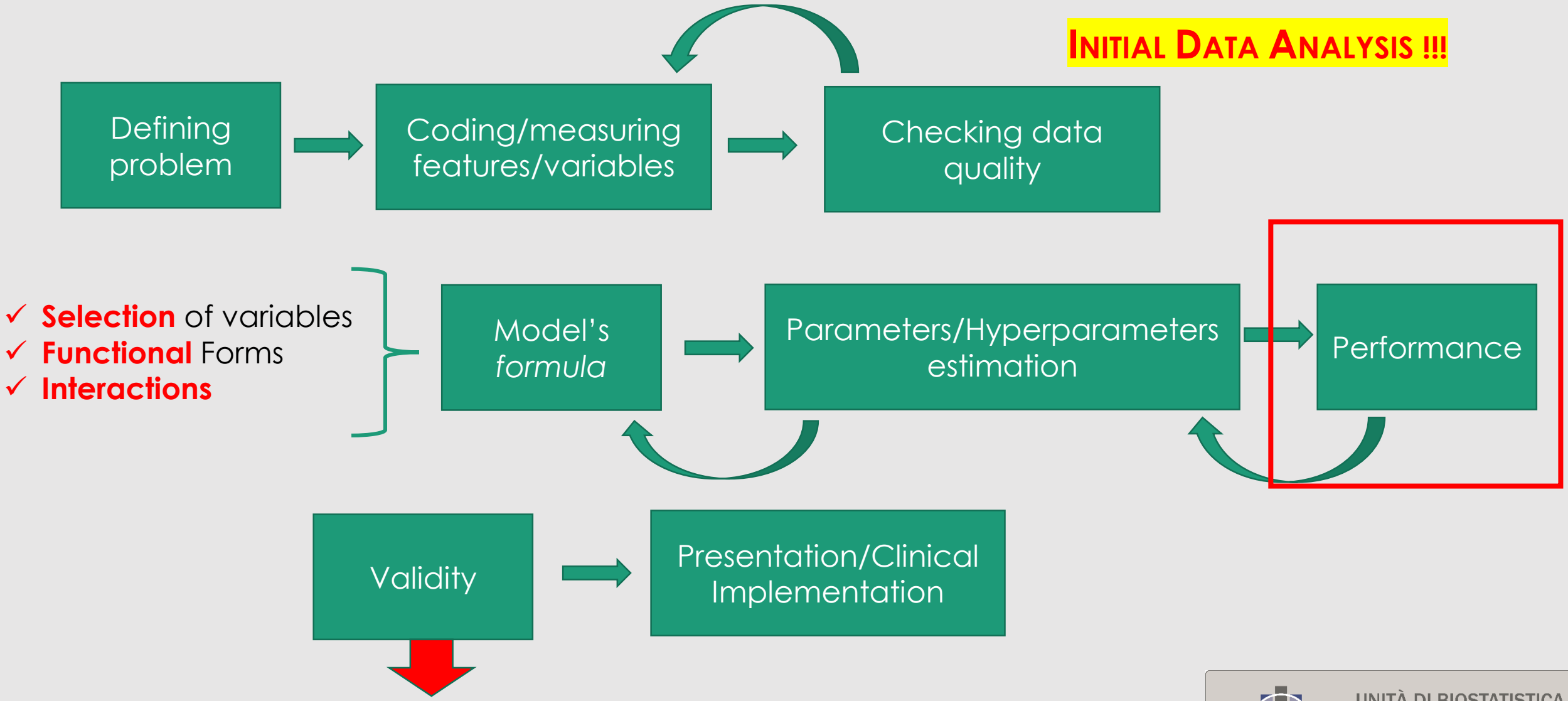
Survival analysis

- Evaluating the **performance** of a survival model
- A note about **competing** risks
- **Bias** in Survival

UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute

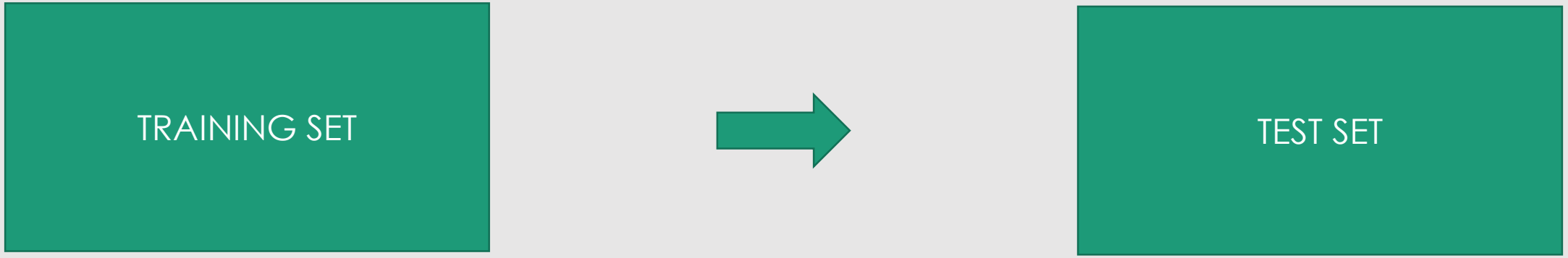
Some steps should be considered in developing prediction models:



UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute

Measures of the accuracy of predictions

Are our predictions **reliable**?



Calibration: does the model predict *accurately*?
calibration **slope**, 1 : perfect calibration

Discrimination: does the model *discriminate* well?
C statistic (**AUC**ROC), 1: perfect discrimination, 0.5 : flipping a coin

Survival model discrimination: *time dependent* ROC curves

The standard approach of ROC curve considers a binary event (disease) status and marker value for an individual **as fixed over time**.

In survival setting, individuals who are disease-free earlier may develop the disease later along the study follow-up.

Thus, an ROC curve **as a function of time** is more appropriate.

M_i : risk score for individual i, (i = 1, …, n), for example from the Cox model: $x_i\beta$

$D_i(t)$: disease status at time t, taking values 1 or 0

For a given threshold c, the time-dependent sensitivity and specificity can be defined respectively by:

$$Sens(c,t) = P(M_i > c | D_i(t) = 1)$$

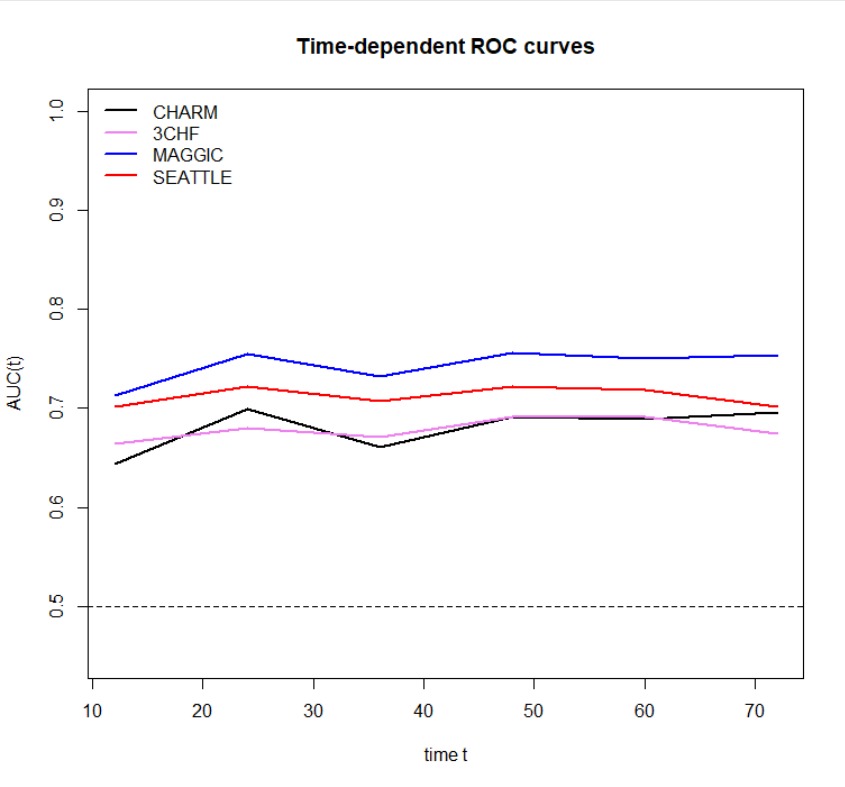
$$Spec(c,t) = P(M_i \leq c | D_i(t) = 0)$$

UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute

The corresponding ROC curve for any time *t* plots *Sens(c,t)* against *1-Spec(c,t)* for thresholds c and the time-dependent AUC is defined by:

$$AUC(t) = \int Sens(c,t)d[1 - Spec(c,t)] \qquad AUC(t) = P(M_i > M_j | T_i \leq t, T_j > t)$$

The AUC(t) is the probability that the estimated risk scores ***from a randomly selected pair*** of diseased and non-diseased individuals at time *t* are correctly relatively ordered.



1. Cases: subjects who experience the event **before** time *t* and controls those who remain event-free through time *t* ***[cumulative/dynamic]***.

2. Cases: subjects who experience an event **at** time *t* ; controls can be compared to incident cases and are subjects with *T>t* ***[incident/dynamic]***

Assessment of survival model calibration

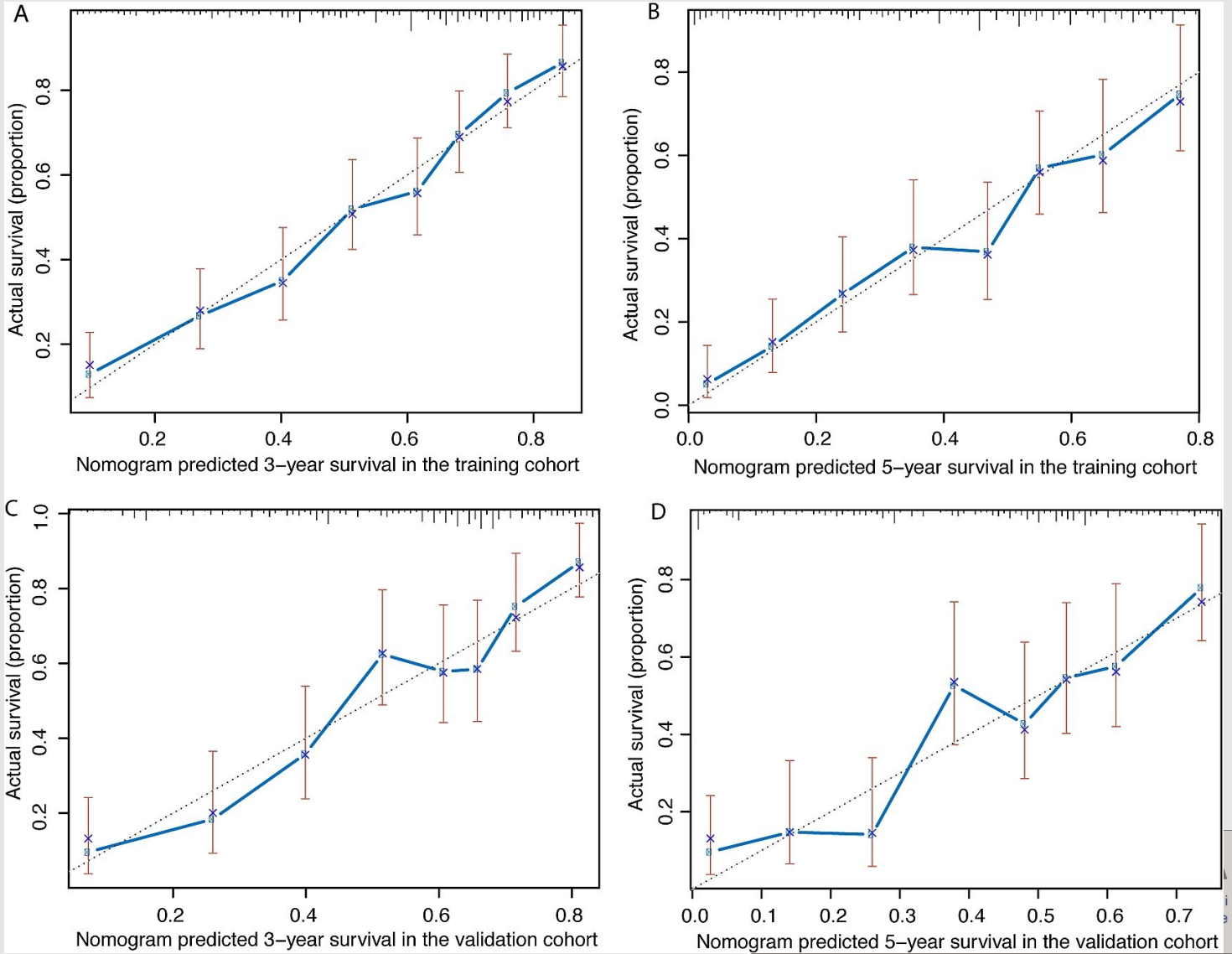
In the context of survival analysis, calibration refers to the ***agreement*** between predicted probabilities and observed event rates or frequencies of the outcome ***within a given duration of time***.

1. Subjects are divided into ***strata*** based on the predicted probability of the occurrence of the event **by time t**.

2. Within each stratum, the mean predicted probability of the occurrence of the event **by time t** is computed.

3. Then, within each stratum, the observed probability of the event **by time t** is computed by the Kaplan-Meier estimator for the subjects in that stratum.

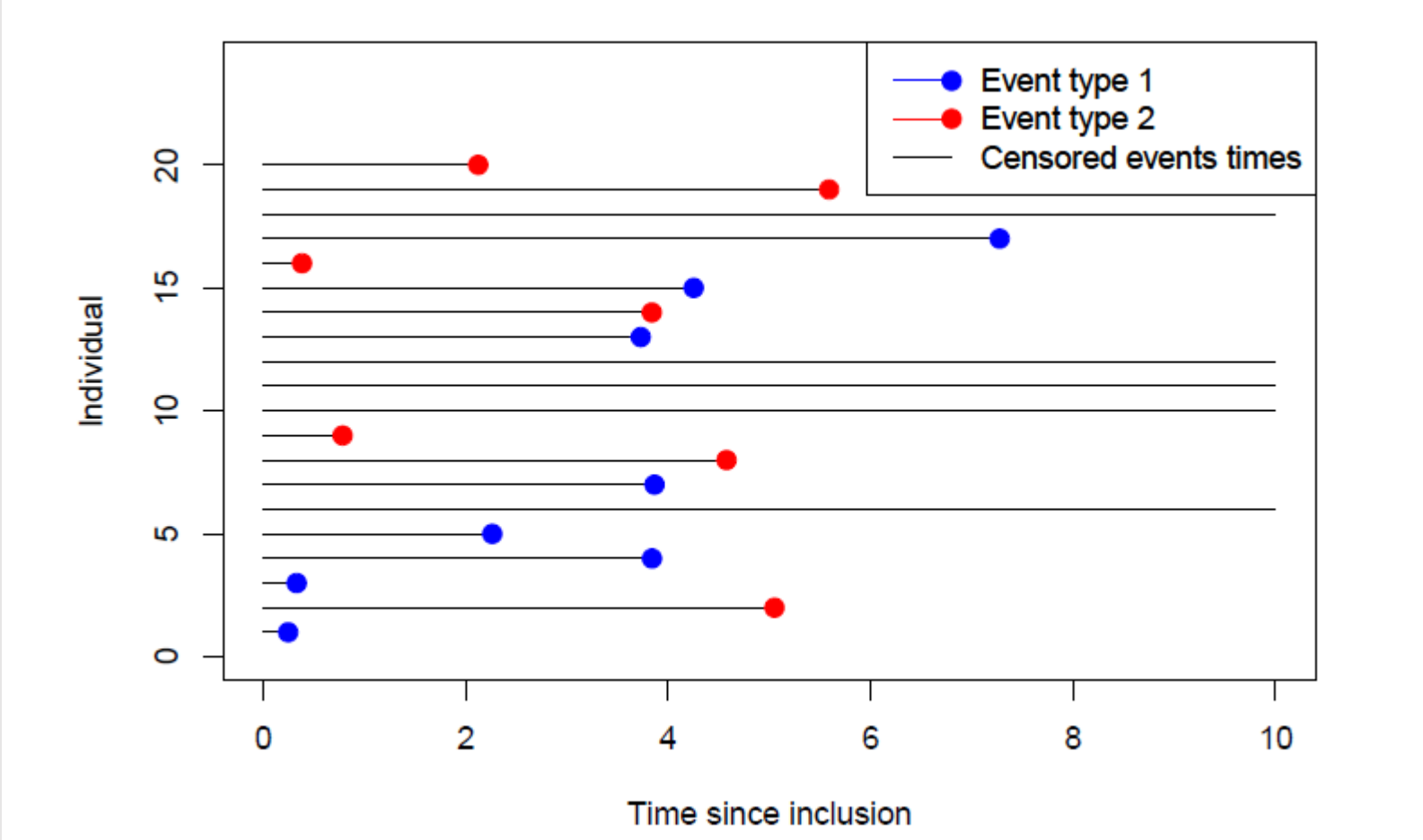
4. The mean predicted and observed probabilities can then be **compared** across strata.



Initial checklist

- **Target population:** who would be eligible to use the model and whatever inclusion/exclusion criteria
- **Time origin:** baseline *time zero* (if there is time involved!)
- **Target of prediction:** event/parameter of interest
- ***Competing risks*** events *after which* the event of interest cannot occur or is not of interest any longer
- **Prediction time horizon:** how far in time from the baseline the prediction is projected (if there is time involved!)
- **Predictor/Prognostic variables:** list of the predictors/features [*measured at baseline*] *(how they were measured / context !)*

Competing risks



- Cancer specific death (with the competing event of death from other causes)
- Return-to-work after traumatic injury (with the competing event of death)

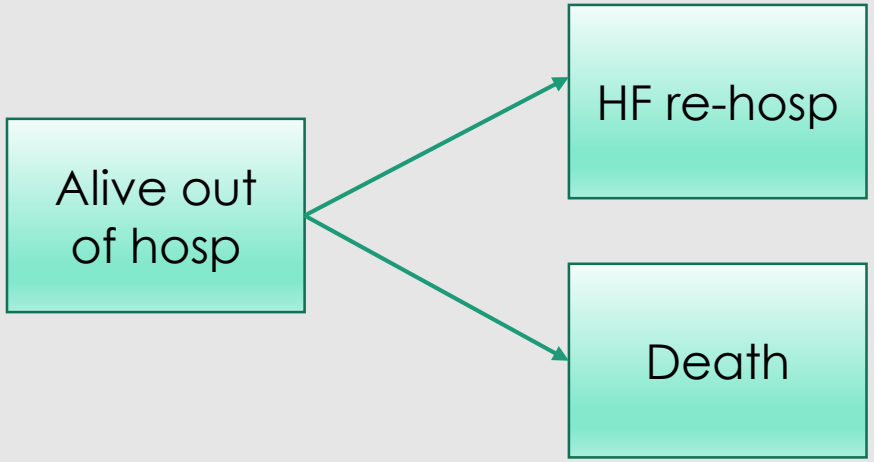
UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute

Clinical research studies often record the time to more than one "first" outcome:

Examples: death, cardiovascular disease (CVD), end stage renal disease (ESRD)

Situations with more than one possible type of event for each subject may be generally described by **multistate models (that also allow for *recurrent* events)**.

The simplest example of a multistate model is the one of competing risks:



A **competing event** is one that precludes* the occurrence of the event of interest:

After transplant or death, patient is no longer at risk for *primary* outcome of interest (ESRD or CVD)

Competing risks arise from **different causes of failure** that are considered as competing events.

We may model competing risks by a process with one ***transient*** state 0, corresponding to *alive and free of the event* and *k* ***absorbing*** states, say corresponding to *event by cause h* (*h*= 1,. . . ,*k*)

* Note that this is different from censoring, which (only) make the event of interest impossible to *observe*…

Remind: the distinctive feature of survival data is **censoring**

Consider one event of interest: death.

Time to the event is **censored** for subjects still alive at the end of their observation.

T = survival time
C =censoring time
Tobs =min(*T*,*C*)

➡ $\delta_i = I(T_i \leq C_i)$ ➡ For each subject *i* we observe a pair of values (T_i, δ_i)

Basic assumption: **non-informative** censoring. At any given point in time subjects who remain **have the same future risk** for the event of interest as **censored** subjects

If a patient **experiences a competing event**, standard survival analysis methods would treat that patient as **censored** for the outcome of interest (e.g., ESRD or CVD).

? Why this could be a problem ?

Particularly in the presence of **strong*** competing risks, as with frail or elderly populations, **standard** survival predictions may substantially **overestimate** the **absolute risk** of the event of interest because subjects with a competing (and thus censored) event are treated ***as if they could experience the event of interest*** *[in the event-time-interval around the censoring date]*.

Such predictions have been said to refer to the risk of failing from the event of interest *in a virtual world where the competing risk is absent*.

The use of the (1-)Kaplan-Meier survival function results in estimates of incidence that are **biased upward**, regardless of whether the competing events are independent of one another (in any case an assumption untestable from the data).

* i.e. high incidence

Therefore, the correct cumulative incidence function for the *k-th* cause is defined as:

$$CIF_k(t) = P(T_k \leq t)$$

as the probability of experiencing the *k-th* event before time t **and before the occurrence** of a different type of event.

<u>Property:</u>

- the sum of the CIF of each of the **individual outcomes** will equal the CIF estimates of the incidence of the **composite outcome** consisting of all of the competing events.

$$S(t) = P(T > t) = 1 - \sum_{k=1}^{J} CIF_k(t)$$

Note that $CIF_k(t)$ **is different** from $1 - KM_k(t)$:

$$CIF_k(t) = P(T \leq t, cause = k) = \int_0^t S(u)h_k(u)du$$

specific hazard for the k-th event of interest

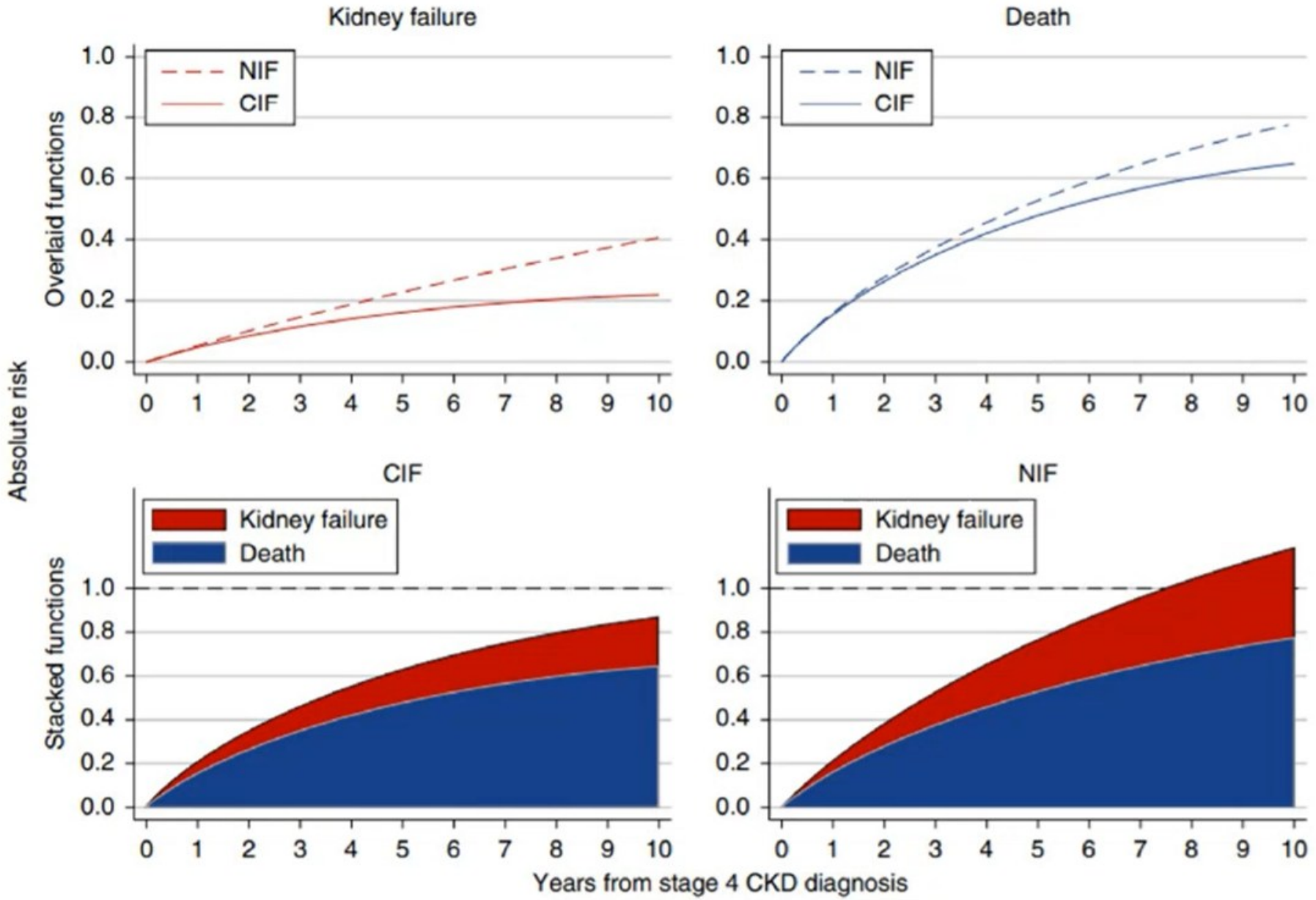
(Aalen-Johansen estimator)

not having failed **from *any other event*** **before**

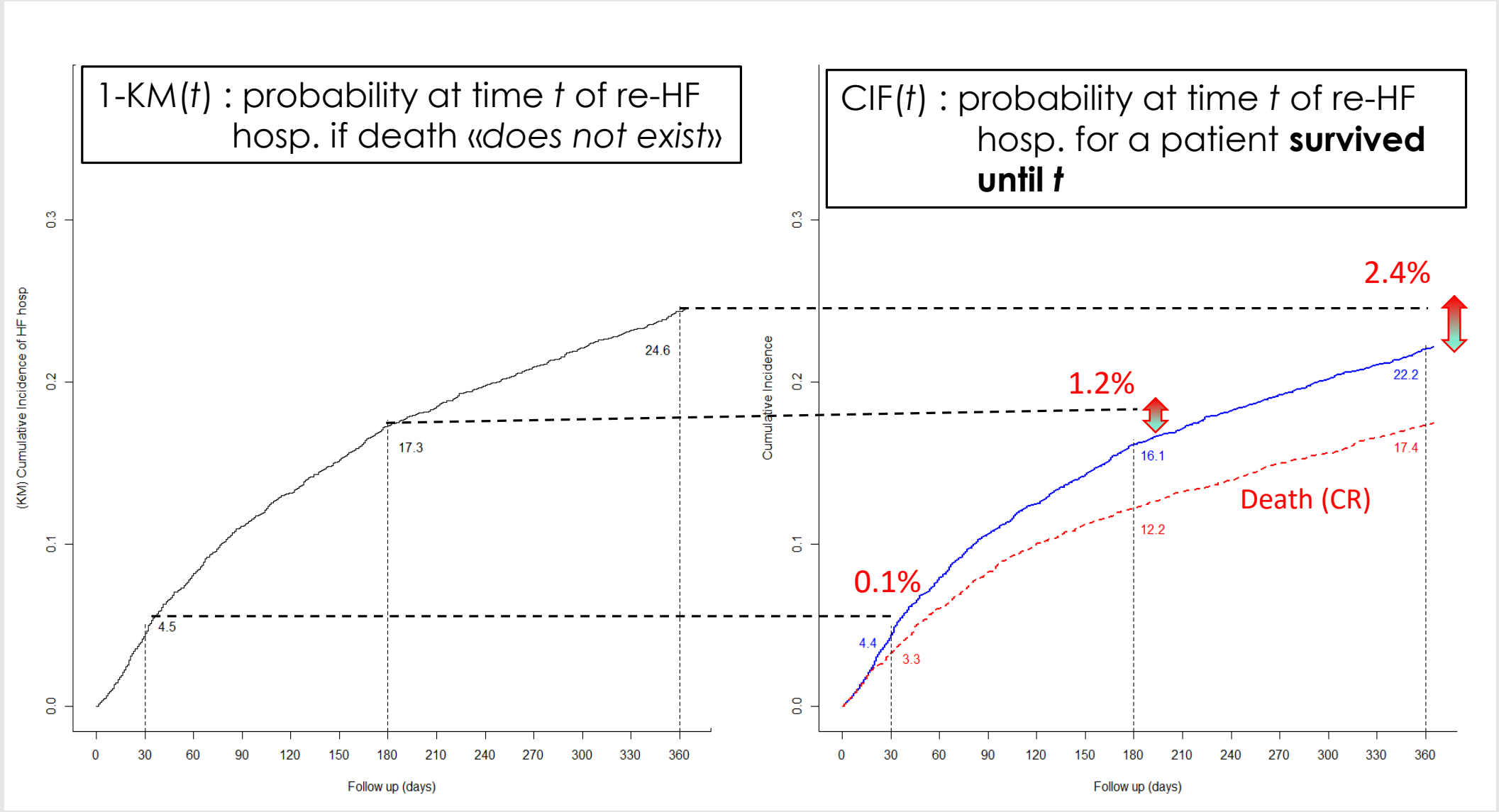
$$1 - KM_k(t) = \int_0^t S(u)_k h_k(u)du$$

Survival estimate **censoring** pts that experience competing risks

If we compute 1-$KM_k(t)$ ***for each event of interest***, the sum of the 1-$KM_k(t)$ **will exceed** that of the KM of the composite end point…


Kidney failure
Death
NIF
CIF
Overlaid functions
Absolute risk
CIF
NIF
Kidney failure
Death
Stacked functions
Years from stage 4 CKD diagnosis
NIF= Naïve Incidence Function
CIF = Comulative Incidence Function

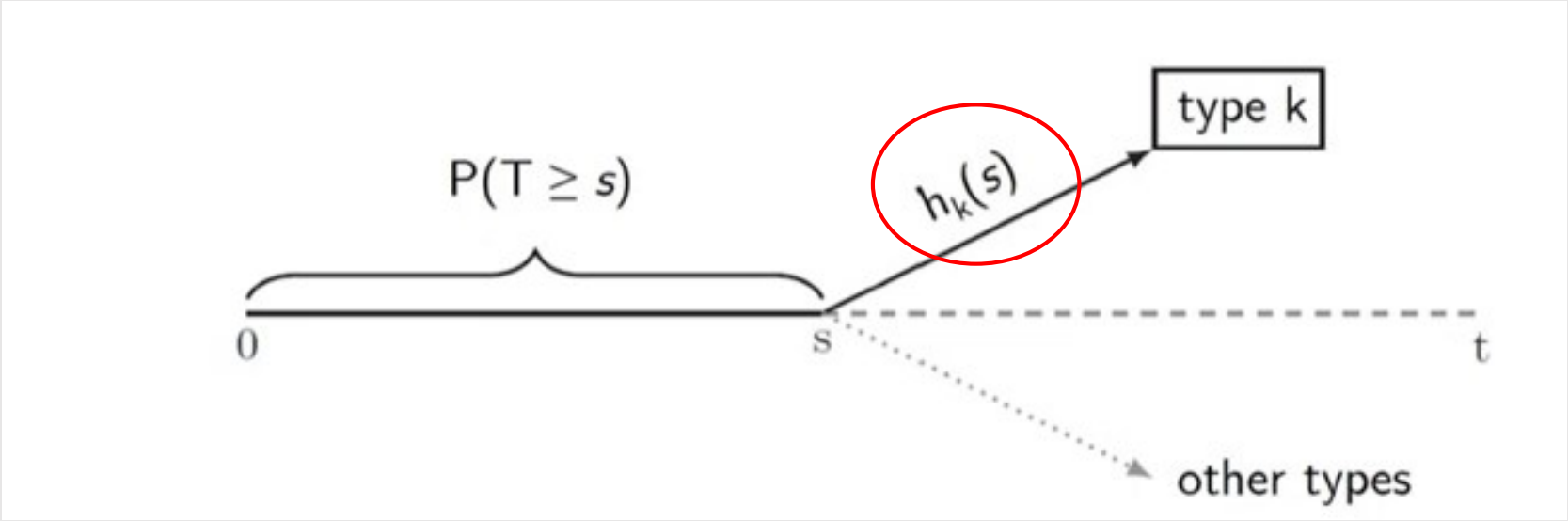
Real data set: death is the ***competing risk*** for re-HF hospitalization



Cause specific hazard function

$$h_j(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} P(t \leq T < t + \Delta t, Y = j | T \geq t)$$

the rate of (only) events by cause j, in small time intervals $t + \Delta t$, **among those who have not yet died by any cause**



Cause specific hazard regression models

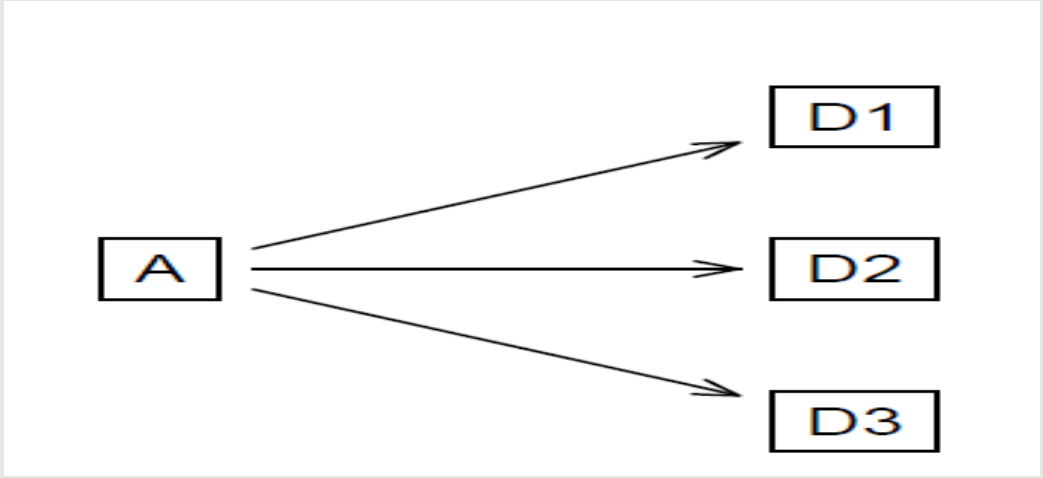
Quite common to fit **Cox models for cause specific hazards:**

$$h_j(t|X) = exp(\boldsymbol{\beta X}) h_{0j}(t)$$

↓

baseline hazard for the jth cause

To go from the **J cause specific hazards** to the **cumulative incidence** an approach formalized in multi-state models is used :



Separate Cox cause-specific models for each transition and then ***probabilities to be in each state (CIF)*** could be computed (to recover the estimated cumulative incidence over time)…

SUMMARY

Ignoring the competing risk **overestimates** the probability of the event of interest.

This problem is specifically related to the **cumulative** risk, not to the **cause-specific hazard rate**.

The competing risk issue comes about when you want to address the **cumulative risk** of a particular event, in which case you probably want to look at the cumulative risks ***of all types*** of events.

When fitting **regression models** in the presence of competing risks, researchers can choose from different families of models:

- modeling the effect of covariates on the **cause-specific hazard** of the outcome
  estimate the effect of the covariates **on the rate of occurrence** of the outcome in those subjects who are currently event free

- modeling the effect of covariates directly on the **cumulative incidence function**
  estimate the effect of covariates **on the absolute risk** of the outcome over time (**Fine and Gray**)

- **multi-state models** : cause-specific transitions + cumulative incidence estimates

Initial checklist

- **Target population:** who would be eligible to use the model and whatever inclusion/exclusion criteria
- **Time origin:** baseline *time zero* (if there is time involved!)
- **Target of prediction:** event/parameter of interest
- **Competing risks** events *after which* the event of interest cannot occur or is not of interest any longer
- **Prediction time horizon:** how far in time from the baseline the prediction is projected (if there is time involved!)
- **Predictor/Prognostic variables:** list of the predictors/features [***measured at baseline?***] *(how they were measured / context !)*

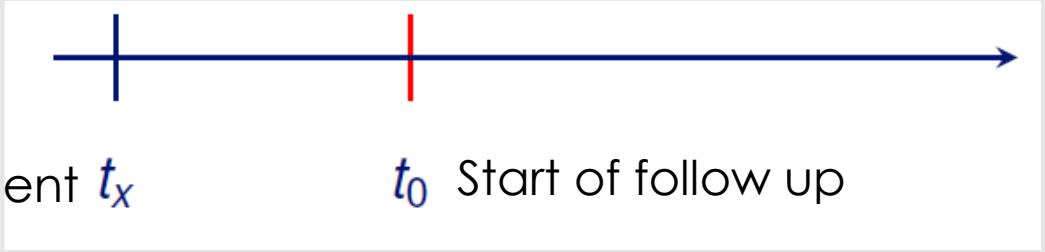
Bias in the survival setting

When using non-experimental data to carry out **causal** investigations, several potential sources of **bias** arise, in particular:

(a) **Selection bias** Do the data capture the target population?

(b) **Immortal time bias** Is exposure status assigned correctly?

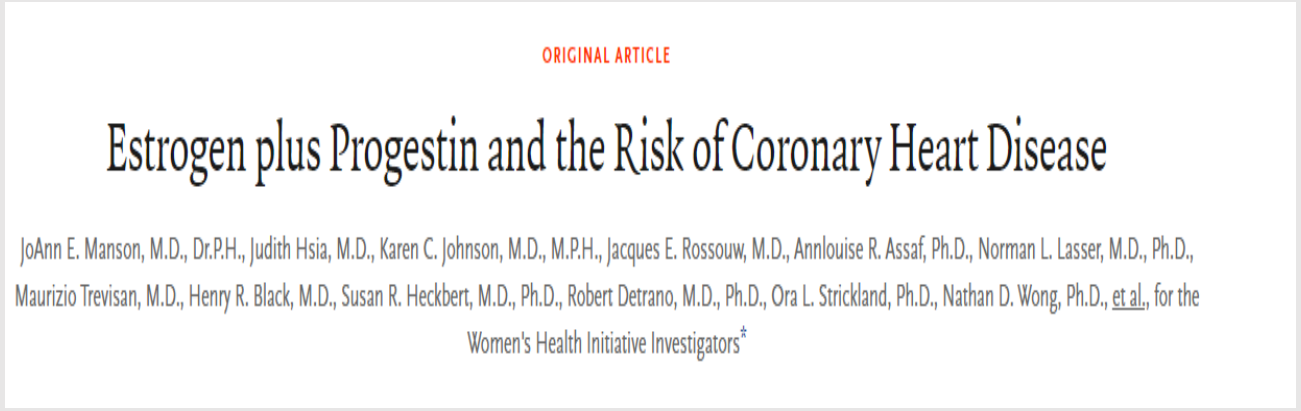
Objective: to compare treated vs. not treated



selection bias
study population
≠target population

It may be the case that individuals who have longer history of treatment (i.e. larger (t_0 - t_x)) **are very different** from those with a short history…(for example, higher probability to include long-term users)

Selection Bias

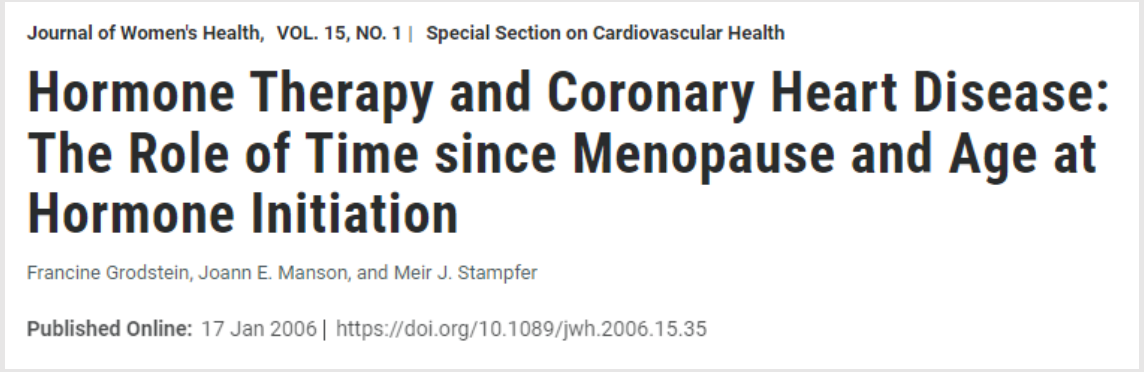


ORIGINAL ARTICLE

Estrogen plus Progestin and the Risk of Coronary Heart Disease

JoAnn E. Manson, M.D., Dr.P.H., Judith Hsia, M.D., Karen C. Johnson, M.D., M.P.H., Jacques E. Rossouw, M.D., Annlouise R. Assaf, Ph.D., Norman L. Lasser, M.D., Ph.D., Maurizio Trevisan, M.D., Henry R. Black, M.D., Susan R. Heckbert, M.D., Ph.D., Robert Detrano, M.D., Ph.D., Ora L. Strickland, Ph.D., Nathan D. Wong, Ph.D., et al., for the Women's Health Initiative Investigators*

2003: RCT, Women's Health Initiative: ITT of initiators compared with non-initiators: HR=1.24.



Journal of Women's Health, VOL. 15, NO. 1 | Special Section on Cardiovascular Health

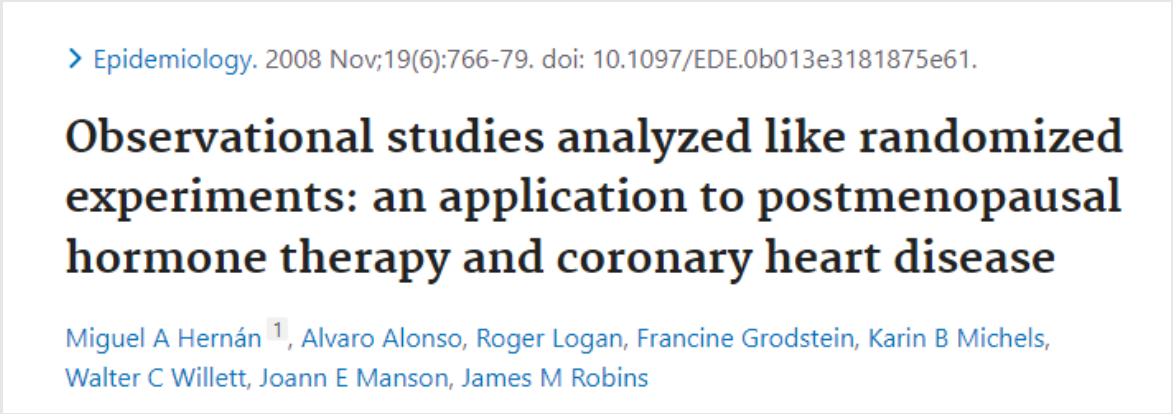
Hormone Therapy and Coronary Heart Disease: The Role of Time since Menopause and Age at Hormone Initiation

Francine Grodstein, Joann E. Manson, and Meir J. Stampfer

Published Online: 17 Jan 2006 | https://doi.org/10.1089/jwh.2006.15.35

2006: Observational study, Nurses Health Study current HRT users compared with never users: HR=0.68.

Was the discrepancy due to unmeasured confounding?



Epidemiology. 2008 Nov;19(6):766-79. doi: 10.1097/EDE.0b013e3181875e61.

Observational studies analyzed like randomized experiments: an application to postmenopausal hormone therapy and coronary heart disease

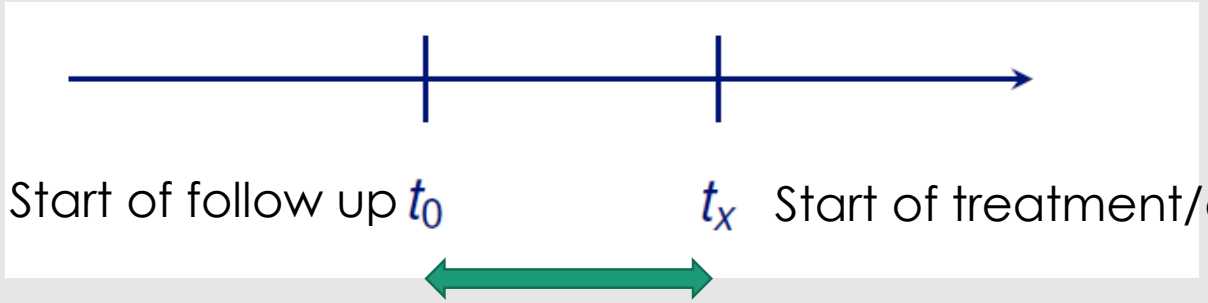
Miguel A Hernán [1], Alvaro Alonso, Roger Logan, Francine Grodstein, Karin B Michels, Walter C Willett, Joann E Manson, James M Robins

2008: New analysis of the Nurses Health Study initiators compared with non-initiators: HR=1.20

The 2006 observational study did not **target** the same population, nor causal effect, as the RCT

Immortal time Bias

Immortal time bias !

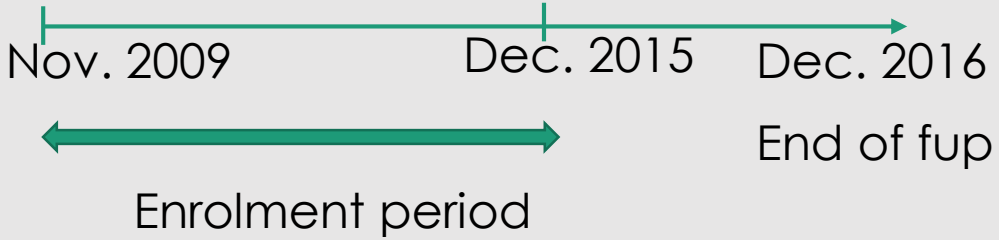


If this information is used to assign individuals as treated/not treated from t_0, those who are seen to have received treatment have a guaranteed survival time of at least t_x .

This makes them immortal for a chunk of their follow-up time (Suissa, 2007)

These errors do not usually arise in RCTs since protocols well define: population, treatment, follow-up, etc.

Suissa. Immortal time bias in observational studies of drug effects. Pharmacoepidemiol Drug Saf 2007; 241–9



ELSEVIER

International Journal of Cardiology

Volume 277, 15 February 2019, Pages 140-146

HF progression among outpatients with HF in a community setting ☆

Annamaria Iorio [a, b], Federico Rea [c, d], Giulia Barbati [e, c], Arjuna Scagnetto [e], Elena Peruzzi [f], Agnese Garavaglia [f], Giovanni Corrao [c, d], Gianfranco Sinagra [b], Andrea Di Lenarda [g]

HF (Heart Failure) patients enrolled from 2009 to 2015.

HF progression: **worsening** of the disease along follow up («exposure»).

Impact of HF progression on mortality?

Index visit (**baseline**): first evaluation with available LVEF (left ventricular ejection fraction) and a coded value of NYHA class (score of HF severity).

HF progression:
(i) Hospital admission for HF or
(ii) Clinical worsening due to the presence of at least 2 criteria compared to the levels observed at the index visit: a ≥1 increase in NYHA class or a ≥10 points decrease in LVEF or a ≥50% (and in any case > 25 mg) increase in furosemide dosage or a new combination of diuretics (thiazides + furosemide)

whatever came first.

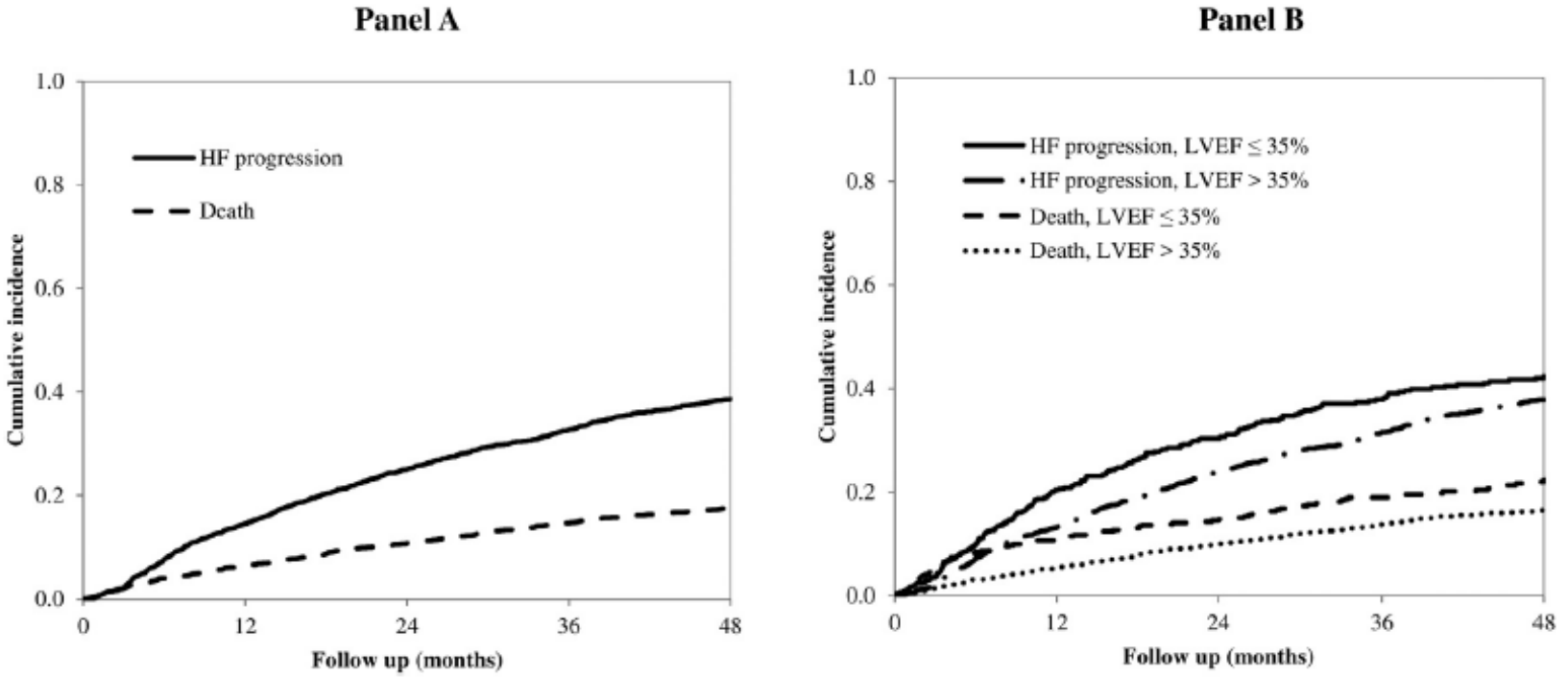
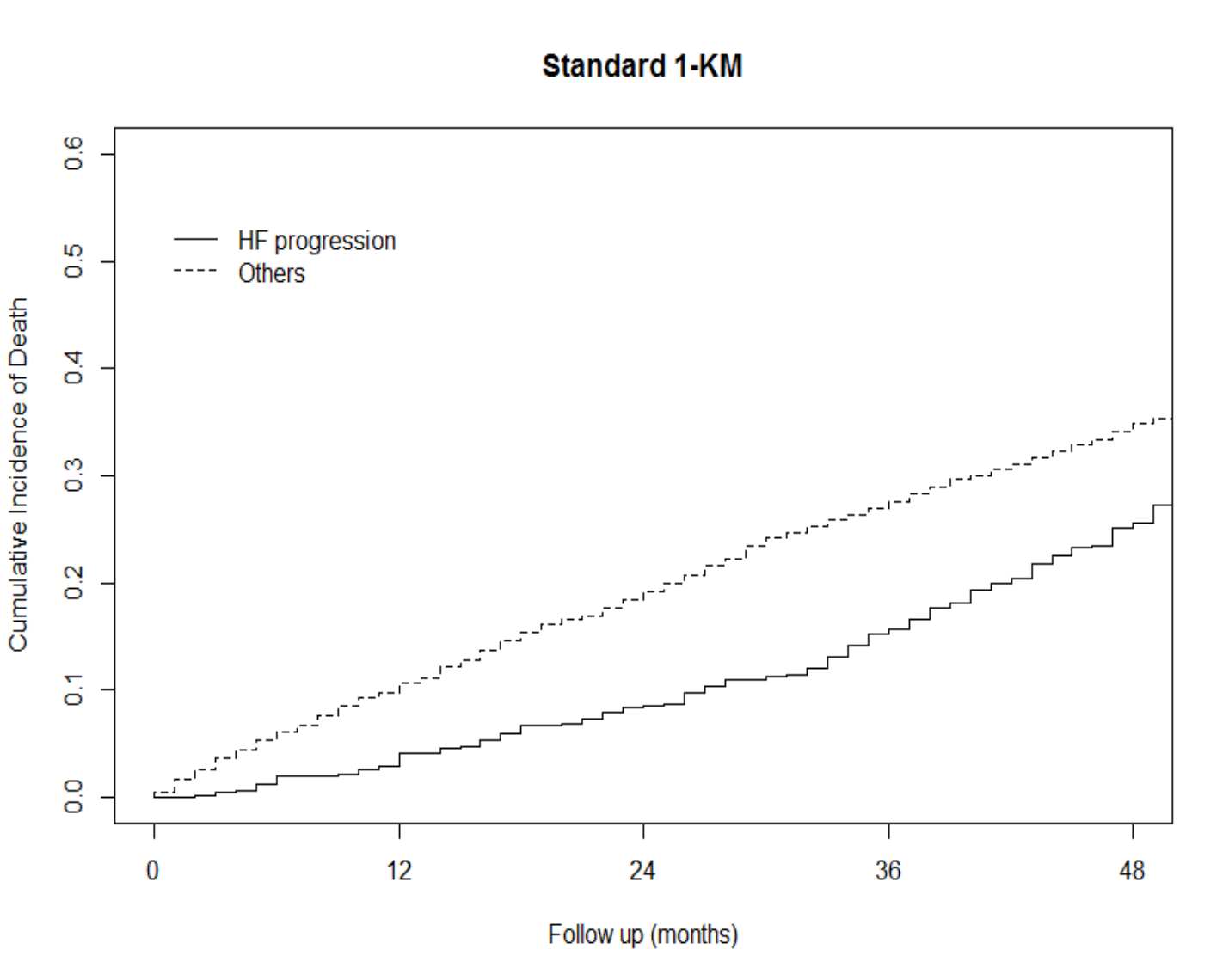


Fig. 1. Cumulative incidence of HF progression and mortality as a competing event among the whole cohort (Panel A) as well as according to LVEF (Panel B). HF: Heart failure; LVEF: Left ventricular ejection fraction.

Impact of HF progression on mortality: wrong approach

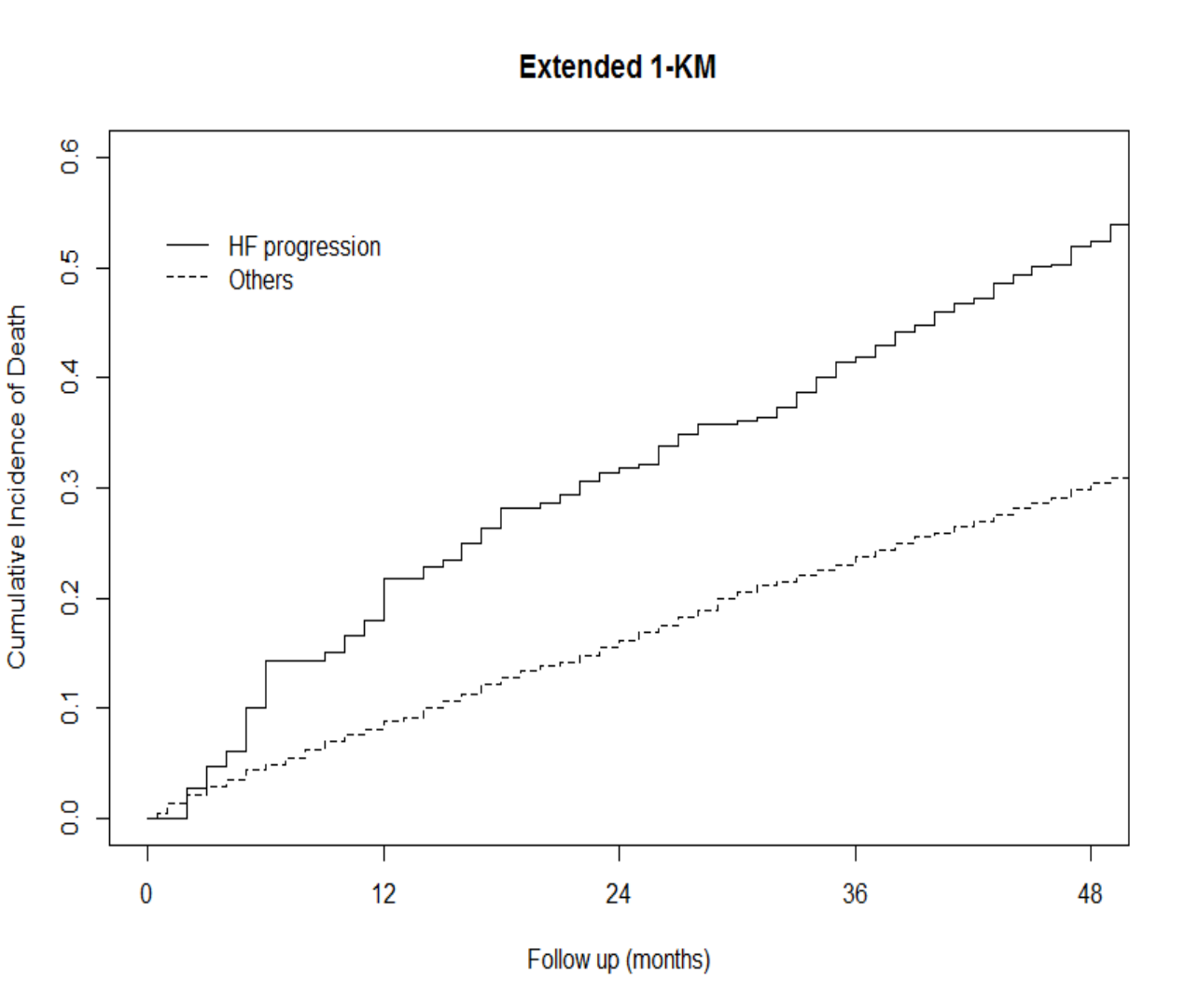


If we consider HF progression as a **time fixed covariate at baseline**, the effect of such disease worsening appears ***protective*** with respect to the risk of death !

Immortal/time dep bias !

UNITÀ DI BIOSTATISTICA
Dipartimento Universitario Clinico di Scienze Mediche Chirurgiche e della Salute

Impact of HF progression on mortality: correct approach



If instead we treat HF progression as a ***time-dependent*** covariate, the effect of such disease worsening appears ***a risk factor*** with respect to the risk of death, as it is expected*.

Technical details of the estimation procedure:

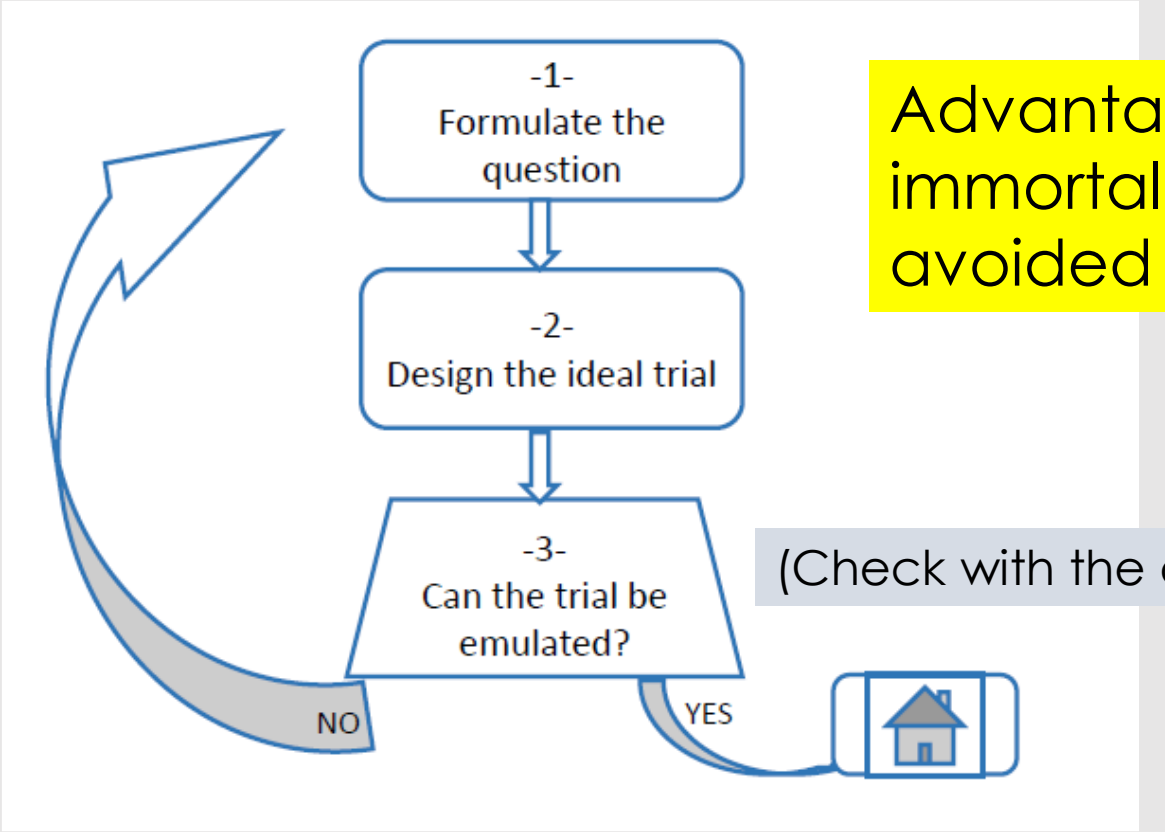
S.M. Snapinn, Q. Jiang, B. Iglewicz "*Illustrating the Impact of a Time-Varying Covariate With an Extended Kaplan-Meier Estimator*", The American Statistician, Vol. 59, No. 4, 2005

*There is no standard test to compare directly these survival curves (patients can contribute to different curves at different times during follow-up)

The framework of **target trial emulation (TTE)** offers guidance for avoiding **errors** in data manipulation and analysis of observational data that may lead to biased results [Hernan and Robins, 2016].

The implementation of TTE is not however as straightforward as it.

It consists of 3 iterative steps:



Advantage: selection and immortal time bias are avoided !!!

(Check with the data you have!)

Hernan and Robins. *Using Big Data to Emulate a Target Trial When a Randomized Trial Is Not Available*. American Journal of Epidemiology, 2016, 183, 758–764


The End


"OUR STATISTICIAN WILL DROP IN AND
EXPLAIN WHY YOU HAVE NOTHING TO WORRY ABOUT."