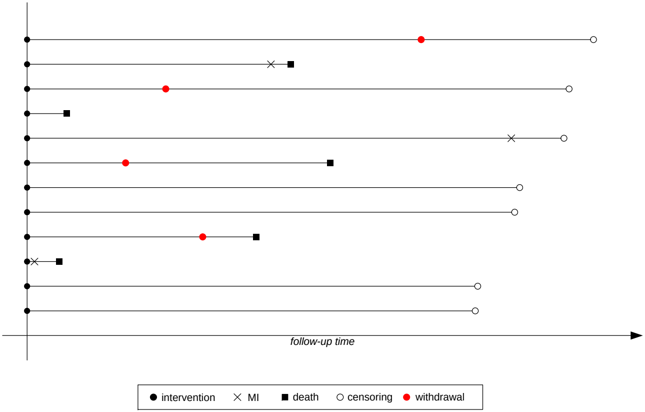


Often we know nothing about what happens to these subjects subsequently.



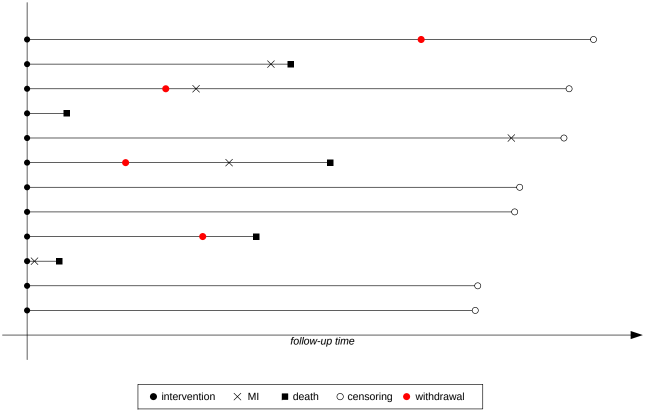
Setting (6)

But in our setting, we DO have data on whether or not death occurred before the end of the study, even for those who withdrew — from a national death index.

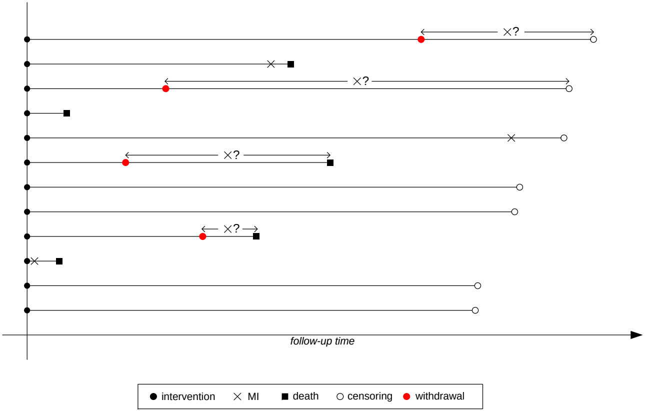


Setting (7)

But for those who in fact had an MI, we don't know this. And for those who did not, we don't know this either.

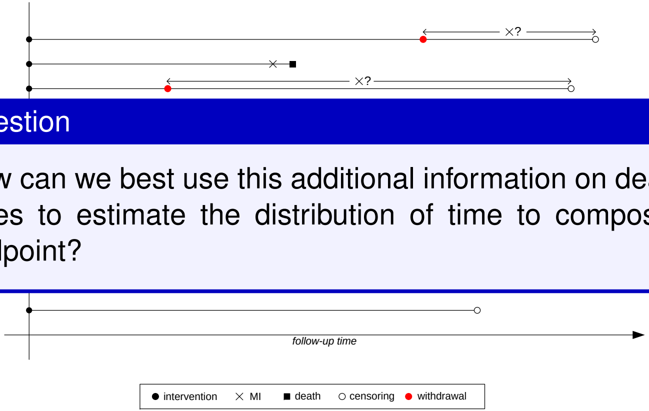


So here is a depiction of our data:



Setting (8)

So here is a depiction of our data:



Question

How can we best use this additional information on death times to estimate the distribution of time to composite endpoint?

So here is a depiction of our data:

Suggestions

Setting (8)

So here is a depiction of our data:

Suggestions

– Ignore the extra information

Setting (8)

So here is a depiction of our data:

Suggestions

– Ignore the extra information *–inefficient*

So here is a depiction of our data:

Suggestions

– Ignore the extra information *–inefficient*

– Treat the extra times-to-death as times-to-composite-endpoint

Setting (8)

So here is a depiction of our data:

Suggestions

– Ignore the extra information *–inefficient*

– Treat the extra times-to-death as times-to-composite-endpoint *–biased since wrongly assumes no MI for dropouts*

Setting (8)

So here is a depiction of our data:

Suggestions

– Ignore the extra information *–inefficient*
– Treat the extra times-to-death as times-to-composite-endpoint *–biased since wrongly assumes no MI for dropouts*
– Impute the missing times-to-MI using all the available information

So here is a depiction of our data:

Suggestions

– Ignore the extra information *–inefficient*

– Treat the extra times-to-death as times-to-composite-endpoint *–biased since wrongly assumes no MI for dropouts*

– Impute the missing times-to-MI using all the available information *–prone to bias from misspecifying the imputation model*

Setting (8)

So here is a depiction of our data:

Suggestions

– Ignore the extra information *–inefficient*
– Treat the extra times-to-death as times-to-composite-endpoint *–biased since wrongly assumes no MI for dropouts*
– Impute the missing times-to-MI using all the available information *–prone to bias from misspecifying the imputation model*
– Ideally want a principled, efficient alternative, not too dependent on extra parametric models

Setting (8)

So here is a depiction of our data:

Suggestions

– Ignore the extra information *–inefficient*

– Treat the extra times-to-death as times-to-composite-endpoint *–biased since wrongly assumes no MI for dropouts*

– Impute the missing times-to-MI using all the available information *–prone to bias from misspecifying the imputation model*

– Ideally want a principled, efficient alternative, not too dependent on extra parametric models *–we will achieve this using the semiparametric theory of augmented inverse probability weighted estimating equations*

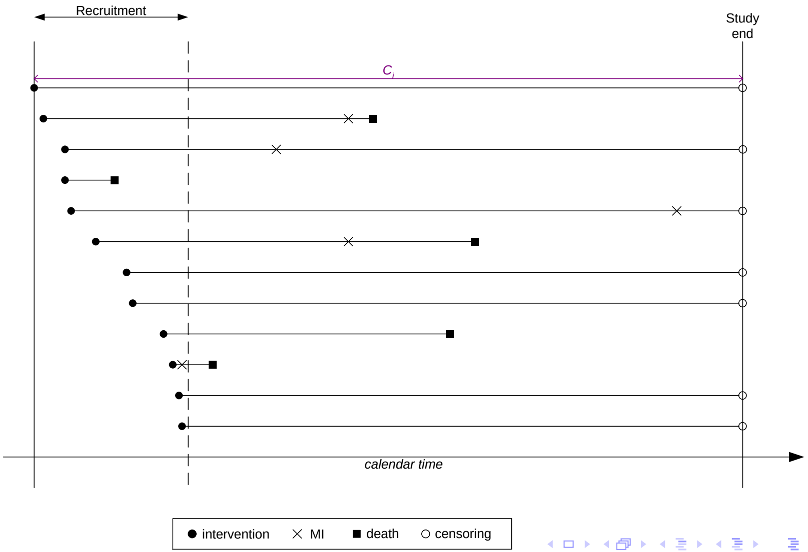
Outline

- Setting
- Notation
- Counting processes
- Semiparametric theory
- Estimators for the distribution of time to composite endpoint
  full data $\rightarrow$ complete cases $\rightarrow$ inverse probability weighted CC $\rightarrow$ augmented IPWCC
- Double robustness and semiparametric efficiency
- Variance estimation
- Simulation study
- Summary and further issues

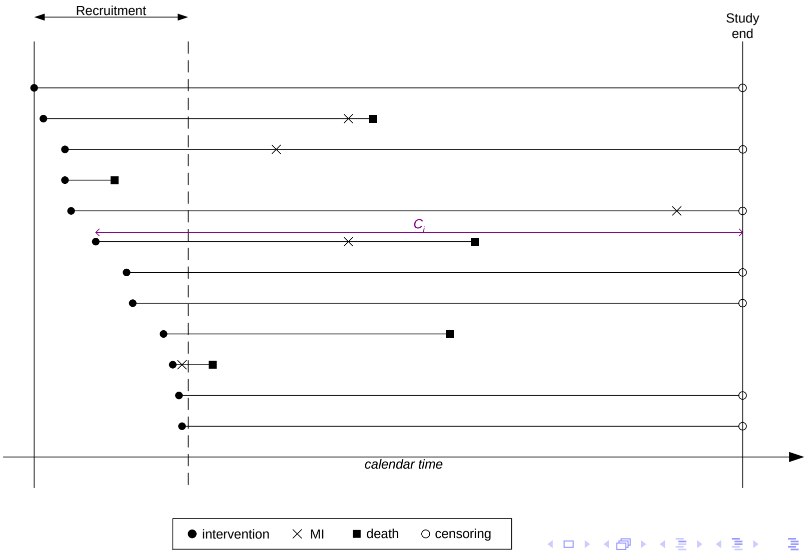
Outline

- Setting
- Notation
- Counting processes
- Semiparametric theory
- Estimators for the distribution of time to composite endpoint
  full data $\rightarrow$ complete cases $\rightarrow$ inverse probability weighted CC $\rightarrow$ augmented IPWCC
- Double robustness and semiparametric efficiency
- Variance estimation
- Simulation study
- Summary and further issues

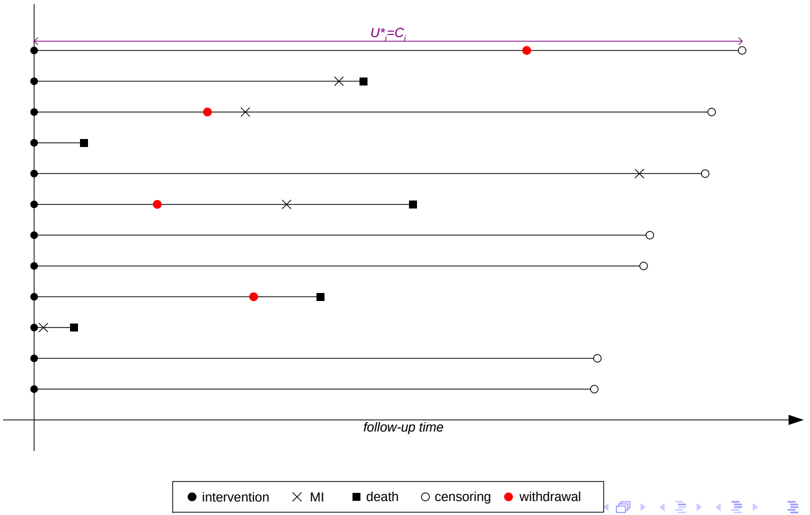
C_i: censoring time



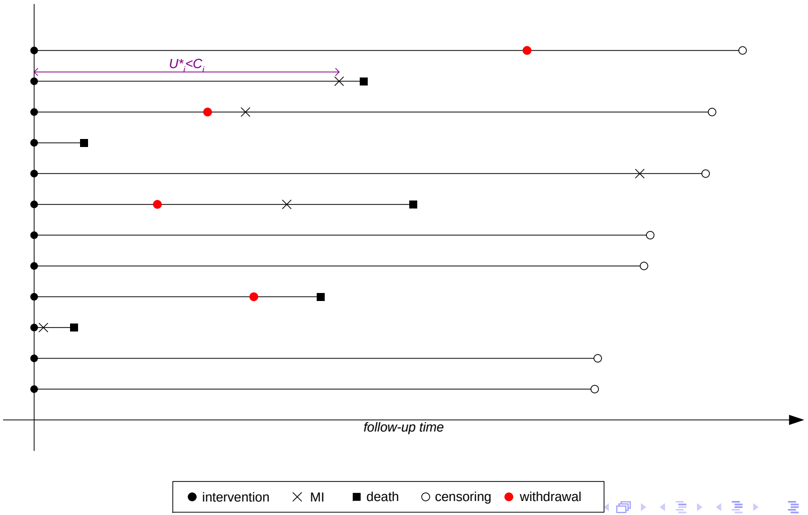
C_i: censoring time



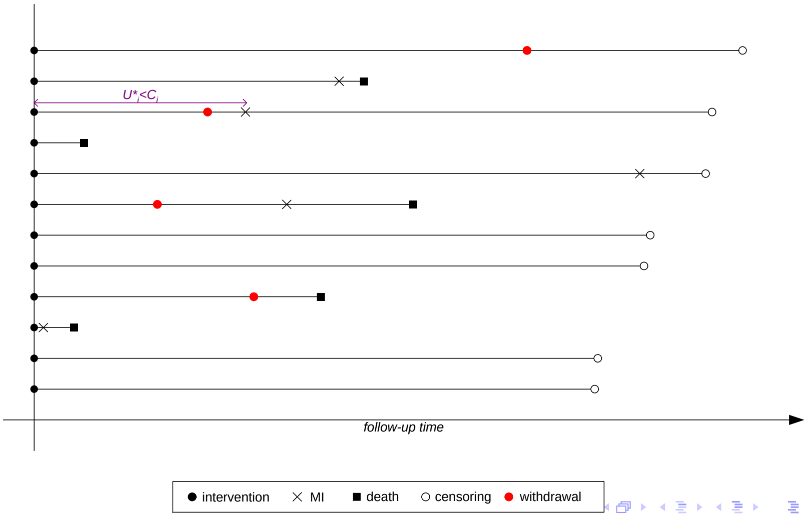
U_i^*: time to first event or censoring



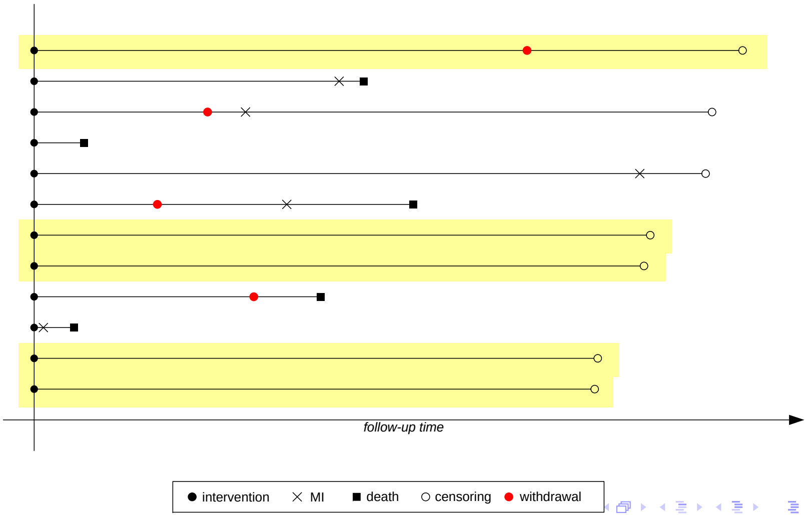
U_i^*: time to first event or censoring



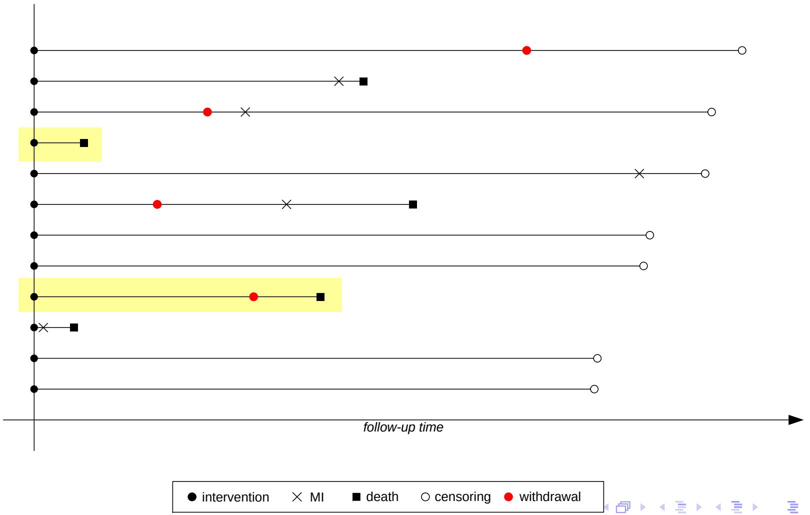
U_i^*: time to first event or censoring



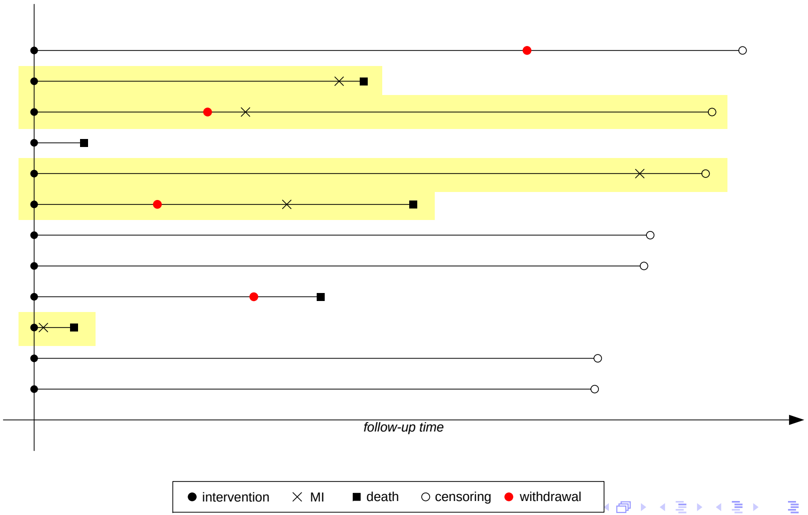
$\Delta_i^* = 0$: U_i^* is a censoring time



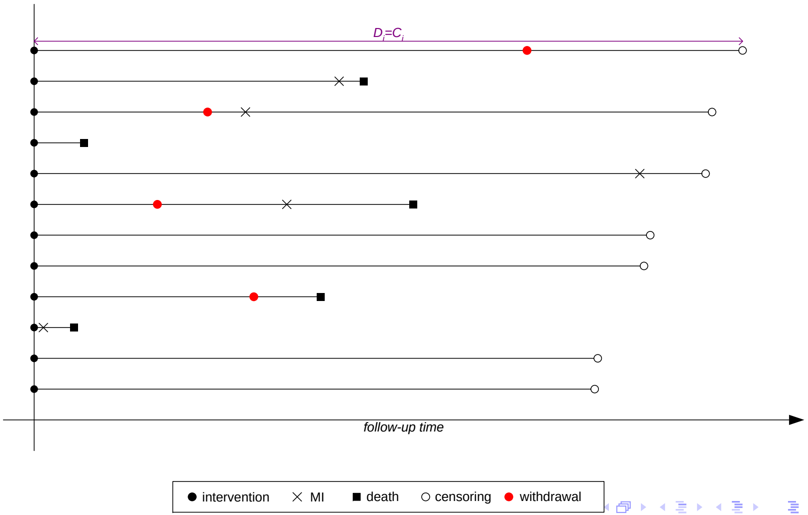
$\Delta_i^* = 1$: U_i^* is a death time



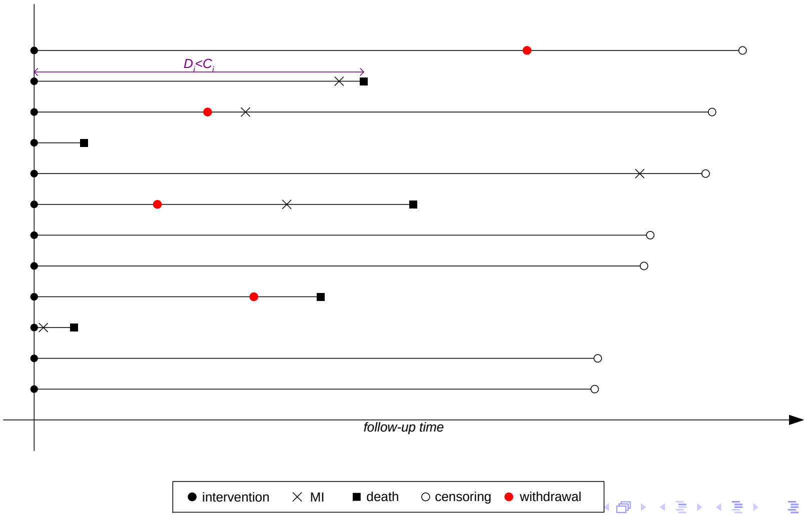
$\Delta_i^* = 2$: U_i^* is a time to MI



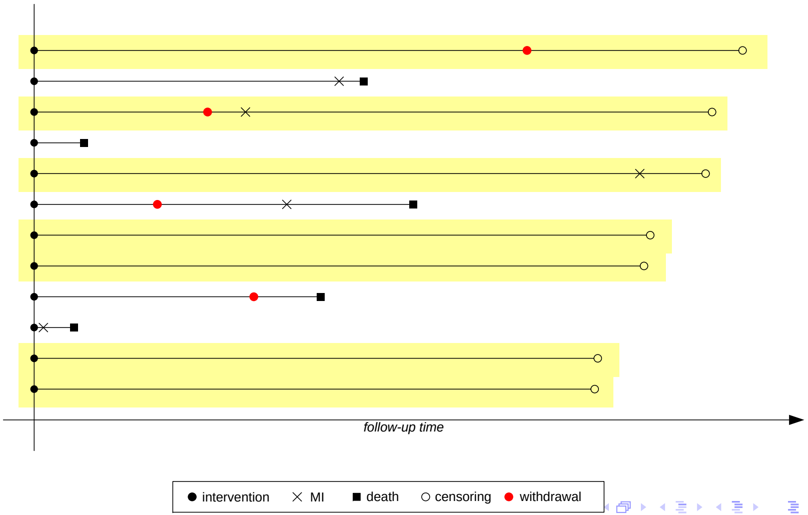
D_i: time to death or censoring



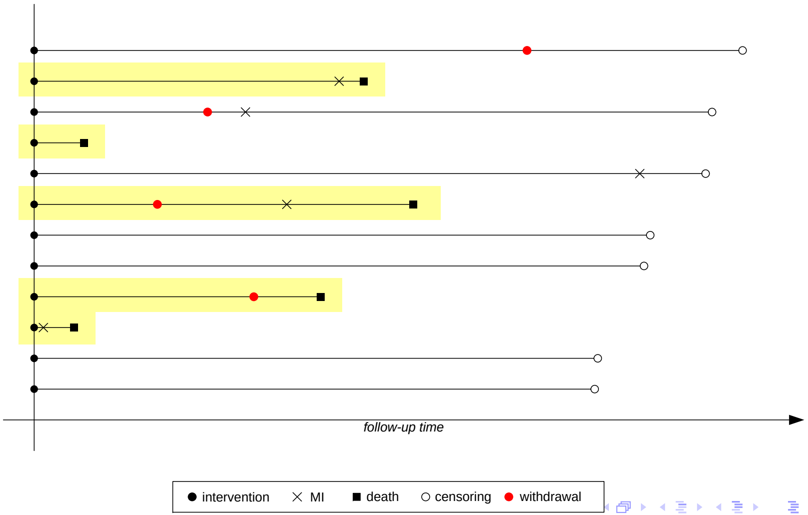
D_i: time to death or censoring



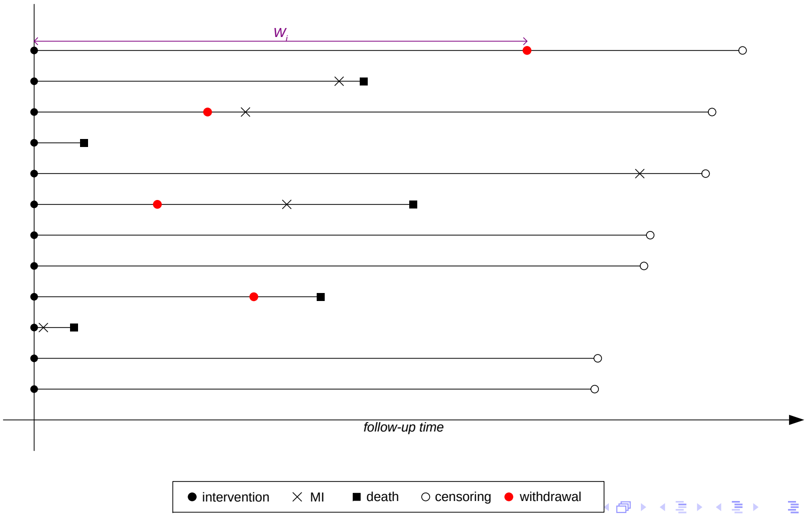
$\Gamma_i = 0$: D_i is a censoring time



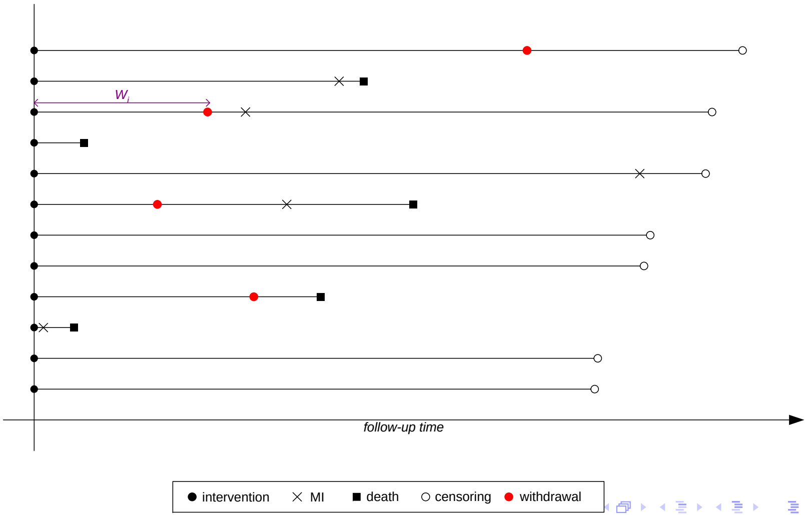
$\Gamma_i = 1$: D_i is a death time



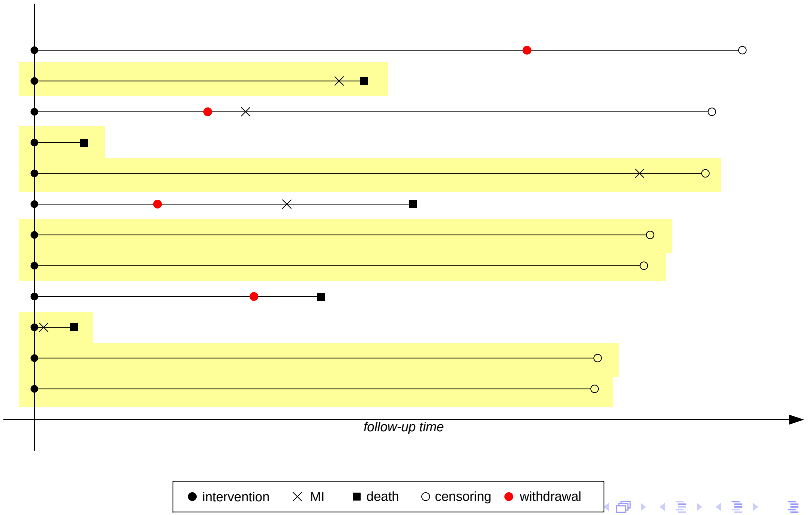
W_i: time to withdrawal



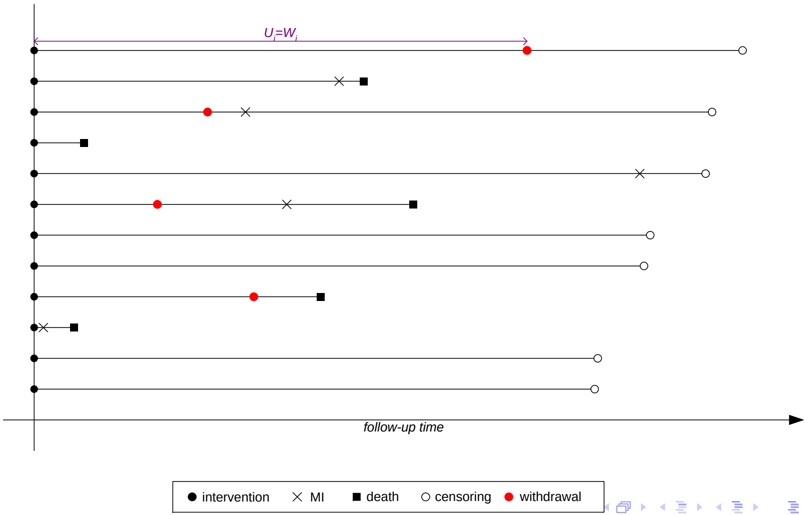
W_i: time to withdrawal



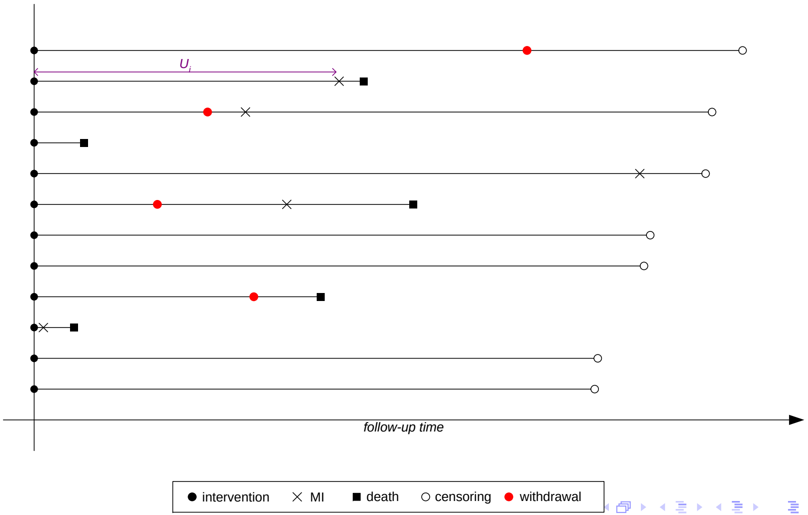
$W_i = \infty$: for those who don't withdraw



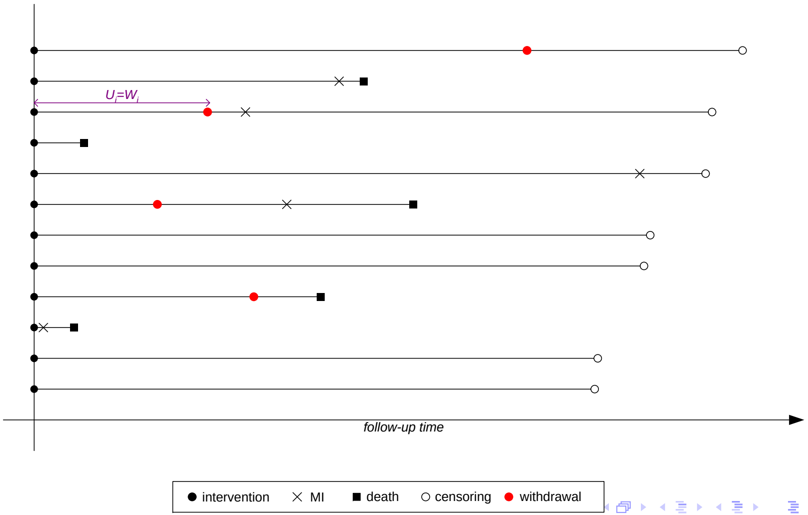
U_i: time to first event or censoring or withdrawal



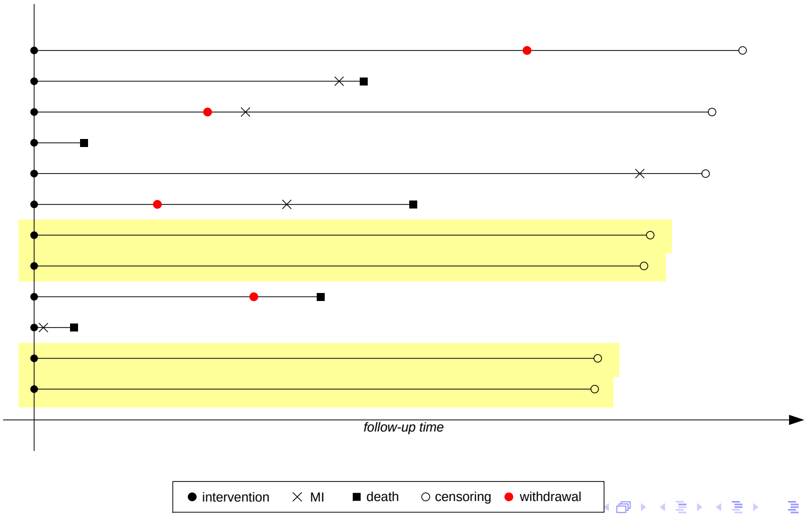
U_i: time to first event or censoring or withdrawal



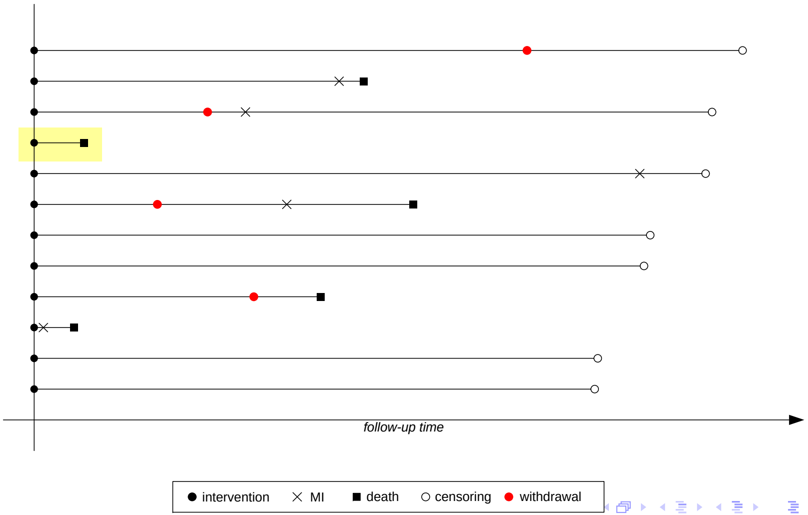
U_i: time to first event or censoring or withdrawal



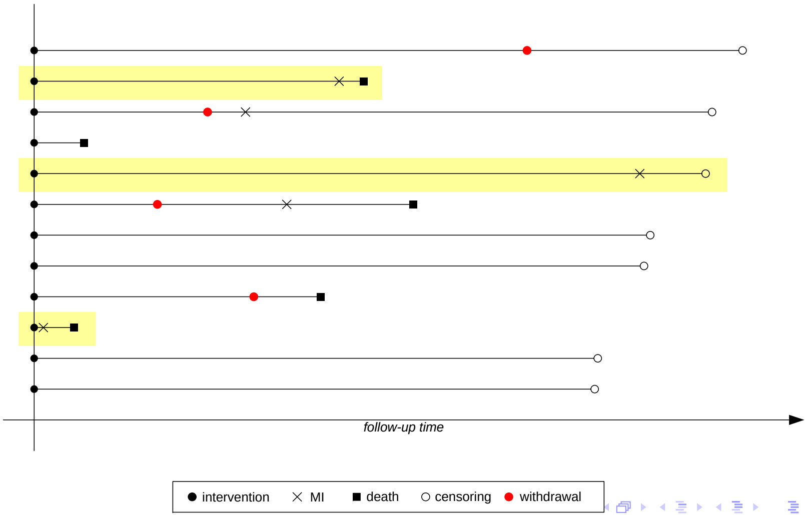
$\Delta_i = 0$: U_i is a censoring time



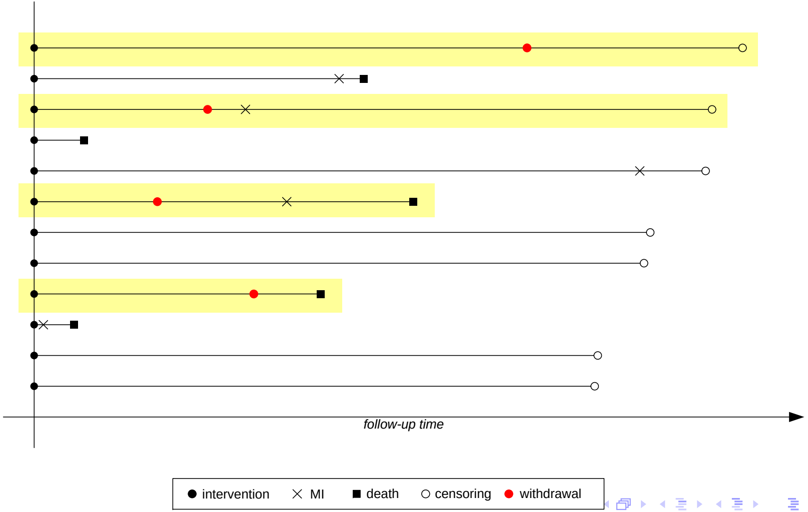
$\Delta_i = 1$: U_i is a death time



$\Delta_i = 2$: U_i is a time to MI



$\Delta_i = -1$: U_i is a withdrawal time



Notation summary (1)

- C_i: time to censoring
- U_i^*: time to MI or death or censoring
- Δ_i^*: event 'indicator' for U_i^*,

$$\Delta_i^* = \begin{cases} 0 \text{ if } U_i^* = C_i \\ 1 \text{ if } U_i^* \text{ is time to death} \\ 2 \text{ if } U_i^* \text{ is time to MI} \end{cases}$$

- D_i: time to death or censoring
- Γ_i: event indicator for D_i,

$$\Gamma_i = \begin{cases} 0 \text{ if } D_i = C_i \\ 1 \text{ if } D_i \text{ is time to death} \end{cases}$$

- $W_i \in (0, U_i^*) \cup \{\infty\}$: time to withdrawal
- U_i: time to MI or death or withdrawal or censoring
- Δ_i: event 'indicator' for U_i,

$$\Delta_i = \begin{cases} -1 \text{ if } U_i = W_i \\ 0 \text{ if } U_i = C_i \\ 1 \text{ if } U_i \text{ is time to death} \\ 2 \text{ if } U_i \text{ is time to MI} \end{cases}$$

Notation summary (1)

Note: starred quantities are not fully-observed

- C_i: time to censoring
- U_i^*: time to MI or death or censoring
- Δ_i^*: event 'indicator' for U_i^*,

$$\Delta_i^* = \begin{cases} 0 \text{ if } U_i^* = C_i \\ 1 \text{ if } U_i^* \text{ is time to death} \\ 2 \text{ if } U_i^* \text{ is time to MI} \end{cases}$$

- D_i: time to death or censoring
- Γ_i: event indicator for D_i,

$$\Gamma_i = \begin{cases} 0 \text{ if } D_i = C_i \\ 1 \text{ if } D_i \text{ is time to death} \end{cases}$$

- $W_i \in (0, U_i^*) \cup \{\infty\}$: time to withdrawal
- U_i: time to MI or death or withdrawal or censoring
- Δ_i: event 'indicator' for U_i,

$$\Delta_i = \begin{cases} -1 \text{ if } U_i = W_i \\ 0 \text{ if } U_i = C_i \\ 1 \text{ if } U_i \text{ is time to death} \\ 2 \text{ if } U_i \text{ is time to MI} \end{cases}$$

Notation summary (2)

— In addition, let $\bar{\mathbf{X}}_i(t)$ be a vector of time-updated covariates for subject i as collected up to time t.

Notation summary (2)

— In addition, let $\bar{\mathbf{X}}_i(t)$ be a vector of time-updated covariates for subject i as collected up to time t.
— $\bar{\mathbf{X}}_i(0)$ are therefore the baseline covariates.

Notation summary (2)

— In addition, let $\bar{\mathbf{X}}_i(t)$ be a vector of time-updated covariates for subject i as collected up to time t.
— $\bar{\mathbf{X}}_i(0)$ are therefore the baseline covariates.
— The **full data** for subject i (what we would see if hypothetically there were no withdrawals) are:

$$\mathcal{F}_i = \left\{ C_i, U_i^*, \Delta_i^*, \bar{\mathbf{X}}_i\left(U_i^*\right), D_i, \Gamma_i \right\}.$$

Notation summary (2)

— In addition, let $\bar{\mathbf{X}}_i(t)$ be a vector of time-updated covariates for subject i as collected up to time t.
— $\bar{\mathbf{X}}_i(0)$ are therefore the baseline covariates.
— The **full data** for subject i (what we would see if hypothetically there were no withdrawals) are:

$$\mathcal{F}_i = \left\{ C_i, U_i^*, \Delta_i^*, \bar{\mathbf{X}}_i(U_i^*), D_i, \Gamma_i \right\}.$$

— The **observed data** for subject i are:

$$\mathcal{O}_i = \left\{ C_i, U_i, \Delta_i, \bar{\mathbf{X}}_i(U_i), D_i, \Gamma_i \right\}.$$

Notation summary (2)

— In addition, let $\bar{\mathbf{X}}_i(t)$ be a vector of time-updated covariates for subject i as collected up to time t.
— $\bar{\mathbf{X}}_i(0)$ are therefore the baseline covariates.
— The **full data** for subject i (what we would see if hypothetically there were no withdrawals) are:

$$\mathcal{F}_i = \left\{C_i, U_i^*, \Delta_i^*, \bar{\mathbf{X}}_i(U_i^*), D_i, \Gamma_i\right\}.$$

— The **observed data** for subject i are:

$$\mathcal{O}_i = \left\{C_i, U_i, \Delta_i, \bar{\mathbf{X}}_i(U_i), D_i, \Gamma_i\right\}.$$

— The **level-r coarsened data** for subject i are:

$$G_r(\mathcal{F}_i) =$$
$$\left\{C_i, I(U_i^* < r), I(U_i^* < r)U_i^*, I(U_i^* < r)\Delta_i^*, \bar{\mathbf{X}}_i\{\min(r, U_i^*)\}, D_i, \Gamma_i\right\}.$$

Notation summary (2)

— In addition, let $\bar{\mathbf{X}}_i(t)$ be a vector of time-updated covariates for subject i as collected up to time t.
— $\bar{\mathbf{X}}_i(0)$ are therefore the baseline covariates.
— The **full data** for subject i (what we would see if hypothetically there were no withdrawals) are:

$$\mathcal{F}_i = \left\{ C_i, U_i^*, \Delta_i^*, \bar{\mathbf{X}}_i\left(U_i^*\right), D_i, \Gamma_i \right\}.$$

— The **observed data** for subject i are:

$$\mathcal{O}_i = \left\{ C_i, U_i, \Delta_i, \bar{\mathbf{X}}_i\left(U_i\right), D_i, \Gamma_i \right\}.$$

— The **level-r coarsened data** for subject i are:

$$G_r\left(\mathcal{F}_i\right) =$$
$$\left\{ C_i, I\left(U_i^* < r\right), I\left(U_i^* < r\right) U_i^*, I\left(U_i^* < r\right) \Delta_i^*, \bar{\mathbf{X}}_i\left\{\min\left(r, U_i^*\right)\right\}, D_i, \Gamma_i \right\}.$$

(think of this as the information we would see on subject i were s/he to withdraw at time r)

Notation summary (3)

— Writing T_i for the uncensored time to composite endpoint, we are interested in estimating the usual related survival estimands:

– the survivor function

$$S(t) = Pr(T_i > t),$$

– the hazard function

$$\lambda(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} Pr(t \leq T_i < t + \Delta t \,|\, T_i \geq t)$$

– and the cumulative hazard function:

$$\Lambda(t) = \int_0^t \lambda(u)\, du,$$

with

$$S(t) = \exp\{-\Lambda(t)\} = \exp\left\{-\int_0^t \lambda(u)\, du\right\}.$$

Outline

- Setting
- Notation
- Counting processes
- Semiparametric theory
- Estimators for the distribution of time to composite endpoint
  full data → complete cases → inverse probability weighted CC → augmented IPWCC
- Double robustness and semiparametric efficiency
- Variance estimation
- Simulation study
- Summary and further issues

Counting processes notation

— We write

$$N_i^*(t) = I(U_i^* \leq t, \Delta_i^* \in \{1,2\})$$

for the counting process associated with the composite endpoint in the absence of withdrawal.

Counting processes notation

— We write

$$N_i^*(t) = I\left(U_i^* \leq t, \Delta_i^* \in \{1,2\}\right)$$

for the counting process associated with the composite endpoint in the absence of withdrawal.

— With withdrawal:

$$N_i(t) = I\left(U_i \leq t, \Delta_i \in \{1,2\}\right).$$

Counting processes notation

— We write

$$N_i^*(t) = I(U_i^* \leq t, \Delta_i^* \in \{1, 2\})$$

for the counting process associated with the composite endpoint in the absence of withdrawal.

— With withdrawal:

$$N_i(t) = I(U_i \leq t, \Delta_i \in \{1, 2\}).$$

— The risk set indicator at time t, in the absence of withdrawal:

$$Y_i^*(t) = I(U_i^* \geq t).$$

Counting processes notation

— We write

$$N_i^*(t) = I\left(U_i^* \leq t, \Delta_i^* \in \{1,2\}\right)$$

for the counting process associated with the composite endpoint in the absence of withdrawal.
— With withdrawal:

$$N_i(t) = I\left(U_i \leq t, \Delta_i \in \{1,2\}\right).$$

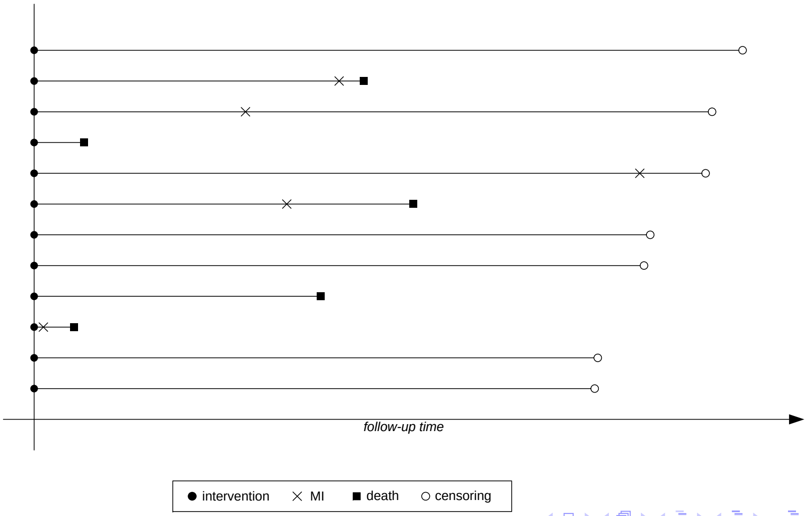
— The risk set indicator at time t, in the absence of withdrawal:

$$Y_i^*(t) = I\left(U_i^* \geq t\right).$$

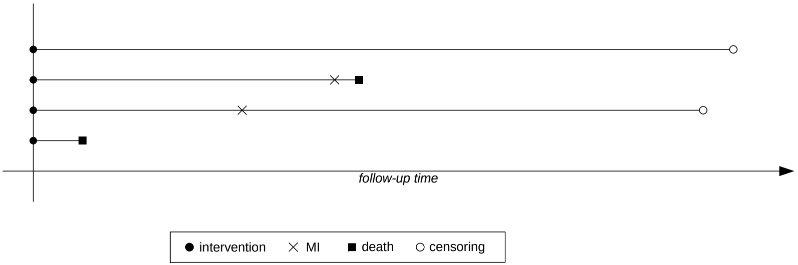
— With withdrawal:

$$Y_i(t) = I\left(U_i \geq t\right).$$

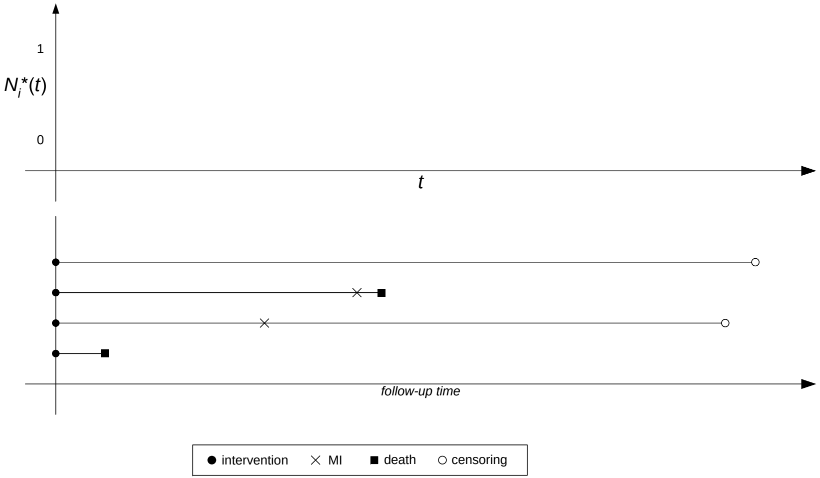
$N_i^*(t)$: counting process for the composite endpoint
(in the absence of withdrawal)



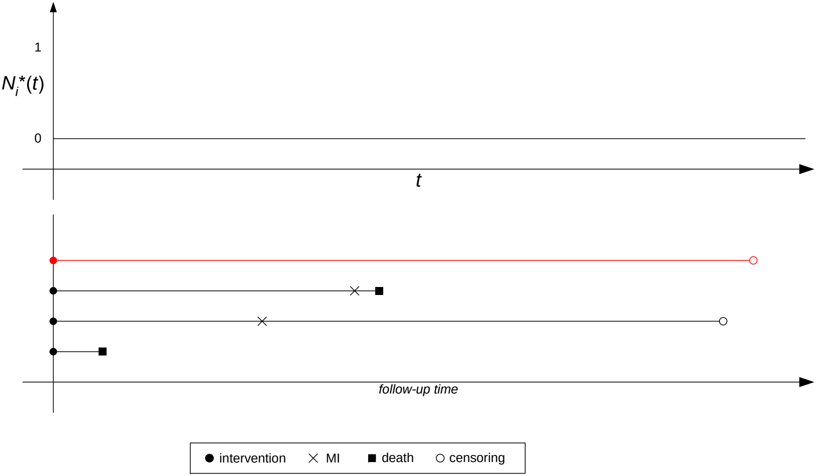
$N_i^*(t)$: counting process for the composite endpoint
(in the absence of withdrawal)



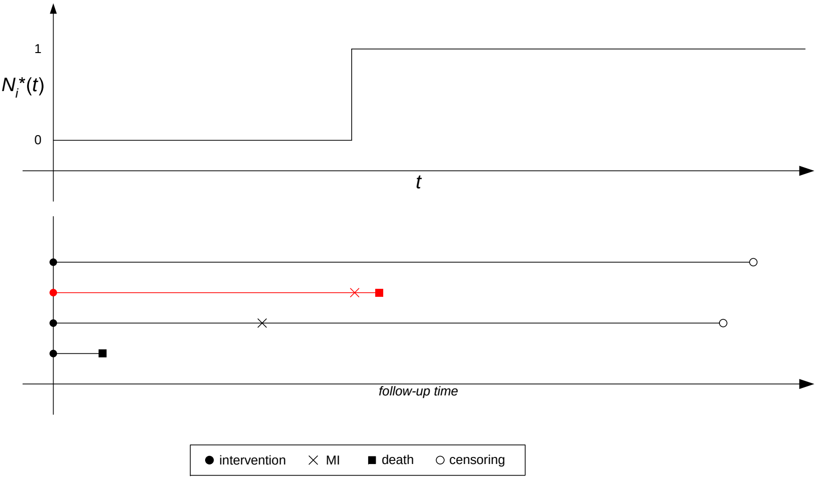
$N_i^*(t)$: counting process for the composite endpoint
(in the absence of withdrawal)



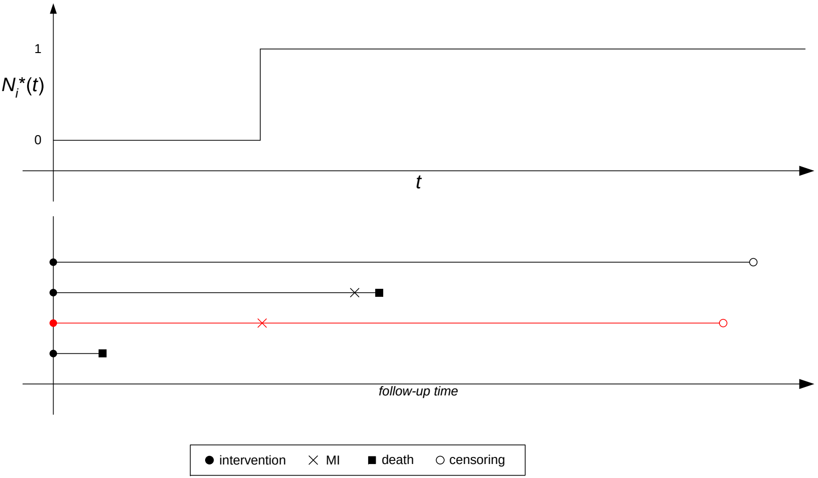
$N_i^*(t)$: counting process for the composite endpoint
(in the absence of withdrawal)



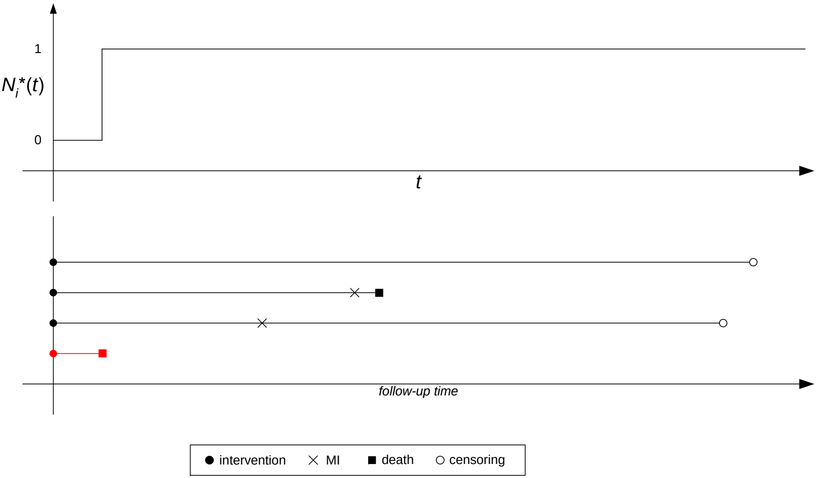
$N_i^*(t)$: counting process for the composite endpoint
(in the absence of withdrawal)



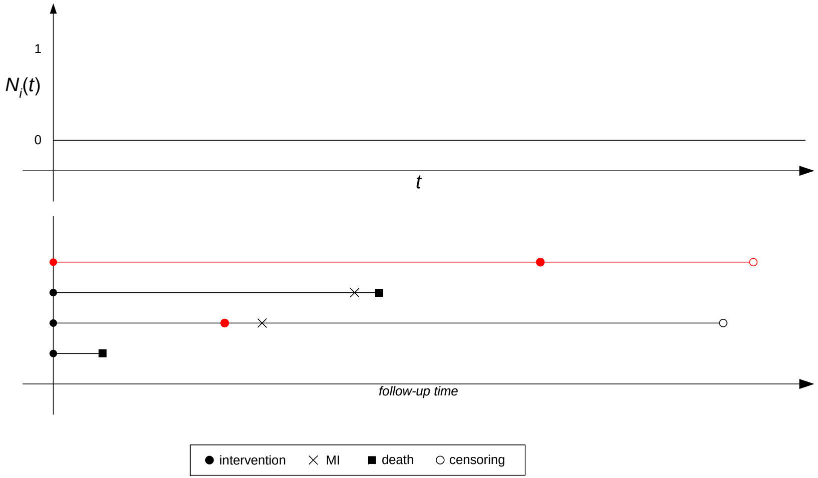
$N_i^*(t)$: counting process for the composite endpoint
(in the absence of withdrawal)



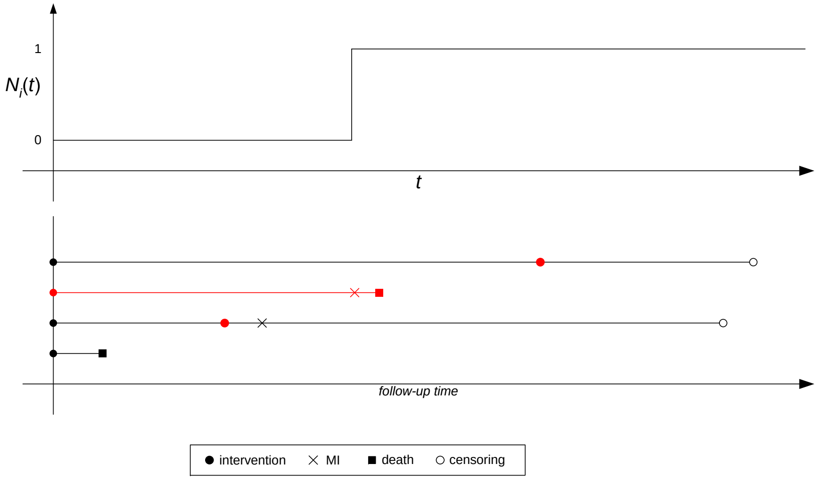
$N_i^*(t)$: counting process for the composite endpoint
(in the absence of withdrawal)



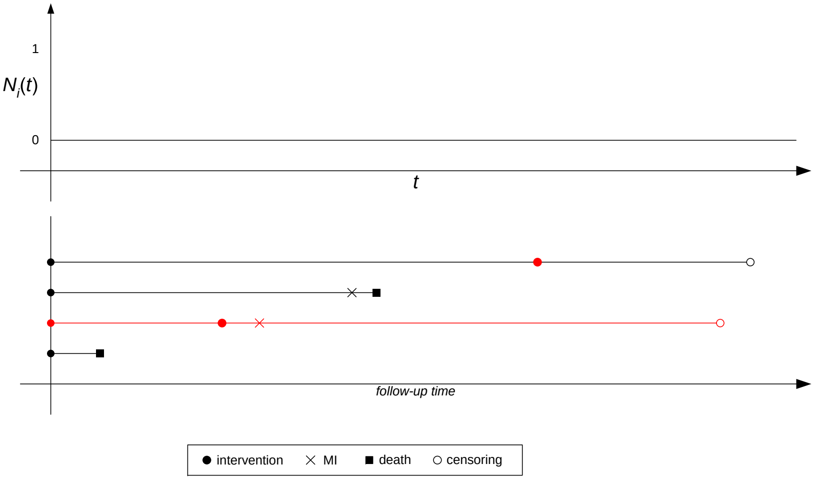
$N_i(t)$: counting process for the composite endpoint
with withdrawal



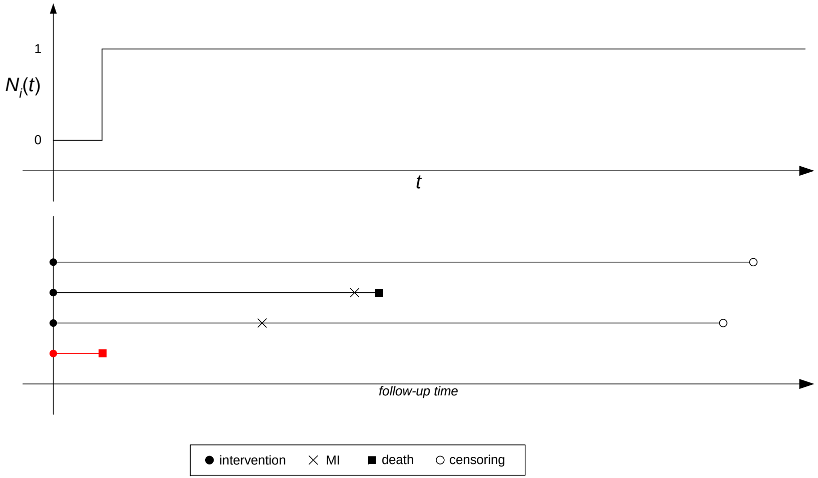
$N_i(t)$: counting process for the composite endpoint
with withdrawal



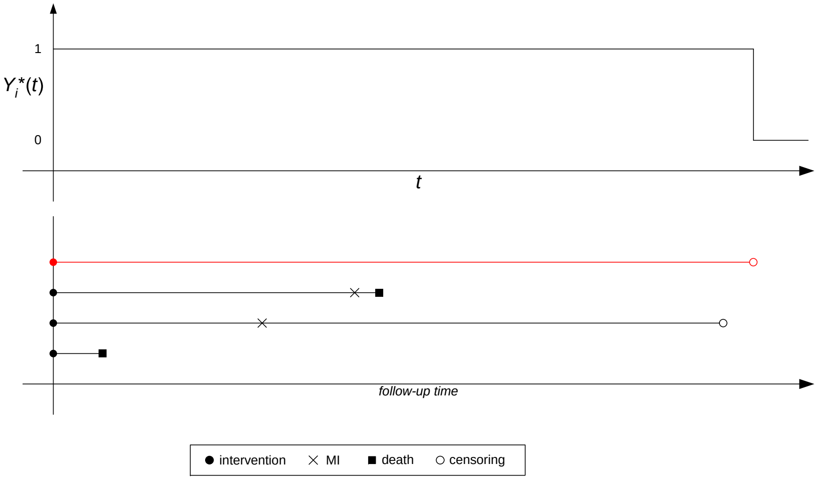
$N_i(t)$: counting process for the composite endpoint
with withdrawal



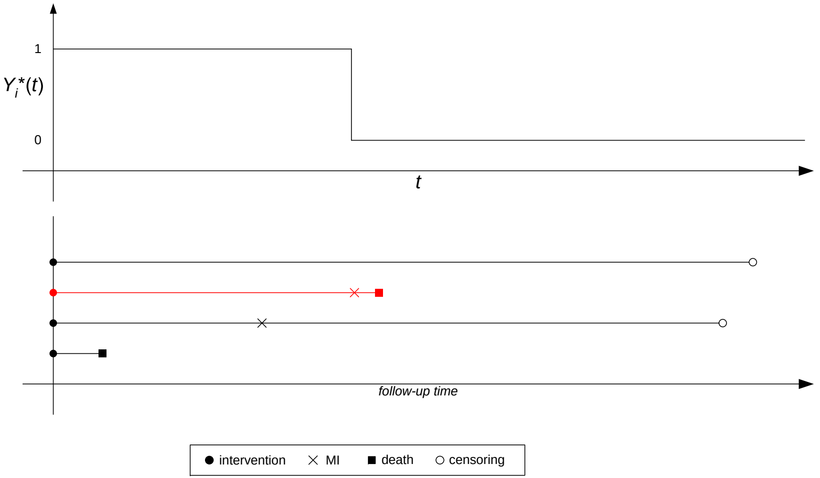
$N_i(t)$: counting process for the composite endpoint
with withdrawal



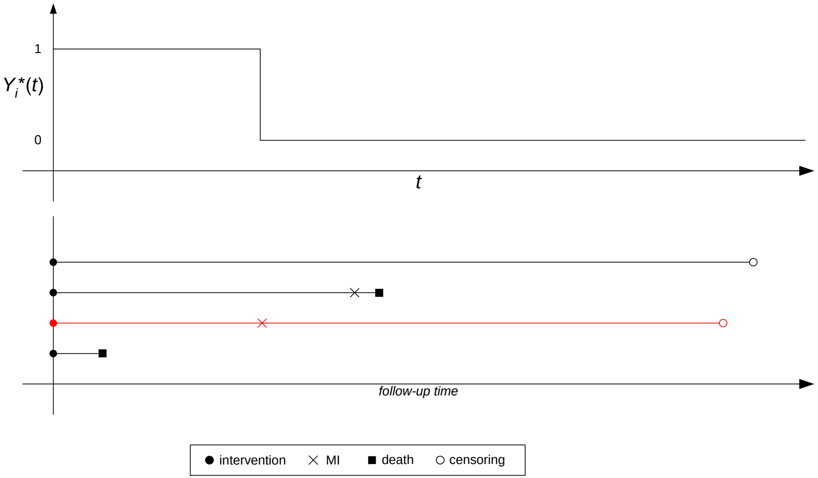
$Y_i^*(t)$: risk set for the composite endpoint
(in the absence of withdrawal)



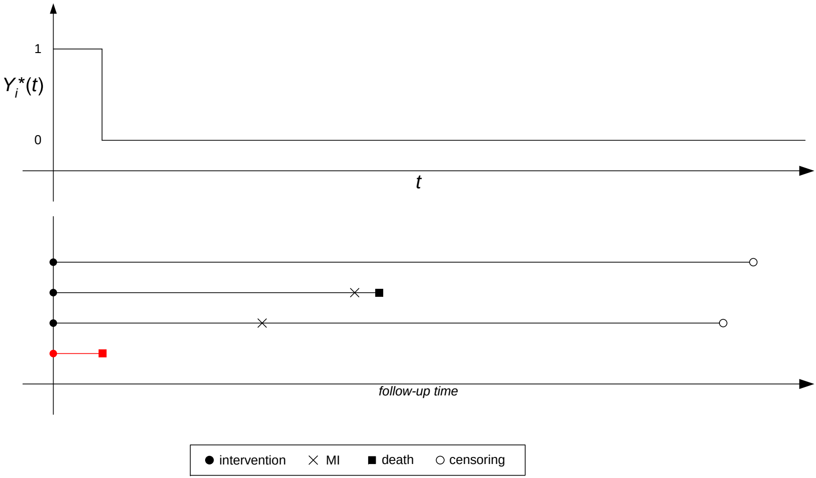
$Y_i^*(t)$: risk set for the composite endpoint
(in the absence of withdrawal)



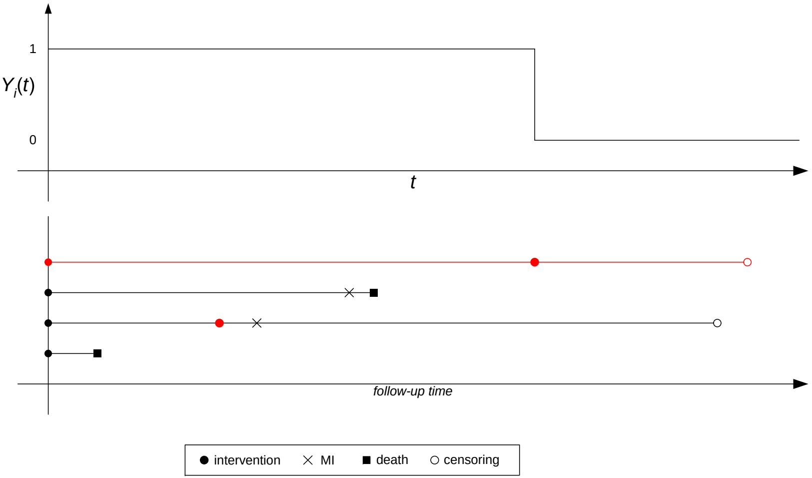
$Y_i^*(t)$: risk set for the composite endpoint
(in the absence of withdrawal)



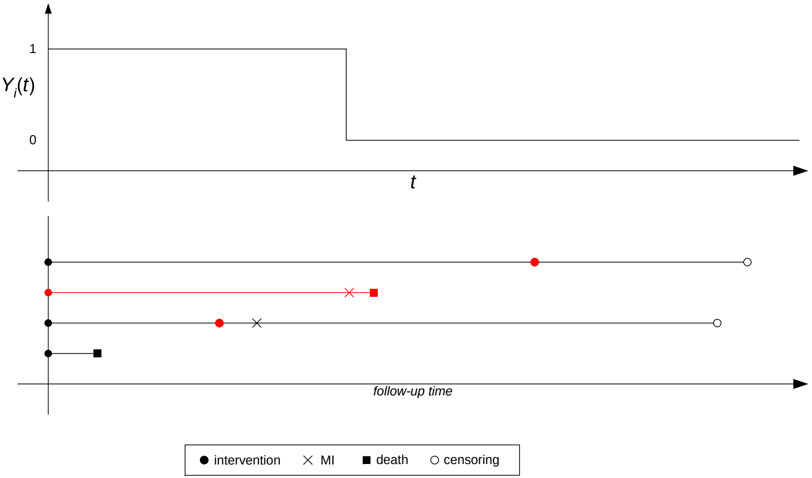
$Y_i^*(t)$: risk set for the composite endpoint
(in the absence of withdrawal)



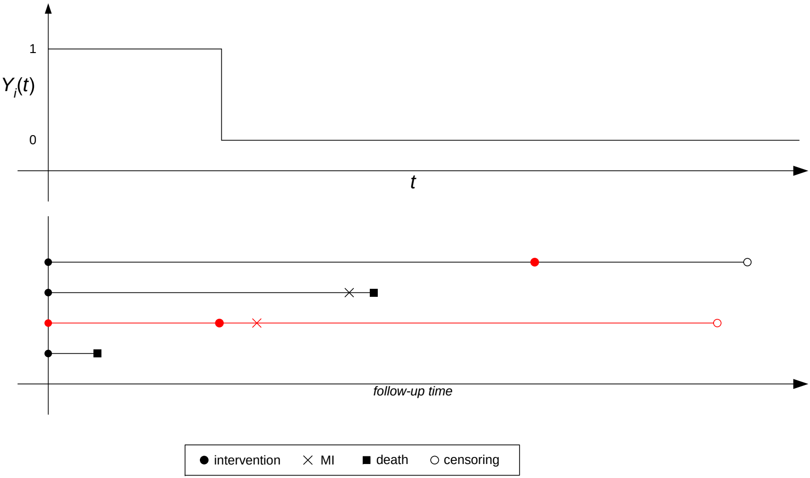
$Y_i(t)$: risk set for the composite endpoint
with withdrawal



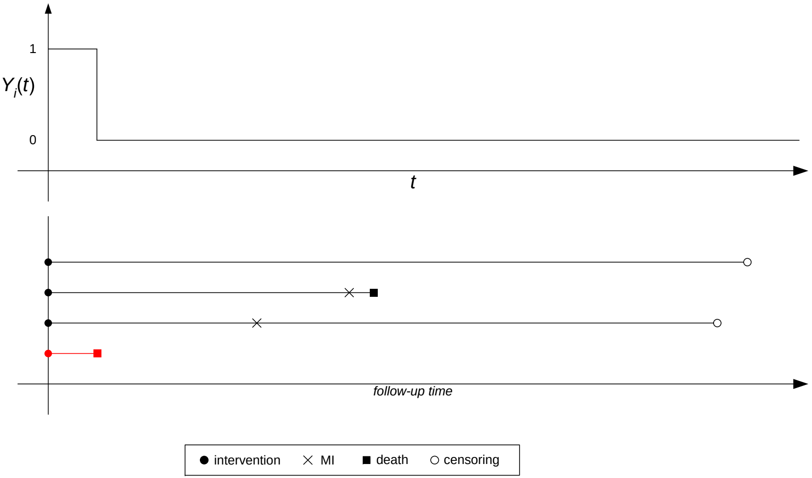
$Y_i(t)$: risk set for the composite endpoint
with withdrawal



$Y_i(t)$: risk set for the composite endpoint
with withdrawal



$Y_i(t)$: risk set for the composite endpoint
with withdrawal



Why is this notation useful?

— It allows us to express the process in this form:

$$\underbrace{dN_i^*(t)}_{\text{`outcome'}} = \underbrace{\overbrace{d\Lambda(t)}^{\lambda(t)\cdot dt}\, Y_i^*(t)}_{\text{`fixed part'}} + \underbrace{d\ \overbrace{\zeta_i(t)}^{\text{martingale}}}_{\text{`error'}}$$

Why is this notation useful?

— It allows us to express the process in this form:

$$\underbrace{dN_i^*(t)}_{\text{'outcome'}} = \underbrace{\overbrace{d\Lambda(t)}^{\lambda(t)\cdot dt}\, Y_i^*(t)}_{\text{'fixed part'}} + \underbrace{d\,\overbrace{\zeta_i(t)}^{\text{martingale}}}_{\text{'error'}}$$

— Properties of martingales are well-understood.

Why is this notation useful?

— It allows us to express the process in this form:

$$\underbrace{dN_i^*(t)}_{\text{'outcome'}} = \underbrace{\overbrace{d\Lambda(t)}^{\lambda(t)\cdot dt} Y_i^*(t)}_{\text{'fixed part'}} + \underbrace{d\ \overbrace{\zeta_i(t)}^{\text{martingale}}}_{\text{'error'}}$$

— Properties of martingales are well-understood.

— This has the feel of a regression model.

Why is this notation useful?

— It allows us to express the process in this form:

$$\underbrace{dN_i^*(t)}_{\text{'outcome'}} = \underbrace{\overbrace{d\Lambda(t)}^{\lambda(t)\cdot dt} Y_i^*(t)}_{\text{'fixed part'}} + \underbrace{d\overbrace{\zeta_i(t)}^{\text{martingale}}}_{\text{'error'}}$$

— Properties of martingales are well-understood.
— This has the feel of a regression model.
— Estimators via estimating equations, eg familiar Nelson–Aalen:

$$\sum_{i=1}^{n} \left\{ dN_i^*(t) - d\Lambda(t)\, Y_i^*(t) \right\} = 0.$$

Why is this notation useful?

— It allows us to express the process in this form:

$$\underbrace{dN_i^*(t)}_{\text{'outcome'}} = \underbrace{\overbrace{d\Lambda(t)}^{\lambda(t)\cdot dt}\, Y_i^*(t)}_{\text{'fixed part'}} + \underbrace{d\,\overbrace{\zeta_i(t)}^{\text{martingale}}}_{\text{'error'}}$$

— Properties of martingales are well-understood.
— This has the feel of a regression model.
— Estimators via estimating equations, eg familiar Nelson–Aalen:

$$\sum_{i=1}^{n} \{dN_i^*(t) - d\Lambda(t)\, Y_i^*(t)\} = 0.$$

References

Andersen PK et al (1993) *Statistical Models based on Counting Processes*.
Aalen OO et al (2008) *Survival and Event History Analysis: A Process Point of View*.

Why is this notation useful?

— It allows us to express the process in this form:

$$\underbrace{dN_i^*(t)}_{\text{'outcome'}} = \underbrace{\overbrace{d\Lambda(t)}^{\lambda(t)\cdot dt}\, Y_i^*(t)}_{\text{'fixed part'}} + \underbrace{d\,\overbrace{\zeta_i(t)}^{\text{martingale}}}_{\text{'error'}}$$

— Properties of martingales are well-understood.
— This has the feel of a regression model.
— Estimators via estimating equations, eg familiar Nelson–Aalen:

$$\sum_{i=1}^{n} \{dN_i^*(t) - d\Lambda(t)\, Y_i^*(t)\} = 0.$$

References

Andersen PK et al (1993) *Statistical Models based on Counting Processes*. *—Hardcore*
Aalen OO et al (2008) *Survival and Event History Analysis: A Process Point of View*. *—More intuitive*

Outline

- Setting
- Notation
- Counting processes
- Semiparametric theory
- Estimators for the distribution of time to composite endpoint
  full data $\rightarrow$ complete cases $\rightarrow$ inverse probability weighted CC $\rightarrow$ augmented IPWCC
- Double robustness and semiparametric efficiency
- Variance estimation
- Simulation study
- Summary and further issues

Statistical models

— A model $\mathcal{M}$ for a set of i.i.d. observations $Z_1, Z_2, \ldots, Z_n$ is a set of densities

$$\mathcal{M} = \{p(z, \theta) : \theta \in \Theta\}$$

that could have given rise to the data, indexed by a parameter θ.

Statistical models

— A model $\mathcal{M}$ for a set of i.i.d. observations $Z_1, Z_2, \ldots, Z_n$ is a set of densities

$$\mathcal{M} = \{p(z, \theta) : \theta \in \Theta\}$$

that could have given rise to the data, indexed by a parameter θ.
— If θ is finite-dimensional, then $\mathcal{M}$ is parametric.

Statistical models

— A model $\mathcal{M}$ for a set of i.i.d. observations $Z_1, Z_2, \ldots, Z_n$ is a set of densities

$$\mathcal{M} = \{p(z, \theta) : \theta \in \Theta\}$$

that could have given rise to the data, indexed by a parameter θ.
— If θ is finite-dimensional, then $\mathcal{M}$ is parametric.
— If θ can be partitioned as

$$\theta = \left(\beta^T, \eta^T\right)^T$$

where β is a finite-dimensional parameter of interest, and η is an infinite-dimensional nuisance parameter, then $\mathcal{M}$ is semiparametric.

Statistical models

— A model $\mathcal{M}$ for a set of i.i.d. observations $Z_1, Z_2, \ldots, Z_n$ is a set of densities

$$\mathcal{M} = \{p(z, \theta) : \theta \in \Theta\}$$

that could have given rise to the data, indexed by a parameter θ.
— If θ is finite-dimensional, then $\mathcal{M}$ is parametric.
— If θ can be partitioned as

$$\theta = \left(\beta^T, \eta^T\right)^T$$

where β is a finite-dimensional parameter of interest, and η is an infinite-dimensional nuisance parameter, then $\mathcal{M}$ is semiparametric.
— In practice, such models arise when η can be any real-valued function (cf baseline hazard function). For this reason, we write ∞-dimensional η as $\eta(\cdot)$.

Statistical models

— A model $\mathcal{M}$ for a set of i.i.d. observations $Z_1, Z_2, \ldots, Z_n$ is a set of densities

$$\mathcal{M} = \{p(z, \theta) : \theta \in \Theta\}$$

that could have given rise to the data, indexed by a parameter θ.
— If θ is finite-dimensional, then $\mathcal{M}$ is parametric.
— If θ can be partitioned as

$$\theta = \left(\beta^T, \eta^T\right)^T$$

where β is a finite-dimensional parameter of interest, and η is an infinite-dimensional nuisance parameter, then $\mathcal{M}$ is semiparametric.
— In practice, such models arise when η can be any real-valued function (cf baseline hazard function). For this reason, we write ∞-dimensional η as $\eta(\cdot)$.
— If $\mathcal{M}$ contains all possible densities for Z then it is nonparametric.

Why be interested in semiparametric models?

— In a perfect randomised clinical trial, very little needs to be modelled. We can usually compare the distribution of the outcome between two groups with minimal parametric assumptions.

Why be interested in semiparametric models?

— In a perfect randomised clinical trial, very little needs to be modelled. We can usually compare the distribution of the outcome between two groups with minimal parametric assumptions.
— As we move away from this ideal, either because the data are incomplete, subject to measurement error, or the treatment is not randomised requiring confounder-adjustments, we often need to use more variables (confounders, or variables conditional on which we believe MAR approximately holds etc).

— In a perfect randomised clinical trial, very little needs to be modelled. We can usually compare the distribution of the outcome between two groups with minimal parametric assumptions.
— As we move away from this ideal, either because the data are incomplete, subject to measurement error, or the treatment is not randomised requiring confounder-adjustments, we often need to use more variables (confounders, or variables conditional on which we believe MAR approximately holds etc).
— The more variables we have, the more difficult it gets to specify a correct parametric model for their joint distribution (or whatever aspects of that joint distribution we require).

Why be interested in semiparametric models?

— In a perfect randomised clinical trial, very little needs to be modelled. We can usually compare the distribution of the outcome between two groups with minimal parametric assumptions.
— As we move away from this ideal, either because the data are incomplete, subject to measurement error, or the treatment is not randomised requiring confounder-adjustments, we often need to use more variables (confounders, or variables conditional on which we believe MAR approximately holds etc).
— The more variables we have, the more difficult it gets to specify a correct parametric model for their joint distribution (or whatever aspects of that joint distribution we require).
— In these 'non-ideal' settings, semiparametric models that leave some of these additional modelling aspects unspecified, are particularly appealing.

Semiparametric estimator

— Given a semiparametric model $\mathcal{M}$, a semiparametric estimator $\hat{\beta}$ of q-dimensional β must be consistent

$$\hat{\beta} - \beta \xrightarrow{\mathcal{P}\{\beta,\eta(\cdot)\}} 0$$

and asymptotically normal

$$n^{\frac{1}{2}}\left(\hat{\beta} - \beta\right) \xrightarrow{\mathcal{D}\{\beta,\eta(\cdot)\}} N\left(0, \Sigma^{q\times q}\left\{\beta, \eta\left(\cdot\right)\right\}\right)$$

for all $p\left\{z, \beta, \eta\left(\cdot\right)\right\} \in \mathcal{M}$.

$*$ $\xrightarrow{\mathcal{P}\{\beta,\eta(\cdot)\}}$: convergence in probability, $\xrightarrow{\mathcal{D}\{\beta,\eta(\cdot)\}}$: convergence in distribution, for density $p\left\{z, \beta, \eta\left(\cdot\right)\right\}$.

Semiparametric estimator

— Given a semiparametric model $\mathcal{M}$, a semiparametric estimator $\hat{\beta}$ of q-dimensional β must be consistent

$$\hat{\beta} - \beta \xrightarrow{\mathcal{P}\{\beta,\eta(\cdot)\}} 0$$

and asymptotically normal

$$n^{\frac{1}{2}} \left(\hat{\beta} - \beta\right) \xrightarrow{\mathcal{D}\{\beta,\eta(\cdot)\}} N\left(0, \Sigma^{q \times q}\left\{\beta, \eta\left(\cdot\right)\right\}\right)$$

for all $p\left\{z, \beta, \eta\left(\cdot\right)\right\} \in \mathcal{M}$.
— So, the larger the model, the smaller the class of semiparametric estimators.

$*$ $\xrightarrow{\mathcal{P}\{\beta,\eta(\cdot)\}}$: convergence in probability, $\xrightarrow{\mathcal{D}\{\beta,\eta(\cdot)\}}$: convergence in distribution, for density $p\left\{z, \beta, \eta\left(\cdot\right)\right\}$.

Asymptotically linear estimators and influence functions

— An estimator $\hat{\beta}$ is asymptotically linear if it can be written as

$$n^{\frac{1}{2}}\left(\hat{\beta}-\beta_0\right)=n^{-\frac{1}{2}}\sum_{i=1}^{n}\varphi\left(Z_i\right)+o_p\left(1\right)$$

where β_0 is the true value of β, $o_p\left(1\right)$ converges in probability to zero, and $\varphi\left(Z_i\right)$ is a $(q \times 1)$ random vector, $E_{\theta_0}\left\{\varphi\left(Z_i\right)\right\}=0$, $E_{\theta_0}\left\{\varphi\left(Z_i\right)\varphi\left(Z_i\right)^T\right\}<\infty$, det $\left[E_{\theta_0}\left\{\varphi\left(Z_i\right)\varphi\left(Z_i\right)^T\right\}\right]\neq 0$.

Asymptotically linear estimators and influence functions

— An estimator $\hat{\beta}$ is asymptotically linear if it can be written as

$$n^{\frac{1}{2}}\left(\hat{\beta}-\beta_0\right) = n^{-\frac{1}{2}}\sum_{i=1}^{n}\varphi\left(Z_i\right)+o_p\left(1\right)$$

where β_0 is the true value of β, $o_p\left(1\right)$ converges in probability to zero, and $\varphi\left(Z_i\right)$ is a $(q \times 1)$ random vector, $E_{\theta_0}\left\{\varphi\left(Z_i\right)\right\}=0$, $E_{\theta_0}\left\{\varphi\left(Z_i\right)\varphi\left(Z_i\right)^T\right\}<\infty$, det $\left[E_{\theta_0}\left\{\varphi\left(Z_i\right)\varphi\left(Z_i\right)^T\right\}\right]\neq 0$.

— $\varphi\left(Z_i\right)$ is the i^{th} **influence function** of $\hat{\beta}$.

Asymptotically linear estimators and influence functions

— An estimator $\hat{\beta}$ is asymptotically linear if it can be written as

$$n^{\frac{1}{2}}\left(\hat{\beta}-\beta_0\right)=n^{-\frac{1}{2}}\sum_{i=1}^{n}\varphi\left(Z_i\right)+o_p\left(1\right)$$

where β_0 is the true value of β, $o_p\left(1\right)$ converges in probability to zero, and $\varphi\left(Z_i\right)$ is a $\left(q\times 1\right)$ random vector, $E_{\theta_0}\left\{\varphi\left(Z_i\right)\right\}=0$, $E_{\theta_0}\left\{\varphi\left(Z_i\right)\varphi\left(Z_i\right)^T\right\}<\infty$, $\det\left[E_{\theta_0}\left\{\varphi\left(Z_i\right)\varphi\left(Z_i\right)^T\right\}\right]\neq 0$.
— $\varphi\left(Z_i\right)$ is the i^{th} **influence function** of $\hat{\beta}$.
— Note that

$$n^{\frac{1}{2}}\left(\hat{\beta}-\beta_0\right)\xrightarrow{\mathcal{D}\left\{\beta_0,\eta_0(\cdot)\right\}}N\left(0,E_{\theta_0}\left\{\varphi\left(Z_i\right)\varphi\left(Z_i\right)^T\right\}\right).$$

Thus the asymptotic properties of an asymptotically linear estimator are governed by its influence function.

Semiparametric theory (in a slide!)

— Maximum likelihood $\rightarrow$ efficient estimators in parametric models.

Semiparametric theory (in a slide!)

— Maximum likelihood $\rightarrow$ efficient estimators in parametric models.
— Historically, estimation in semiparametric models was less systematic (partial likelihood, quasi likelihood, ...)

Semiparametric theory (in a slide!)

— Maximum likelihood $\rightarrow$ efficient estimators in parametric models.
— Historically, estimation in semiparametric models was less systematic (partial likelihood, quasi likelihood, . . .)
— These are good estimators, but context-specific and optimality properties not always clear.

Semiparametric theory (in a slide!)

— Maximum likelihood $\rightarrow$ efficient estimators in parametric models.
— Historically, estimation in semiparametric models was less systematic (partial likelihood, quasi likelihood, ...)
— These are good estimators, but context-specific and optimality properties not always clear.
— Semiparametric theory fixes this. Influence functions live in Hilbert spaces, whose geometry is well-understood. Loosely, distances in Hilbert spaces represent variances, so efficient estimators have influence functions close to the origin.

Semiparametric theory (in a slide!)

— Maximum likelihood $\rightarrow$ efficient estimators in parametric models.
— Historically, estimation in semiparametric models was less systematic (partial likelihood, quasi likelihood, ...)
— These are good estimators, but context-specific and optimality properties not always clear.
— Semiparametric theory fixes this. Influence functions live in Hilbert spaces, whose geometry is well-understood. Loosely, distances in Hilbert spaces represent variances, so efficient estimators have influence functions close to the origin.
— It gives a neat way of constructing good estimators. If we can find 'any old' semiparametric estimator, we can use geometric projections to find a better/best estimator.

Semiparametric theory (in a slide!)

— Maximum likelihood $\rightarrow$ efficient estimators in parametric models.
— Historically, estimation in semiparametric models was less systematic (partial likelihood, quasi likelihood, ...)
— These are good estimators, but context-specific and optimality properties not always clear.
— Semiparametric theory fixes this. Influence functions live in Hilbert spaces, whose geometry is well-understood. Loosely, distances in Hilbert spaces represent variances, so efficient estimators have influence functions close to the origin.
— It gives a neat way of constructing good estimators. If we can find 'any old' semiparametric estimator, we can use geometric projections to find a better/best estimator.
— Useful for missing data, where a 'bad' semiparametric estimator is easy to construct (IPW). This is then improved by augmentation.

Semiparametric theory (in a slide!)

— Maximum likelihood $\rightarrow$ efficient estimators in parametric models.
— Historically, estimation in semiparametric models was less systematic (partial likelihood, quasi likelihood, ...)
— These are good estimators, but context-specific and optimality properties not always clear.
— Semiparametric theory fixes this. Influence functions live in Hilbert spaces, whose geometry is well-understood. Loosely, distances in Hilbert spaces represent variances, so efficient estimators have influence functions close to the origin.
— It gives a neat way of constructing good estimators. If we can find 'any old' semiparametric estimator, we can use geometric projections to find a better/best estimator.
— Useful for missing data, where a 'bad' semiparametric estimator is easy to construct (IPW). This is then improved by augmentation.

Key reference

Tsiatis (2006) *Semiparametric Theory and Missing Data*.

Outline

- Setting
- Notation
- Counting processes
- Semiparametric theory
- Estimators for the distribution of time to composite endpoint
  full data $\rightarrow$ complete cases $\rightarrow$ inverse probability weighted CC $\rightarrow$ augmented IPWCC
- Double robustness and semiparametric efficiency
- Variance estimation
- Simulation study
- Summary and further issues

Outline

- Setting
- Notation
- Counting processes
- Semiparametric theory
- Estimators for the distribution of time to composite endpoint
  full data $\rightarrow$ complete cases $\rightarrow$ inverse probability weighted CC $\rightarrow$ augmented IPWCC
- Double robustness and semiparametric efficiency
- Variance estimation
- Simulation study
- Summary and further issues

Full data estimator

— Were there no withdrawals, we would assume independent censoring ($C_i \perp\!\!\!\perp T_i$) and no further assumptions.

Full data estimator

— Were there no withdrawals, we would assume independent censoring ($C_i \perp\!\!\!\perp T_i$) and no further assumptions.
— Call this model $\mathcal{M}_{\text{full}}$.

Full data estimator

— Were there no withdrawals, we would assume independent censoring ($C_i \perp\!\!\!\perp T_i$) and no further assumptions.
— Call this model $\mathcal{M}_{\text{full}}$.
— Then we would solve:

$$\sum_{i=1}^{n} \{dN_i^*(t) - d\Lambda(t)\, Y_i^*(t)\} = 0$$

Full data estimator

— Were there no withdrawals, we would assume independent censoring ($C_i \perp\!\!\!\perp T_i$) and no further assumptions.
— Call this model $\mathcal{M}_{\text{full}}$.
— Then we would solve:

$$\sum_{i=1}^{n} \{dN_i^* (t) - d\Lambda (t) Y_i^* (t)\} = 0$$

— This leads to the Nelson–Aalen cumulative hazard estimator:

$$\hat{\Lambda}^{\text{full}} (t) = \int_0^t \frac{\sum_{i=1}^n dN_i^* (u)}{\sum_{i=1}^n Y_i^* (u)}$$

Full data estimator

— Were there no withdrawals, we would assume independent censoring ($C_i \perp\!\!\!\perp T_i$) and no further assumptions.
— Call this model $\mathcal{M}_{\text{full}}$.
— Then we would solve:

$$\sum_{i=1}^{n} \{dN_i^*(t) - d\Lambda(t)\, Y_i^*(t)\} = 0$$

— This leads to the Nelson–Aalen cumulative hazard estimator:

$$\hat{\Lambda}^{\text{full}}(t) = \int_0^t \frac{\sum_{i=1}^{n} dN_i^*(u)}{\sum_{i=1}^{n} Y_i^*(u)}$$

— And the Breslow estimator of the survivor function:

$$\hat{S}^{\text{full}}(t) = \exp\left\{-\hat{\Lambda}^{\text{full}}(t)\right\}.$$

Full data estimator

— Were there no withdrawals, we would assume independent censoring ($C_i \perp\!\!\!\perp T_i$) and no further assumptions.
— Call this model $\mathcal{M}_{\text{full}}$.
— Then we would solve:

$$\sum_{i=1}^{n} \{dN_i^* (t) - d\Lambda (t) \, Y_i^* (t)\} = 0$$

— This leads to the Nelson–Aalen cumulative hazard estimator:

$$\hat{\Lambda}^{\text{full}} (t) = \int_0^t \frac{\sum_{i=1}^n dN_i^* (u)}{\sum_{i=1}^n Y_i^* (u)}$$

— And the Breslow estimator of the survivor function:

$$\hat{S}^{\text{full}} (t) = \exp \left\{ -\hat{\Lambda}^{\text{full}} (t) \right\}.$$

— $\hat{\Lambda}^{\text{full}} (t)$ is the *only* semiparametric estimator of $\Lambda (t)$ for $\mathcal{M}_{\text{full}}$, and thus it is (trivially) semiparametric efficient (see Tsiatis (2006) for precise definition).

Outline

- Setting
- Notation
- Counting processes
- Semiparametric theory
- Estimators for the distribution of time to composite endpoint
  full data → complete cases → inverse probability weighted CC → augmented IPWCC
- Double robustness and semiparametric efficiency
- Variance estimation
- Simulation study
- Summary and further issues

Complete case estimator (1)

— In the presence of withdrawals, we don't have the data to solve

$$\sum_{i=1}^{n} \{dN_i^* (t) - d\Lambda (t) Y_i^* (t)\} = 0.$$

Complete case estimator (1)

— In the presence of withdrawals, we don't have the data to solve

$$\sum_{i=1}^{n} \{dN_i^* (t) - d\Lambda (t) \, Y_i^* (t)\} = 0.$$

— To proceed we must assume something about the withdrawal mechanism.

Complete case estimator (1)

— In the presence of withdrawals, we don't have the data to solve

$$\sum_{i=1}^{n} \left\{ dN_i^* (t) - d\Lambda (t) \, Y_i^* (t) \right\} = 0.$$

— To proceed we must assume something about the withdrawal mechanism.

— Suppose we assume that **withdrawal is independent**:

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} Pr \left(t \leq W_i < t + \Delta t \,|\, W_i \geq t, U_i^*, \Delta_i^* \right) = I \left(U_i^* \geq t \right) \kappa (t) \, .$$

Complete case estimator (1)

— In the presence of withdrawals, we don't have the data to solve

$$\sum_{i=1}^{n} \{dN_i^* (t) - d\Lambda (t) Y_i^* (t)\} = 0.$$

— To proceed we must assume something about the withdrawal mechanism.

— Suppose we assume that **withdrawal is independent**:

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} Pr (t \leq W_i < t + \Delta t \,|\, W_i \geq t, U_i^*, \Delta_i^*) = I (U_i^* \geq t) \, \kappa (t) \, .$$

— Call this model (independent withdrawal, together with independent censoring, but nothing else) $\mathcal{M}_{\text{ind}}$.

Complete case estimator (1)

— In the presence of withdrawals, we don't have the data to solve

$$\sum_{i=1}^{n} \{dN_i^* (t) - d\Lambda (t) Y_i^* (t)\} = 0.$$

— To proceed we must assume something about the withdrawal mechanism.

— Suppose we assume that **withdrawal is independent**:

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} Pr (t \leq W_i < t + \Delta t \,|W_i \geq t, U_i^*, \Delta_i^*) = I (U_i^* \geq t) \kappa (t) .$$

— Call this model (independent withdrawal, together with independent censoring, but nothing else) $\mathcal{M}_{\text{ind}}$.

— Then withdrawal is 'just another form of censoring' and so the Nelson–Aalen estimator for the observed data is consistent.

Complete case estimator (1)

— In the presence of withdrawals, we don't have the data to solve

$$\sum_{i=1}^{n} \{dN_i^*(t) - d\Lambda(t) Y_i^*(t)\} = 0.$$

— To proceed we must assume something about the withdrawal mechanism.

— Suppose we assume that **withdrawal is independent**:

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} Pr(t \leq W_i < t + \Delta t \,|\, W_i \geq t, U_i^*, \Delta_i^*) = I(U_i^* \geq t)\kappa(t).$$

— Call this model (independent withdrawal, together with independent censoring, but nothing else) $\mathcal{M}_{\text{ind}}$.

— Then withdrawal is 'just another form of censoring' and so the Nelson–Aalen estimator for the observed data is consistent.

— That is, we solve:

$$\sum_{i=1}^{n} \{dN_i(t) - d\Lambda(t) Y_i(t)\} = 0.$$

Complete case estimator (2)

— We now confirm that this estimator is consistent.

Complete case estimator (2)

— We now confirm that this estimator is consistent.
— To do so, we need to show that

$$E\left\{dN_i(t) - d\Lambda(t)\, Y_i(t)\right\} = 0.$$

Complete case estimator (2)

— We now confirm that this estimator is consistent.
— To do so, we need to show that

$$E\left\{dN_i(t) - d\Lambda(t)\, Y_i(t)\right\} = 0.$$

— First note that

$$dN_i(t) - d\Lambda(t)\, Y_i(t) = I(W_i > t)\left\{dN_i^*(t) - d\Lambda(t)\, Y_i^*(t)\right\}$$

Complete case estimator (2)

— We now confirm that this estimator is consistent.
— To do so, we need to show that

$$E\left\{dN_i(t) - d\Lambda(t)\, Y_i(t)\right\} = 0.$$

— First note that

$$dN_i(t) - d\Lambda(t)\, Y_i(t) = I(W_i > t)\left\{dN_i^*(t) - d\Lambda(t)\, Y_i^*(t)\right\}$$

— and thus

$$\begin{aligned} E\left\{dN_i(t) - \right. & \left. d\Lambda(t)\, Y_i(t)\right\} \\ &= E\left[I(W_i > t)\left\{dN_i^*(t) - d\Lambda(t)\, Y_i^*(t)\right\}\right] \\ &= E\left(E\left[I(W_i > t)\left\{dN_i^*(t) - d\Lambda(t)\, Y_i^*(t)\right\} \middle| U_i^*, \Delta_i^*\right]\right) \\ &= E\left[Pr\left(W_i > t \,|\, U_i^*, \Delta_i^*\right)\left\{dN_i^*(t) - d\Lambda(t)\, Y_i^*(t)\right\}\right]. \end{aligned}$$

— Independent withdrawal implies:

$$
\begin{aligned}
Pr\left(W_i > t \,|\, U_i^*, \Delta_i^*\right) &= \exp\left\{-\int_0^t I\left(U_i^* \geq u\right)\kappa\left(u\right)du\right\} \\
&= \exp\left\{-\int_0^{\min(t,U_i^*)} \kappa\left(u\right)du\right\} \\
&= I\left(U_i^* \geq t\right)\exp\left\{-\int_0^t \kappa\left(u\right)du\right\} \\
&\quad + I\left(U_i^* < t\right)\exp\left\{-\int_0^{U_i^*} \kappa\left(u\right)du\right\}.
\end{aligned}
$$

Complete case estimator (4)

— Thus,

$$E\left\{dN_i(t) - d\Lambda(t)\,Y_i(t)\right\} = E\left(\left[I\left(U_i^* \geq t\right)\exp\left\{-\int_0^t \kappa(u)\,du\right\} + I\left(U_i^* < t\right)\exp\left\{-\int_0^{U_i^*} \kappa(u)\,du\right\}\right]\left\{dN_i^*(t) - d\Lambda(t)\,Y_i^*(t)\right\}\right).$$

Complete case estimator (5)

— But

$$I\left(U_i^* \geq t\right)\left\{dN_i^*\left(t\right) - d\Lambda\left(t\right)Y_i^*\left(t\right)\right\} = dN_i^*\left(t\right) - d\Lambda\left(t\right)Y_i^*\left(t\right)$$

and

$$I\left(U_i^* < t\right)\left\{dN_i^*\left(t\right) - d\Lambda\left(t\right)Y_i^*\left(t\right)\right\} = 0$$

Complete case estimator (5)

— But

$$I\left(U_i^* \geq t\right)\left\{dN_i^*\left(t\right) - d\Lambda\left(t\right)Y_i^*\left(t\right)\right\} = dN_i^*\left(t\right) - d\Lambda\left(t\right)Y_i^*\left(t\right)$$

and

$$I\left(U_i^* < t\right)\left\{dN_i^*\left(t\right) - d\Lambda\left(t\right)Y_i^*\left(t\right)\right\} = 0$$

— So this simplifies to give:

$$\begin{aligned}
E\left\{dN_i\left(t\right)\right. & \left. - d\Lambda\left(t\right)Y_i\left(t\right)\right\} \\
& = E\left[\exp\left\{-\int_0^t \kappa\left(u\right)du\right\}\left\{dN_i^*\left(t\right) - d\Lambda\left(t\right)Y_i^*\left(t\right)\right\}\right] \\
& = \exp\left\{-\int_0^t \kappa\left(u\right)du\right\}\underbrace{E\left\{dN_i^*\left(t\right) - d\Lambda\left(t\right)Y_i^*\left(t\right)\right\}}_{=0} \\
& = 0
\end{aligned}$$

as required.

Complete case estimator (6)

— This leads to:

$$\hat{\Lambda}^{CC}(t) = \int_0^t \frac{\sum_{i=1}^n dN_i(u)}{\sum_{i=1}^n Y_i(u)}.$$

Complete case estimator (6)

— This leads to:

$$\hat{\Lambda}^{CC}(t) = \int_0^t \frac{\sum_{i=1}^n dN_i(u)}{\sum_{i=1}^n Y_i(u)}.$$

— And the corresponding Breslow estimator of $S(t)$ (which we won't keep stating).

Complete case estimator (6)

— This leads to:

$$\hat{\Lambda}^{CC}(t) = \int_0^t \frac{\sum_{i=1}^n dN_i(u)}{\sum_{i=1}^n Y_i(u)}.$$

— And the corresponding Breslow estimator of $S(t)$ (which we won't keep stating).
— $\hat{\Lambda}^{CC}(t)$ is a semiparametric estimator under model $\mathcal{M}_{\text{ind}}$.

Complete case estimator (6)

— This leads to:

$$\hat{\Lambda}^{CC}(t) = \int_0^t \frac{\sum_{i=1}^n dN_i(u)}{\sum_{i=1}^n Y_i(u)}.$$

— And the corresponding Breslow estimator of $S(t)$ (which we won't keep stating).
— $\hat{\Lambda}^{CC}(t)$ is a semiparametric estimator under model $\mathcal{M}_{\text{ind}}$.
— But it's not the only one, nor the most efficient.

Complete case estimator (6)

— This leads to:

$$\hat{\Lambda}^{CC}(t) = \int_0^t \frac{\sum_{i=1}^n dN_i(u)}{\sum_{i=1}^n Y_i(u)}.$$

— And the corresponding Breslow estimator of $S(t)$ (which we won't keep stating).

— $\hat{\Lambda}^{CC}(t)$ is a semiparametric estimator under model $\mathcal{M}_{\text{ind}}$.

— But it's not the only one, nor the most efficient.

— It does not use the data on (D_i, Γ_i) for those who withdraw.

Complete case estimator (6)

— This leads to:

$$\hat{\Lambda}^{CC}(t) = \int_0^t \frac{\sum_{i=1}^n dN_i(u)}{\sum_{i=1}^n Y_i(u)}.$$

— And the corresponding Breslow estimator of $S(t)$ (which we won't keep stating).
— $\hat{\Lambda}^{CC}(t)$ is a semiparametric estimator under model $\mathcal{M}_{\text{ind}}$.
— But it's not the only one, nor the most efficient.
— It does not use the data on (D_i, Γ_i) for those who withdraw.
— Later we'll augment the estimating function to include these additional data without restricting the model further, improving efficiency.

Complete case estimator (6)

— This leads to:

$$\hat{\Lambda}^{CC}(t) = \int_0^t \frac{\sum_{i=1}^n dN_i(u)}{\sum_{i=1}^n Y_i(u)}.$$

— And the corresponding Breslow estimator of $S(t)$ (which we won't keep stating).
— $\hat{\Lambda}^{CC}(t)$ is a semiparametric estimator under model $\mathcal{M}_{\text{ind}}$.
— But it's not the only one, nor the most efficient.
— It does not use the data on (D_i, Γ_i) for those who withdraw.
— Later we'll augment the estimating function to include these additional data without restricting the model further, improving efficiency.
— First, however, we consider relaxing the assumption of independent withdrawal.

Outline

- Setting
- Notation
- Counting processes
- Semiparametric theory
- Estimators for the distribution of time to composite endpoint
  full data → complete cases → inverse probability weighted CC →
  augmented IPWCC
- Double robustness and semiparametric efficiency
- Variance estimation
- Simulation study
- Summary and further issues

Inverse probability weighted CC estimator (1)

— Independent withdrawal often implausible.

— We may wish to relax this to an assumption of **covariate-driven withdrawal at random**:

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} Pr\left(t \leq W_i < t + \Delta t \,|\, W_i \geq t, \mathcal{F}_i\right) = I\left(U_i^* \geq t\right) \lambda\left\{t, \bar{\mathbf{X}}_i(t)\right\}$$

Inverse probability weighted CC estimator (1)

— Independent withdrawal often implausible.
— We may wish to relax this to an assumption of **covariate-driven withdrawal at random**:

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} Pr\left(t \leq W_i < t + \Delta t \,|W_i \geq t, \mathcal{F}_i\right) = I\left(U_i^* \geq t\right) \lambda \left\{t, \bar{\mathbf{X}}_i\left(t\right)\right\}$$

— Call this model (covariate-driven withdrawal, independent censoring, nothing else) $\mathcal{M}_{\text{CDW}}$.

Inverse probability weighted CC estimator (1)

— Independent withdrawal often implausible.
— We may wish to relax this to an assumption of **covariate-driven withdrawal at random**:

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} Pr\left(t \leq W_i < t + \Delta t \,|\, W_i \geq t, \mathcal{F}_i\right) = I\left(U_i^* \geq t\right) \lambda\left\{t, \bar{\mathbf{X}}_i(t)\right\}$$

— Call this model (covariate-driven withdrawal, independent censoring, nothing else) $\mathcal{M}_{\text{CDW}}$.
— Note that this is a ‘missing at random’ mechanism since $I\left(U_i^* \geq t\right), I\left(U_i^* \geq t\right) \bar{\mathbf{X}}_i(t) \in G_t\left(\mathcal{F}_i\right)$.

Inverse probability weighted CC estimator (1)

— Independent withdrawal often implausible.
— We may wish to relax this to an assumption of **covariate-driven withdrawal at random**:

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} Pr\left(t \leq W_i < t + \Delta t \,|\, W_i \geq t, \mathcal{F}_i\right) = I\left(U_i^* \geq t\right) \lambda\left\{t, \bar{\mathbf{X}}_i\left(t\right)\right\}$$

— Call this model (covariate-driven withdrawal, independent censoring, nothing else) $\mathcal{M}_{\text{CDW}}$.
— Note that this is a 'missing at random' mechanism since $I\left(U_i^* \geq t\right), I\left(U_i^* \geq t\right) \bar{\mathbf{X}}_i\left(t\right) \in G_t\left(\mathcal{F}_i\right)$.
— Recall that under independent withdrawal we had:

$$\begin{aligned} E\left\{dN_i\left(t\right) - d\Lambda\left(t\right) Y_i\left(t\right)\right\} &= E\left[Pr\left(W_i > t \,|\, U_i^*, \Delta_i^*\right)\left\{dN_i^*\left(t\right) - d\Lambda\left(t\right) Y_i^*\left(t\right)\right\}\right] \\ &= \exp\left\{-\int_0^t \kappa\left(u\right) du\right\} E\left\{dN_i^*\left(t\right) - d\Lambda\left(t\right) Y_i^*\left(t\right)\right\} \end{aligned}$$

Inverse probability weighted CC estimator (1)

— Independent withdrawal often implausible.
— We may wish to relax this to an assumption of **covariate-driven withdrawal at random**:

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} Pr\left(t \leq W_i < t + \Delta t \,|\, W_i \geq t, \mathcal{F}_i\right) = I\left(U_i^* \geq t\right) \lambda \left\{t, \bar{\mathbf{X}}_i\left(t\right)\right\}$$

— Call this model (covariate-driven withdrawal, independent censoring, nothing else) $\mathcal{M}_{\mathrm{CDW}}$.
— Note that this is a 'missing at random' mechanism since $I\left(U_i^* \geq t\right), I\left(U_i^* \geq t\right) \bar{\mathbf{X}}_i\left(t\right) \in G_t\left(\mathcal{F}_i\right)$.
— Recall that under independent withdrawal we had:

$$\begin{aligned} E\left\{dN_i\left(t\right) - d\Lambda\left(t\right) Y_i\left(t\right)\right\} &= E\left[Pr\left(W_i > t \,|\, U_i^*, \Delta_i^*\right)\left\{dN_i^*\left(t\right) - d\Lambda\left(t\right) Y_i^*\left(t\right)\right\}\right] \\ &= \exp\left\{-\int_0^t \kappa\left(u\right) du\right\} E\left\{dN_i^*\left(t\right) - d\Lambda\left(t\right) Y_i^*\left(t\right)\right\} \end{aligned}$$

Inverse probability weighted CC estimator (1)

— Independent withdrawal often implausible.
— We may wish to relax this to an assumption of **covariate-driven withdrawal at random**:

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} Pr\left(t \leq W_i < t + \Delta t \,|\, W_i \geq t, \mathcal{F}_i\right) = I\left(U_i^* \geq t\right) \lambda \left\{t, \bar{\mathbf{X}}_i\left(t\right)\right\}$$

— Call this model (covariate-driven withdrawal, independent censoring, nothing else) $\mathcal{M}_{\mathrm{CDW}}$.
— Note that this is a 'missing at random' mechanism since $I\left(U_i^* \geq t\right), I\left(U_i^* \geq t\right) \bar{\mathbf{X}}_i\left(t\right) \in G_t\left(\mathcal{F}_i\right)$.
— Recall that under independent withdrawal we had:

$$\begin{aligned} E\left\{dN_i\left(t\right) - d\Lambda\left(t\right) Y_i\left(t\right)\right\} &= E\left[Pr\left(W_i > t \,|\, U_i^*, \Delta_i^*\right)\left\{dN_i^*\left(t\right) - d\Lambda\left(t\right) Y_i^*\left(t\right)\right\}\right] \\ &= \exp\left\{-\int_0^t \kappa\left(u\right) du\right\} E\left\{dN_i^*\left(t\right) - d\Lambda\left(t\right) Y_i^*\left(t\right)\right\} \end{aligned}$$

— This will not hold, when $Pr\left(W_i > t \,|\, \mathcal{F}_i\right)$ depends on $\bar{\mathbf{X}}_i\left(t\right)$, and $\bar{\mathbf{X}}_i\left(t\right)$ is associated with U_i^*, Δ_i^*.

Inverse probability weighted CC estimator (2)

— So, under covariate-driven withdrawal at random, $\hat{\Lambda}^{\mathrm{CC}}(t)$ is not a consistent estimator of $\Lambda(t)$.

Inverse probability weighted CC estimator (2)

— So, under covariate-driven withdrawal at random, $\hat{\Lambda}^{CC}(t)$ is not a consistent estimator of $\Lambda(t)$.
— Under covariate-driven withdrawal at random, we can correct for this inconsistency by simple inverse probability weighting.

Inverse probability weighted CC estimator (2)

— So, under covariate-driven withdrawal at random, $\hat{\Lambda}^{CC}(t)$ is not a consistent estimator of $\Lambda(t)$.
— Under covariate-driven withdrawal at random, we can correct for this inconsistency by simple inverse probability weighting.
— The estimating equation becomes:

$$\sum_{i=1}^{n} \frac{dN_i(t) - d\Lambda(t)\, Y_i(t)}{Pr\left\{ W_i > t \,\middle|\, U_i^* \geq t, \bar{\mathbf{X}}_i(t) \right\}} = 0,$$

Inverse probability weighted CC estimator (2)

— So, under covariate-driven withdrawal at random, $\hat{\Lambda}^{\mathrm{CC}}(t)$ is not a consistent estimator of $\Lambda(t)$.
— Under covariate-driven withdrawal at random, we can correct for this inconsistency by simple inverse probability weighting.
— The estimating equation becomes:

$$\sum_{i=1}^{n} \frac{dN_i(t) - d\Lambda(t)\, Y_i(t)}{Pr\left\{W_i > t \,\middle|\, U_i^* \geq t, \bar{\mathbf{X}}_i(t)\right\}} = 0,$$

— And the IPWCC estimator is:

$$\hat{\Lambda}^{\mathrm{IPWCC}}(t) = \int_0^t \frac{\sum_{i=1}^n \frac{dN_i(u)}{Pr\left\{W_i > t \middle| U_i^* \geq t, \bar{\mathbf{X}}_i(t)\right\}}}{\sum_{i=1}^n \frac{Y_i(u)}{Pr\left\{W_i > t \middle| U_i^* \geq t, \bar{\mathbf{X}}_i(t)\right\}}}.$$

Inverse probability weighted CC estimator (3)

— To see that this estimator is consistent, note that:

$$
\begin{aligned}
E\left\{\frac{dN_i(t) - d\Lambda(t)\,Y_i(t)}{Pr\left\{W_i > t \,\middle|\, U_i^* \geq t, \bar{\mathbf{X}}_i(t)\right\}}\right\}
&= E\left\{\frac{I(W_i > t)\left\{dN_i^*(t) - d\Lambda(t)\,Y_i^*(t)\right\}}{Pr\left\{W_i > t \,\middle|\, U_i^* \geq t, \bar{\mathbf{X}}_i(t)\right\}}\right\} \\
&= E\left[E\left\{\left.\frac{I(W_i > t)\left\{dN_i^*(t) - d\Lambda(t)\,Y_i^*(t)\right\}}{Pr\left\{W_i > t \,\middle|\, U_i^* \geq t, \bar{\mathbf{X}}_i(t)\right\}}\right| \mathcal{F}_i\right\}\right] \\
&= E\left[\frac{Pr\left(W_i > t \mid \mathcal{F}_i\right)\left\{dN_i^*(t) - d\Lambda(t)\,Y_i^*(t)\right\}}{Pr\left\{W_i > t \,\middle|\, U_i^* \geq t, \bar{\mathbf{X}}_i(t)\right\}}\right] \\
&= E\left\{dN_i^*(t) - d\Lambda(t)\,Y_i^*(t)\right\} = 0
\end{aligned}
$$

under covariate-driven withdrawal at random.

Inverse probability weighted CC estimator (4)

— $\hat{\Lambda}^{\text{IPWCC}}(t)$ is a semiparametric estimator of $\Lambda(t)$ under model $\mathcal{M}_{\text{CDW}}$, but again it is not the only one, and is not semiparametric efficient: it ignores data on (D_i, Γ_i) for those who withdraw.

Inverse probability weighted CC estimator (4)

— $\hat{\Lambda}^{\text{IPWCC}}(t)$ is a semiparametric estimator of $\Lambda(t)$ under model $\mathcal{M}_{\text{CDW}}$, but again it is not the only one, and is not semiparametric efficient: it ignores data on (D_i, Γ_i) for those who withdraw.

— Also, in practice must specify a model for

$$Pr\left\{W_i > t \,\middle|\, U_i^* \geq t, \bar{\mathbf{X}}_i(t)\right\} = K\left\{t, \bar{\mathbf{X}}_i(t); \gamma\right\}$$

Inverse probability weighted CC estimator (4)

— $\hat{\Lambda}^{\text{IPWCC}}(t)$ is a semiparametric estimator of $\Lambda(t)$ under model $\mathcal{M}_{\text{CDW}}$, but again it is not the only one, and is not semiparametric efficient: it ignores data on (D_i, Γ_i) for those who withdraw.

— Also, in practice must specify a model for

$$Pr\left\{W_i > t \,\middle|\, U_i^* \geq t, \bar{\mathbf{X}}_i(t)\right\} = K\left\{t, \bar{\mathbf{X}}_i(t); \gamma\right\}$$

— Let $\mathcal{M}_{\text{CM}}$ (for 'coarsening model') be the set of densities for which this holds.

Inverse probability weighted CC estimator (4)

— $\hat{\Lambda}^{\text{IPWCC}}(t)$ is a semiparametric estimator of $\Lambda(t)$ under model $\mathcal{M}_{\text{CDW}}$, but again it is not the only one, and is not semiparametric efficient: it ignores data on (D_i, Γ_i) for those who withdraw.
— Also, in practice must specify a model for

$$Pr\left\{W_i > t \,\middle|\, U_i^* \geq t, \bar{\mathbf{X}}_i(t)\right\} = K\left\{t, \bar{\mathbf{X}}_i(t); \gamma\right\}$$

— Let $\mathcal{M}_{\text{CM}}$ (for 'coarsening model') be the set of densities for which this holds.
— The feasible IPWCC estimator is then:

$$\hat{\Lambda}^{\text{f-IPWCC}}(t) = \int_0^t \frac{\sum_{i=1}^n \frac{dN_i(u)}{K\left\{t, \bar{\mathbf{X}}_i(t); \hat{\gamma}\right\}}}{\sum_{i=1}^n \frac{Y_i(u)}{K\left\{t, \bar{\mathbf{X}}_i(t); \hat{\gamma}\right\}}}$$

Inverse probability weighted CC estimator (4)

— $\hat{\Lambda}^{\text{IPWCC}}(t)$ is a semiparametric estimator of $\Lambda(t)$ under model $\mathcal{M}_{\text{CDW}}$, but again it is not the only one, and is not semiparametric efficient: it ignores data on (D_i, Γ_i) for those who withdraw.
— Also, in practice must specify a model for

$$Pr\left\{W_i > t \,\middle|\, U_i^* \geq t, \bar{\mathbf{X}}_i(t)\right\} = K\left\{t, \bar{\mathbf{X}}_i(t); \gamma\right\}$$

— Let $\mathcal{M}_{\text{CM}}$ (for 'coarsening model') be the set of densities for which this holds.
— The feasible IPWCC estimator is then:

$$\hat{\Lambda}^{\text{f-IPWCC}}(t) = \int_0^t \frac{\sum_{i=1}^n \frac{dN_i(u)}{K\left\{t, \bar{\mathbf{X}}_i(t); \hat{\gamma}\right\}}}{\sum_{i=1}^n \frac{Y_i(u)}{K\left\{t, \bar{\mathbf{X}}_i(t); \hat{\gamma}\right\}}}$$

— $\hat{\Lambda}^{\text{f-IPWCC}}(t)$ is a semiparametric estimator under $\mathcal{M}_{\text{CDW}} \cap \mathcal{M}_{\text{CM}}$. Also, provided that γ are estimated sufficiently efficiently by $\hat{\gamma}$, $\hat{\Lambda}^{\text{f-IPWCC}}(t)$ is more efficient than $\hat{\Lambda}^{\text{IPWCC}}(t)$. (Why?)

Extended IPWCC

— So far made no use of data on (D_i, Γ_i) for those who withdraw.

Extended IPWCC

— So far made no use of data on (D_i, Γ_i) for those who withdraw.
— The simplest way to incorporate this additional information is to extend the weights model to $\mathcal{M}_{\mathrm{ECM}}$:

$$Pr\left\{W_i > t \,\middle|\, U_i^* \geq t, \bar{\mathbf{X}}_i\left(t\right), D_i, \Gamma_i\right\} = \tilde{K}\left\{t, \bar{\mathbf{X}}_i\left(t\right), D_i, \Gamma_i; \tilde{\gamma}\right\}$$

Extended IPWCC

— So far made no use of data on (D_i, Γ_i) for those who withdraw.
— The simplest way to incorporate this additional information is to extend the weights model to $\mathcal{M}_{\text{ECM}}$:

$$Pr\left\{W_i > t \,\middle|\, U_i^* \geq t, \bar{\mathbf{X}}_i(t), D_i, \Gamma_i\right\} = \tilde{K}\left\{t, \bar{\mathbf{X}}_i(t), D_i, \Gamma_i; \tilde{\gamma}\right\}$$

— The feasible extended IPWCC estimator is:

$$\hat{\Lambda}^{\text{f-IPWCC-ext}}(t) = \int_0^t \frac{\sum_{i=1}^n \frac{dN_i(u)}{\tilde{K}\left\{t, \bar{\mathbf{X}}_i, D_i, \Gamma_i(t); \hat{\tilde{\gamma}}\right\}}}{\sum_{i=1}^n \frac{Y_i(u)}{\tilde{K}\left\{t, \bar{\mathbf{X}}_i, D_i, \Gamma_i(t); \hat{\tilde{\gamma}}\right\}}}$$

Extended IPWCC

— So far made no use of data on (D_i, Γ_i) for those who withdraw.
— The simplest way to incorporate this additional information is to extend the weights model to $\mathcal{M}_{\text{ECM}}$:

$$Pr\left\{W_i > t \,\middle|\, U_i^* \geq t, \bar{\mathbf{X}}_i(t), D_i, \Gamma_i\right\} = \tilde{K}\left\{t, \bar{\mathbf{X}}_i(t), D_i, \Gamma_i; \tilde{\gamma}\right\}$$

— The feasible extended IPWCC estimator is:

$$\hat{\Lambda}^{\text{f-IPWCC-ext}}(t) = \int_0^t \frac{\sum_{i=1}^n \frac{dN_i(u)}{\tilde{K}\left\{t, \bar{\mathbf{X}}_i, D_i, \Gamma_i(t); \hat{\tilde{\gamma}}\right\}}}{\sum_{i=1}^n \frac{Y_i(u)}{\tilde{K}\left\{t, \bar{\mathbf{X}}_i, D_i, \Gamma_i(t); \hat{\tilde{\gamma}}\right\}}}$$

— This has the advantage of being consistent under **covariate-and-death-time-driven withdrawal at random** $(\mathcal{M}_{\text{CDDW}})$:

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} Pr\left(t \leq W_i < t + \Delta t \,|\, W_i \geq t, \mathcal{F}_i\right) = I\left(U_i^* \geq t\right) \nu\left\{t, \bar{\mathbf{X}}_i(t), D_i, \Gamma_i\right\}$$

Extended IPWCC

— So far made no use of data on (D_i, Γ_i) for those who withdraw.
— The simplest way to incorporate this additional information is to extend the weights model to $\mathcal{M}_{\text{ECM}}$:

$$Pr\left\{W_i > t \,\middle|\, U_i^* \geq t, \bar{\mathbf{X}}_i(t), D_i, \Gamma_i\right\} = \tilde{K}\left\{t, \bar{\mathbf{X}}_i(t), D_i, \Gamma_i; \tilde{\gamma}\right\}$$

— The feasible extended IPWCC estimator is:

$$\hat{\Lambda}^{\text{f-IPWCC-ext}}(t) = \int_0^t \frac{\sum_{i=1}^n \frac{dN_i(u)}{\tilde{K}\left\{t, \bar{\mathbf{X}}_i, D_i, \Gamma_i(t); \hat{\tilde{\gamma}}\right\}}}{\sum_{i=1}^n \frac{Y_i(u)}{\tilde{K}\left\{t, \bar{\mathbf{X}}_i, D_i, \Gamma_i(t); \hat{\tilde{\gamma}}\right\}}}$$

— This has the advantage of being consistent under **covariate-and-death-time-driven withdrawal at random** ($\mathcal{M}_{\text{CDDW}}$):

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} Pr\left(t \leq W_i < t + \Delta t \,|\, W_i \geq t, \mathcal{F}_i\right) = I\left(U_i^* \geq t\right) \nu \left\{t, \bar{\mathbf{X}}_i(t), D_i, \Gamma_i\right\}$$

— And is also more efficient than $\hat{\Lambda}^{\text{f-IPWCC}}(t)$ even under $\mathcal{M}_{\text{CDW}} \cap \mathcal{M}_{\text{CM}}$. (Why?)

Extended IPWCC

— So far made no use of data on (D_i, Γ_i) for those who withdraw.
— The simplest way to incorporate this additional information is to extend the weights model to $\mathcal{M}_{\text{ECM}}$:

$$Pr\left\{W_i > t \,\middle|\, U_i^* \geq t, \bar{\mathbf{X}}_i(t), D_i, \Gamma_i\right\} = \tilde{K}\left\{t, \bar{\mathbf{X}}_i(t), D_i, \Gamma_i; \tilde{\gamma}\right\}$$

— The feasible extended IPWCC estimator is:

$$\hat{\Lambda}^{\text{f-IPWCC-ext}}(t) = \int_0^t \frac{\sum_{i=1}^n \frac{dN_i(u)}{\tilde{K}\left\{t, \bar{\mathbf{X}}_i, D_i, \Gamma_i(t); \hat{\tilde{\gamma}}\right\}}}{\sum_{i=1}^n \frac{Y_i(u)}{\tilde{K}\left\{t, \bar{\mathbf{X}}_i, D_i, \Gamma_i(t); \hat{\tilde{\gamma}}\right\}}}$$

— This has the advantage of being consistent under **covariate-and-death-time-driven withdrawal at random** $(\mathcal{M}_{\text{CDDW}})$:

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} Pr\left(t \leq W_i < t + \Delta t \,|\, W_i \geq t, \mathcal{F}_i\right) = I\left(U_i^* \geq t\right) \nu\left\{t, \bar{\mathbf{X}}_i(t), D_i, \Gamma_i\right\}$$

— And is also more efficient than $\hat{\Lambda}^{\text{f-IPWCC}}(t)$ even under $\mathcal{M}_{\text{CDW}} \cap \mathcal{M}_{\text{CM}}$. (Why?)
— But it's not semiparametric efficient. We can do better…

Outline

- Setting
- Notation
- Counting processes
- Semiparametric theory
- Estimators for the distribution of time to composite endpoint
  full data $\rightarrow$ complete cases $\rightarrow$ inverse probability weighted CC $\rightarrow$ augmented IPWCC
- Double robustness and semiparametric efficiency
- Variance estimation
- Simulation study
- Summary and further issues

Augmented IPWCC (1)

— Consider augmenting the IPWCC estimating equation to:

$$\sum_{i=1}^{n}\left[\frac{dN_i(t)-d\Lambda(t)\,Y_i(t)}{Pr\left\{W_i>t\,\middle|\,U_i^*\geq t,\bar{\mathbf{X}}_i(t)\right\}}\right.$$
$$\left.+\int_0^t \frac{dM_i(u)}{Pr\left\{W_i>u\,\middle|\,U_i^*\geq u,\bar{\mathbf{X}}_i(u)\right\}}h\left\{u,G_u\left(\mathcal{F}_i\right)\right\}\right]=0$$

where

$$dM_i(u)=\lim_{\Delta u\to 0}\left[I\left(u\leq W_i<u+\Delta u\right)-\lambda\left\{u,\bar{\mathbf{X}}_i(u)\right\}I\left(W_i\geq u\right)\right]I\left(U_i^*\geq u\right)$$

and $h\left\{u,G_u\left(\mathcal{F}_i\right)\right\}$ is an arbitrary function at time u of $G_u\left(\mathcal{F}_i\right)$.

Augmented IPWCC (2)

— Under covariate-driven withdrawal at random

$$E\left\{dM_i\left(u\right)\middle|G_u\left(\mathcal{F}_i\right)\right\} = 0.$$

Augmented IPWCC (2)

— Under covariate-driven withdrawal at random

$$E\left\{dM_i(u)\,|\,G_u(\mathcal{F}_i)\right\} = 0.$$

— Thus the augmented estimator is consistent under $\mathcal{M}_{\text{CDW}}$ for any choice of $h\{u, G_u(\mathcal{F}_i)\}$.

Augmented IPWCC (2)

— Under covariate-driven withdrawal at random

$$E\left\{dM_i(u)\,|\,G_u(\mathcal{F}_i)\right\} = 0.$$

— Thus the augmented estimator is consistent under $\mathcal{M}_{\text{CDW}}$ for any choice of $h\{u, G_u(\mathcal{F}_i)\}$.
— Semiparametric theory shows that the optimal (most efficient) choice of $h\{u, G_u(\mathcal{F}_i)\}$ is

$$h_{\text{opt}}\{u, G_u(\mathcal{F}_i)\} = E\left\{ dN_i^*(t) - d\Lambda(t)\, Y_i^*(t) \middle|\, G_u(\mathcal{F}_i)\right\}$$

— This conditional expectation is equal to:

$$\frac{I(C_i > t)\, I(U_i^* > u)}{H\left\{u, \bar{\mathbf{X}}_i(u), D_i, \Gamma_i\right\}} \Big(I(D_i = t)\, H\left\{t, \bar{\mathbf{X}}_i(u), D_i, \Gamma_i\right\} + \left\{ I(C_i \leq D_i) + I(C_i > D_i)\, I(D_i > t) \right\} \cdot \left[dH\left\{t, \bar{\mathbf{X}}_i(u), D_i, \Gamma_i\right\} - d\Lambda(t)\, H\left\{t, \bar{\mathbf{X}}_i(u), D_i, \Gamma_i\right\} \right] \Big)$$

where $I(D_i = t)$ is used as shorthand for $\lim_{\Delta t \to 0} I(t \leq D_i < t + \Delta t)$, $\mu\left(u, \bar{\mathbf{X}}_i(u), D_i, \Gamma_i\right)$ is the cause-specific conditional hazard of MI given $\bar{\mathbf{X}}_i(u)$, D_i, and Γ_i,

Augmented IPWCC (3)

— This conditional expectation is equal to:

$$\frac{I(C_i > t)\, I(U_i^* > u)}{H\left\{u, \bar{\mathbf{X}}_i(u), D_i, \Gamma_i\right\}} \Big(I(D_i = t)\, H\left\{t, \bar{\mathbf{X}}_i(u), D_i, \Gamma_i\right\}$$
$$+ \{I(C_i \leq D_i) + I(C_i > D_i)\, I(D_i > t)\} \cdot$$
$$\left[dH\left\{t, \bar{\mathbf{X}}_i(u), D_i, \Gamma_i\right\} - d\Lambda(t)\, H\left\{t, \bar{\mathbf{X}}_i(u), D_i, \Gamma_i\right\}\right]\Big)$$

where $I(D_i = t)$ is used as shorthand for $\lim_{\Delta t \to 0} I(t \leq D_i < t + \Delta t)$, $\mu\left(u, \bar{\mathbf{X}}_i(u), D_i, \Gamma_i\right)$ is the cause-specific conditional hazard of MI given $\bar{\mathbf{X}}_i(u)$, D_i, and Γ_i,

Augmented IPWCC (4)

$$H\left\{u, \bar{\mathbf{X}}_i(u), D_i, \Gamma_i\right\} = \exp\left\{-\int_0^u \mu\left(r, \bar{\mathbf{X}}_i(r), D_i, \Gamma_i\right) dr\right\},$$

$$H\left\{t, \bar{\mathbf{X}}_i(u), D_i, \Gamma_i\right\} = \int_{\bar{\mathbf{x}} \in \bar{\mathcal{X}}(t)} H\left\{t, \bar{\mathbf{X}}_i(t) = \bar{\mathbf{x}}, D_i, \Gamma_i\right\} f_{\bar{\mathbf{X}}(t)|\bar{\mathbf{X}}(u),D,\Gamma}\left\{\bar{\mathbf{x}}, \bar{\mathbf{X}}(u), D_i, \Gamma_i\right\} d\bar{\mathbf{x}}$$

and

$$dH\left\{t, \bar{\mathbf{X}}_i(u), D_i, \Gamma_i\right\} = \lim_{\Delta t \to 0}\left[H\left\{t + \Delta t, \bar{\mathbf{X}}_i(u), D_i, \Gamma_i\right\} - H\left\{t, \bar{\mathbf{X}}_i(u), D_i, \Gamma_i\right\}\right].$$

Augmented IPWCC (5)

— These are substituted into the estimating equation, which can then be solved for $d\Lambda(t)$, leading to the AIPW estimator $\hat{\Lambda}^{\text{AIPW}}(t)$ (further ugly details omitted!).

Augmented IPWCC (5)

— These are substituted into the estimating equation, which can then be solved for $d\Lambda(t)$, leading to the AIPW estimator $\hat{\Lambda}^{\text{AIPW}}(t)$ (further ugly details omitted!).
— To make it feasible ($\hat{\Lambda}^{\text{f-AIPW}}(t)$) we need a coarsening model ($\mathcal{M}_{\text{CM}}$), but also a model ($\mathcal{M}_{\text{CSM}}$) for $\mu\left(u, \bar{\mathbf{X}}_i(u), D_i, \Gamma_i\right)$, the cause-specific conditional hazard of MI, and for the conditional density of the time-updated covariates $f_{\bar{\mathbf{x}}(t)|\bar{\mathbf{x}}(u),D,\Gamma}\left\{\bar{\mathbf{x}}, \bar{\mathbf{X}}(u), D_i, \Gamma_i\right\}$, $\mathcal{M}_{\text{TUCM}}$.

Augmented IPWCC (5)

— These are substituted into the estimating equation, which can then be solved for $d\Lambda(t)$, leading to the AIPW estimator $\hat{\Lambda}^{\text{AIPW}}(t)$ (further ugly details omitted!).

— To make it feasible ($\hat{\Lambda}^{\text{f-AIPW}}(t)$) we need a coarsening model ($\mathcal{M}_{\text{CM}}$), but also a model ($\mathcal{M}_{\text{CSM}}$) for $\mu\left(u, \bar{\mathbf{X}}_i(u), D_i, \Gamma_i\right)$, the cause-specific conditional hazard of MI, and for the conditional density of the time-updated covariates $f_{\bar{\mathbf{X}}(t)|\bar{\mathbf{X}}(u),D,\Gamma}\left\{\bar{\mathbf{x}}, \bar{\mathbf{X}}(u), D_i, \Gamma_i\right\}$, $\mathcal{M}_{\text{TUCM}}$.

— We can also extend it ($\hat{\Lambda}^{\text{f-AIPW-ext}}(t)$), by including (D_i, Γ_i) in the coarsening model ($\mathcal{M}_{\text{ECM}}$).

Outline

- Setting
- Notation
- Counting processes
- Semiparametric theory
- Estimators for the distribution of time to composite endpoint
  full data $\rightarrow$ complete cases $\rightarrow$ inverse probability weighted CC $\rightarrow$ augmented IPWCC
- Double robustness and semiparametric efficiency
- Variance estimation
- Simulation study
- Summary and further issues

Double robustness and semiparametric efficiency (1)

— It can be shown that $\hat{\Lambda}^{\text{f-AIPW}}(t)$ is semiparametric under

$$\mathcal{M}_{\text{CDW}} \cap \{\mathcal{M}_{\text{CM}} \cup (\mathcal{M}_{\text{CSM}} \cap \mathcal{M}_{\text{TUCM}})\}$$

Double robustness and semiparametric efficiency (1)

— It can be shown that $\hat{\Lambda}^{\text{f-AIPW}}(t)$ is semiparametric under

$$\mathcal{M}_{\text{CDW}} \cap \{\mathcal{M}_{\text{CM}} \cup (\mathcal{M}_{\text{CSM}} \cap \mathcal{M}_{\text{TUCM}})\}$$

— And that $\hat{\Lambda}^{\text{f-AIPW-ext}}(t)$ is semiparametric under

$$\mathcal{M}_{\text{CDDW}} \cap \{\mathcal{M}_{\text{ECM}} \cup (\mathcal{M}_{\text{CSM}} \cap \mathcal{M}_{\text{TUCM}})\}.$$

Double robustness and semiparametric efficiency (1)

— It can be shown that $\hat{\Lambda}^{\text{f-AIPW}}(t)$ is semiparametric under

$$\mathcal{M}_{\text{CDW}} \cap \{\mathcal{M}_{\text{CM}} \cup (\mathcal{M}_{\text{CSM}} \cap \mathcal{M}_{\text{TUCM}})\}$$

— And that $\hat{\Lambda}^{\text{f-AIPW-ext}}(t)$ is semiparametric under

$$\mathcal{M}_{\text{CDDW}} \cap \{\mathcal{M}_{\text{ECM}} \cup (\mathcal{M}_{\text{CSM}} \cap \mathcal{M}_{\text{TUCM}})\}.$$

— Furthermore, $\hat{\Lambda}^{\text{f-AIPW}}(t)$ is semiparametric efficient under

$$\mathcal{M}_{\text{CDW}} \cap \mathcal{M}_{\text{CM}} \cap \mathcal{M}_{\text{CSM}} \cap \mathcal{M}_{\text{TUCM}}$$

Double robustness and semiparametric efficiency (1)

— It can be shown that $\hat{\Lambda}^{\text{f-AIPW}}(t)$ is semiparametric under

$$\mathcal{M}_{\text{CDW}} \cap \{\mathcal{M}_{\text{CM}} \cup (\mathcal{M}_{\text{CSM}} \cap \mathcal{M}_{\text{TUCM}})\}$$

— And that $\hat{\Lambda}^{\text{f-AIPW-ext}}(t)$ is semiparametric under

$$\mathcal{M}_{\text{CDDW}} \cap \{\mathcal{M}_{\text{ECM}} \cup (\mathcal{M}_{\text{CSM}} \cap \mathcal{M}_{\text{TUCM}})\}.$$

— Furthermore, $\hat{\Lambda}^{\text{f-AIPW}}(t)$ is semiparametric efficient under

$$\mathcal{M}_{\text{CDW}} \cap \mathcal{M}_{\text{CM}} \cap \mathcal{M}_{\text{CSM}} \cap \mathcal{M}_{\text{TUCM}}$$

— and $\hat{\Lambda}^{\text{f-AIPW-ext}}(t)$ is semiparametric efficient under

$$\mathcal{M}_{\text{CDDW}} \cap \mathcal{M}_{\text{ECM}} \cap \mathcal{M}_{\text{CSM}} \cap \mathcal{M}_{\text{TUCM}}.$$

Double robustness and semiparametric efficiency (1)

— It can be shown that $\hat{\Lambda}^{\text{f-AIPW}}(t)$ is semiparametric under

$$\mathcal{M}_{\text{CDW}} \cap \{\mathcal{M}_{\text{CM}} \cup (\mathcal{M}_{\text{CSM}} \cap \mathcal{M}_{\text{TUCM}})\}$$

— And that $\hat{\Lambda}^{\text{f-AIPW-ext}}(t)$ is semiparametric under

$$\mathcal{M}_{\text{CDDW}} \cap \{\mathcal{M}_{\text{ECM}} \cup (\mathcal{M}_{\text{CSM}} \cap \mathcal{M}_{\text{TUCM}})\}.$$

— Furthermore, $\hat{\Lambda}^{\text{f-AIPW}}(t)$ is semiparametric efficient under

$$\mathcal{M}_{\text{CDW}} \cap \mathcal{M}_{\text{CM}} \cap \mathcal{M}_{\text{CSM}} \cap \mathcal{M}_{\text{TUCM}}$$

— and $\hat{\Lambda}^{\text{f-AIPW-ext}}(t)$ is semiparametric efficient under

$$\mathcal{M}_{\text{CDDW}} \cap \mathcal{M}_{\text{ECM}} \cap \mathcal{M}_{\text{CSM}} \cap \mathcal{M}_{\text{TUCM}}.$$

— If we correctly specify either the coarsening model or both the cause-specific and time-updated covariates models (or all three), then the AIPW estimator will be consistent.

Double robustness and semiparametric efficiency (1)

— It can be shown that $\hat{\Lambda}^{\text{f-AIPW}}(t)$ is semiparametric under

$$\mathcal{M}_{\text{CDW}} \cap \{\mathcal{M}_{\text{CM}} \cup (\mathcal{M}_{\text{CSM}} \cap \mathcal{M}_{\text{TUCM}})\}$$

— And that $\hat{\Lambda}^{\text{f-AIPW-ext}}(t)$ is semiparametric under

$$\mathcal{M}_{\text{CDDW}} \cap \{\mathcal{M}_{\text{ECM}} \cup (\mathcal{M}_{\text{CSM}} \cap \mathcal{M}_{\text{TUCM}})\}.$$

— Furthermore, $\hat{\Lambda}^{\text{f-AIPW}}(t)$ is semiparametric efficient under

$$\mathcal{M}_{\text{CDW}} \cap \mathcal{M}_{\text{CM}} \cap \mathcal{M}_{\text{CSM}} \cap \mathcal{M}_{\text{TUCM}}$$

— and $\hat{\Lambda}^{\text{f-AIPW-ext}}(t)$ is semiparametric efficient under

$$\mathcal{M}_{\text{CDDW}} \cap \mathcal{M}_{\text{ECM}} \cap \mathcal{M}_{\text{CSM}} \cap \mathcal{M}_{\text{TUCM}}.$$

— If we correctly specify either the coarsening model or both the cause-specific and time-updated covariates models (or all three), then the AIPW estimator will be consistent.

— This property is known as double robustness.

— Double robustness is especially important in our setting, where it is probably unrealistic to hope that the cause-specific and time-updated covariates models are correctly specified.

Double robustness and semiparametric efficiency (2)

— Double robustness is especially important in our setting, where it is probably unrealistic to hope that the cause-specific and time-updated covariates models are correctly specified.
— Under only the assumption that the coarsening model is correctly specified, the f-AIPW estimator is consistent.

Double robustness and semiparametric efficiency (2)

— Double robustness is especially important in our setting, where it is probably unrealistic to hope that the cause-specific and time-updated covariates models are correctly specified.
— Under only the assumption that the coarsening model is correctly specified, the f-AIPW estimator is consistent.
— Furthermore, if we correctly specify all three models, then the f-AIPW estimator is optimally efficient.

Double robustness and semiparametric efficiency (2)

— Double robustness is especially important in our setting, where it is probably unrealistic to hope that the cause-specific and time-updated covariates models are correctly specified.
— Under only the assumption that the coarsening model is correctly specified, the f-AIPW estimator is consistent.
— Furthermore, if we correctly specify all three models, then the f-AIPW estimator is optimally efficient.
— In practice, when the cause-specific and time-updated covariates models are not correctly specified, experience suggests that augmentation will lead to efficiency gains as long as model misspecification is not too severe.

Outline

- Setting
- Notation
- Counting processes
- Semiparametric theory
- Estimators for the distribution of time to composite endpoint
  full data $\rightarrow$ complete cases $\rightarrow$ inverse probability weighted CC $\rightarrow$ augmented IPWCC
- Double robustness and semiparametric efficiency
- Variance estimation
- Simulation study
- Summary and further issues

— Another beauty of semiparametric theory is that it gives an automatic strategy for deriving variance estimators.

Variance estimation

— Another beauty of semiparametric theory is that it gives an automatic strategy for deriving variance estimators.
— The asymptotic variance of the estimator is equal to the variance of its influence function.

Variance estimation

— Another beauty of semiparametric theory is that it gives an automatic strategy for deriving variance estimators.

— The asymptotic variance of the estimator is equal to the variance of its influence function.

— See paper for details.

Outline

- Setting
- Notation
- Counting processes
- Semiparametric theory
- Estimators for the distribution of time to composite endpoint
  full data → complete cases → inverse probability weighted CC → augmented IPWCC
- Double robustness and semiparametric efficiency
- Variance estimation
- Simulation study
- Summary and further issues

— 1000 datasets, sample size = 100.
— One binary baseline covariate, X, with $Pr(X = 1) = 0.5$. No time-updated covariates.

— 1000 datasets, sample size = 100.
— One binary baseline covariate, X, with $Pr(X = 1) = 0.5$. No time-updated covariates.
— Subjects enter the study uniformly at random over 2 years, administrative censoring at 5 years.

Simulation study
Design

— 1000 datasets, sample size = 100.
— One binary baseline covariate, X, with $Pr(X = 1) = 0.5$. No time-updated covariates.
— Subjects enter the study uniformly at random over 2 years, administrative censoring at 5 years.
— Conditional on X, time to MI is Weibull with shape 0.5 and scale $\{10 \exp(1.5 - 3X)\}$.

— 1000 datasets, sample size = 100.
— One binary baseline covariate, X, with $Pr(X = 1) = 0.5$. No time-updated covariates.
— Subjects enter the study uniformly at random over 2 years, administrative censoring at 5 years.
— Conditional on X, time to MI is Weibull with shape 0.5 and scale $\{10 \exp(1.5 - 3X)\}$.
— Conditional on X, time to death is exponential with hazard $0.24 \exp(-1.5 + 3X)$. This time to death is compared with time to MI. If MI occurs first then the time to death is discarded, and the time to death is re-generated as the MI time plus a draw from exponential with hazard $0.6 \exp(-1.5 + 3X)$.

— 1000 datasets, sample size = 100.

— One binary baseline covariate, X, with $Pr(X = 1) = 0.5$. No time-updated covariates.

— Subjects enter the study uniformly at random over 2 years, administrative censoring at 5 years.

— Conditional on X, time to MI is Weibull with shape 0.5 and scale $\{10\exp(1.5 - 3X)\}$.

— Conditional on X, time to death is exponential with hazard $0.24\exp(-1.5 + 3X)$. This time to death is compared with time to MI. If MI occurs first then the time to death is discarded, and the time to death is re-generated as the MI time plus a draw from exponential with hazard $0.6\exp(-1.5 + 3X)$.

— Conditional on X, withdrawal is exponential with hazard $\exp(-0.5 + X)$.

We compare five estimators of the survivor distribution:

1. the full data estimator, $\hat{S}^{\text{full}}(t)$.
2. the complete case estimator, $\hat{S}^{\text{CC}}(t)$.
3. the IPWCC estimator, $\hat{S}^{\text{f-IPWCC}}(t)$, with only X used to predict the weights using Cox PH model. This coarsening model is correctly specified.
4. the IPWCC estimator, $\hat{S}^{\text{f-IPWCC-ext}}(t)$, with X and (D, Γ) used to predict the weights using a Cox PH model. Correctly specified but more elaborate than necessary.
5. the AIPW estimator, $\hat{S}^{\text{f-AIPW-ext}}(t)$. X and (D, Γ) used in the model for the weights, and for the cause-specific MI model in a Cox PH model. This CS model is not correctly specified.

Simulation study

Results: $S(0.5)$

Estimator of survivor function	Mean	SE	% increase in SE compared with full data	Coverage of 95% CI
full	0.622	0.0464		
CC	0.631	0.0479	3.3%	
f-IPWCC	0.622	0.0486	4.9%	
f-IPWCC-ext	0.622	0.0482	4.0%	
f-AIPW-ext	0.622	0.0475	2.3%	94.7%

Estimator of survivor function	Mean	SE	% increase in SE compared with full data	Coverage of 95% CI
full	0.430	0.0485		
CC	0.463	0.0536	10.4%	
f-IPWCC	0.432	0.0535	10.2%	
f-IPWCC-ext	0.432	0.0524	7.9%	
f-AIPW-ext	0.430	0.0513	5.6%	95.9%

Simulation study
Results: $S(2.5)$

Estimator of survivor function	Mean	SE	% increase in SE compared with full data	Coverage of 95% CI
full	0.359	0.0483		
CC	0.401	0.0564	16.7%	
f-IPWCC	0.363	0.0548	13.3%	
f-IPWCC-ext	0.363	0.0541	11.9%	
f-AIPW-ext	0.360	0.0512	6.0%	96.7%

Estimator of survivor function	Mean	SE	% increase in SE compared with full data	Coverage of 95% CI
full	0.318	0.0480		
CC	0.362	0.0596	24.3%	
f-IPWCC	0.324	0.0569	18.6%	
f-IPWCC-ext	0.326	0.0556	16.0%	
f-AIPW-ext	0.320	0.0522	8.9%	95.6%

Estimator of survivor function	Mean	SE	% increase in SE compared with full data	Coverage of 95% CI
full	0.288	0.0495		
CC	0.332	0.0689	39.1%	
f-IPWCC	0.297	0.0640	30.0%	
f-IPWCC-ext	0.301	0.0619	25.0%	
f-AIPW-ext	0.289	0.0573	15.8%	93.8%

Outline

- Setting
- Notation
- Counting processes
- Semiparametric theory
- Estimators for the distribution of time to composite endpoint
  full data $\rightarrow$ complete cases $\rightarrow$ inverse probability weighted CC $\rightarrow$ augmented IPWCC
- Double robustness and semiparametric efficiency
- Variance estimation
- Simulation study
- Summary and further issues

Summary

— Semiparametric theory of augmented inverse probability weighted estimating equations is powerful, especially for complex, ‘non-ideal’ settings.

Summary

— Semiparametric theory of augmented inverse probability weighted estimating equations is powerful, especially for complex, ‘non-ideal’ settings.

— We’ve shown how partial information on components of a composite endpoint can be incorporated into the estimation of the time to composite endpoint in a principled way, when other components of the composite endpoint are not observed due to withdrawal.

Summary

— Semiparametric theory of augmented inverse probability weighted estimating equations is powerful, especially for complex, ‘non-ideal’ settings.
— We’ve shown how partial information on components of a composite endpoint can be incorporated into the estimation of the time to composite endpoint in a principled way, when other components of the composite endpoint are not observed due to withdrawal.
— An appeal of this approach is that, although further models are required (for the cause-specific hazard of the incompletely-observed event, and for the evolution of the time-updated covariate process, if this is to be modelled), the consistency of our estimator does not rely on having correctly specified these models.

Summary

— Semiparametric theory of augmented inverse probability weighted estimating equations is powerful, especially for complex, 'non-ideal' settings.

— We've shown how partial information on components of a composite endpoint can be incorporated into the estimation of the time to composite endpoint in a principled way, when other components of the composite endpoint are not observed due to withdrawal.

— An appeal of this approach is that, although further models are required (for the cause-specific hazard of the incompletely-observed event, and for the evolution of the time-updated covariate process, if this is to be modelled), the consistency of our estimator does not rely on having correctly specified these models.

— Efficiency gains are guaranteed if the additional models are correctly specified, and typically will be seen even if this is not the case.

Summary

— Semiparametric theory of augmented inverse probability weighted estimating equations is powerful, especially for complex, ‘non-ideal’ settings.
— We’ve shown how partial information on components of a composite endpoint can be incorporated into the estimation of the time to composite endpoint in a principled way, when other components of the composite endpoint are not observed due to withdrawal.
— An appeal of this approach is that, although further models are required (for the cause-specific hazard of the incompletely-observed event, and for the evolution of the time-updated covariate process, if this is to be modelled), the consistency of our estimator does not rely on having correctly specified these models.
— Efficiency gains are guaranteed if the additional models are correctly specified, and typically will be seen even if this is not the case.
— In simulations, AIPW seen to recover up to 50% of the efficiency lost through withdrawal in standard approaches.

Further issues

— Although the approach can deal in theory with time-updated covariates, in practice incorporating these into the cause-specific model for the incompletely-observed event will be problematic, since further models are required, along with the calculation of a typically intractable integral.

Further issues

— Although the approach can deal in theory with time-updated covariates, in practice incorporating these into the cause-specific model for the incompletely-observed event will be problematic, since further models are required, along with the calculation of a typically intractable integral.
— A pragmatic solution would be to omit time-updated covariates from the cause-specific model, but further work is required to understand the sacrifice involved in doing so.

Further issues

— Although the approach can deal in theory with time-updated covariates, in practice incorporating these into the cause-specific model for the incompletely-observed event will be problematic, since further models are required, along with the calculation of a typically intractable integral.

— A pragmatic solution would be to omit time-updated covariates from the cause-specific model, but further work is required to understand the sacrifice involved in doing so.

— Future work: the comparison of the distributions of time to composite endpoint in two independent groups, via a weighted log-rank test.