# Tutorial on causal inference – Part III

## Trial emulation and survival analysis for disease incidence registers: A case study on the causal effect of pre-emptive kidney transplantation

Ingeborg Waernbaum

Department of Statistics, Uppsala University

*joint with Camila Olarte Parra, Staffan Schön and Els Goetghebeur*

**Virtual RSS Conference 2020, $9^{th}$ September 2020**

# A causal analysis in practice

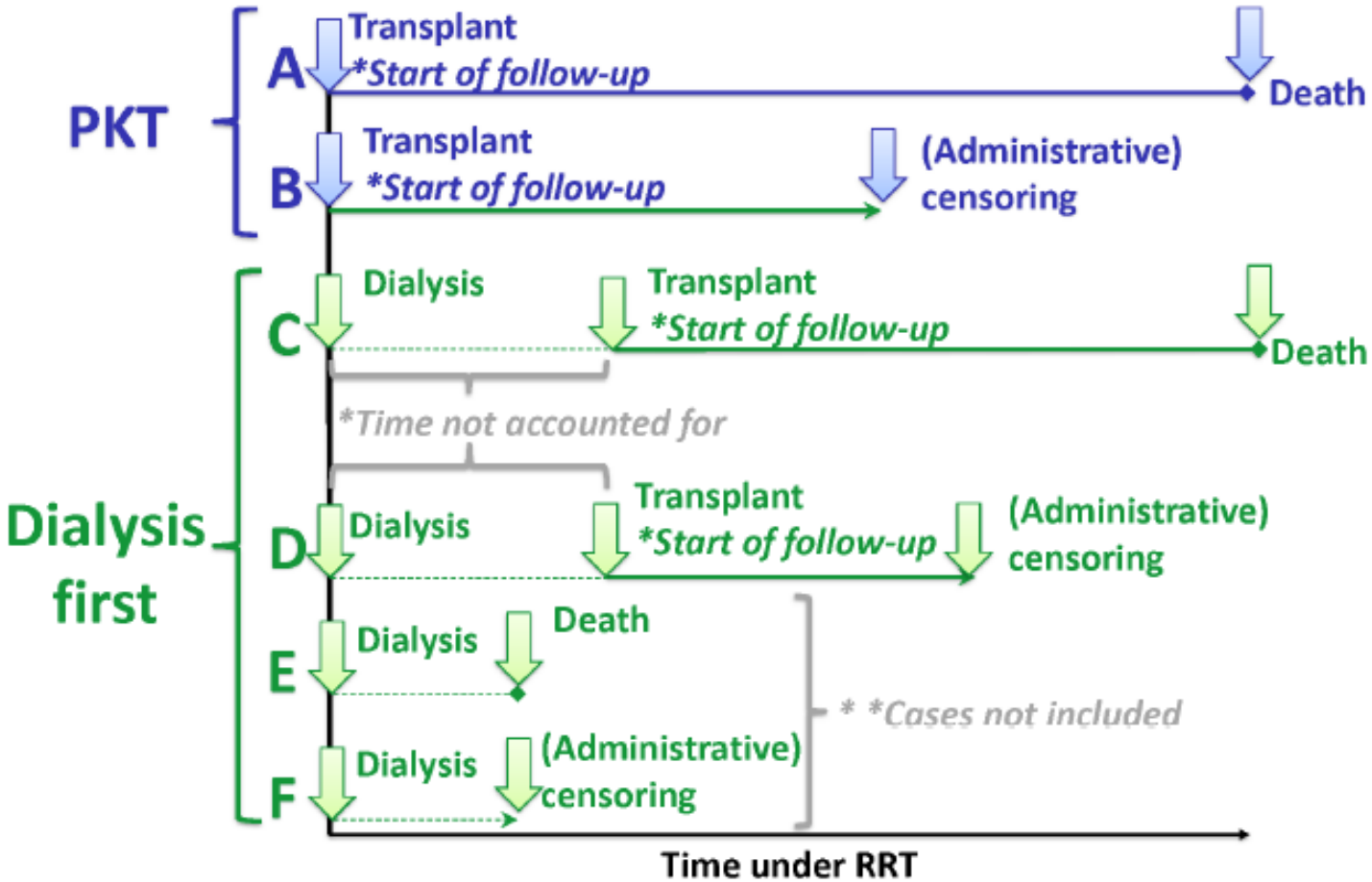
- Parts I and II: Key steps and simulation learner for a binary exposure.
- Here, we apply the key steps in a study of causal effects of pre-emptive kidney transplantation on survival time analyzed with data from a population-based disease incidence register.
- Applying the steps, illustration of target trial emulation using observational data. We target estimands: ITT and as-treated estimands defined by (potential) survival outcomes.
- The example illustrates further challenges with observational data, specifically disease incidence registers: Missing covariate data in the early years, patient profiles and/or level of care changes over calender time, e.g, biasing unadjusted K-M curves.

# Kidney failure: End Stage Renal Disease - ESRD

- SNR, Swedish Kidney Registry, population based incidence register for patients with ESRD.
- When kidneys fail, patients cannot survive without Renal Replacement Therapy (RRT).
- National Quality Register, started in 1991, high coverage of incident cases.
- Two types of therapies: Dialysis and Transplantation.

# Types of treatments: Transplantation and dialysis

- Average life expectancy on dialysis 5-10 years, transplantation 12-20 years.
- Sources of bias in previous studies: Immortal time bias (time not accounted for if $t_0$ is set at transplant). Cases not included (if death or censoring before transplant).
- Pre-emptive kidney transplantation (PKT): Transplantation without previous dialysis.


PKT
A
Transplant
*Start of follow-up
Death
B
Transplant
*Start of follow-up
(Administrative)
censoring
Dialysis
first
C
Dialysis
Transplant
*Start of follow-up
Death
*Time not accounted for
D
Dialysis
Transplant
*Start of follow-up
(Administrative)
censoring
E
Dialysis
Death
F
Dialysis
(Administrative)
censoring
* *Cases not included
Time under RRT

# Notation

- Binary exposure/treatment, $PKT$ (1=PKT treatment, 0 = dialysis first). PKT (n=1,097) or dialysis first (n=18,430).
- $Z$, baseline variables, measured before $t_0$ (start of follow-up ).
- Outcomes, $T_1$, $T_0$, potential outcomes under exposure/no exposure.
- "Observed" outcome $T = PKT \cdot T_1 + (1 - PKT) \cdot T_0$ (consistency assumption), Censoring, $C$.
- Propensity score: $e(Z) = P(PKT = 1|Z)$
- Assumptions: No unmeasured confounding at baseline $T_1, T_0 \perp\!\!\!\perp PKT \mid Z$, Non-informative censoring $C \perp\!\!\!\perp T \mid Z, PKT$, Missing at random, $Z_{mis} \perp\!\!\!\perp T \mid Z_{obs}, PKT$.

# Target trial emulation - mimicking an RCT that would answer our causal question.

Estimands in RCTs:

*Intention-to-treat effect*, effect of being assigned PKT at a specific time point, $t_0$. Marginalizing over subsequent treatments as they occur in the population.

*Per-protocol effect*, targets the effect of adhering to a treatment as established by the researcher.

*As-treated effect*, estimand defined by treatment actually received, e.g., taking into account the sequence of treatments.

# Key steps of causal inference - tutorial

## Definitions

- (*Define exposure*) Binary exposure, PKT - vs dialysis first (possibly followed by transplant).
- (*Define the outcome*) - $T$, Time from RRT onset ($t_0$) to death.
- (*Define the population(s) of interest*) - All swedish patients in need of RRT (1991-2017), eligible for either treatment (thus satisfying the positivity assumption). Full cohort, PKT, dialysis first.
- (*Formalize potential outcomes*) - $T_1$, Potential time until death under PKT, $T_0$ potential time until death if start is instead with dialysis. Starts at $t_0$, (time of treatment assignment).

# Key steps of causal inference

## Targets of estimation

- (*Specify the 'estimand'*) - *ITT* Average potential survival curve differences, $\frac{1}{n}\sum_{i\in\{P_i\}}(S_1(t;Z_i) - S_0(t;Z_i))$, *As-treated*: $\exp(-\psi)$ causal parameter in a structural accelerated failure time model (factor of survival time lost under dialysis).

## Estimation

- (*State assumptions for its identification & estimation*) No unmeasured confounding at baseline, non-informative censoring, missing at random, model assumptions.
- (*Estimate the target causal effect*) ITT: Survival curve estimates, As-treated analysis: AFT model estimates (adjusted for baseline confounding).
- (*Perform sensitivity analyses as needed*) Results from several alternative analyses, e.g., with and without imputation, additional (or removed) confounders etc.

# Study population selection

Exclusion: Patients over 75 years, from foreign or unkown region, patients receiving RRT abroad, patients with cancer.

| Study population | PKT | Dialysis first |
|---|---|---|
| Number of adult patients from SRR 1991-2017 | 1,214 (100.0) | 28,312 (100.0) |
| Number of patients older than 75 years | 4 (0.3) | 7,108 (25.1) |
| Number of patients from foreign or unknown region | 18 (1.5) | 170 (0.6) |
| Number of patients who receive RRT abroad | 57 (4.7) | 79 (0.3) |
| Number of patients who died or got censored on same day of RRT onset | 1 (0.1) | 20 (0.1) |
| Number of patients with a history of cancer or unknown cancer status | 37 (3.0) | 2,501 (8.8) |
| Total sample | 1,097 (90.4) | 18,434 (65.1) |

# Survival descriptives

| (Sub)population | Patients, n (%) | Deaths n (%) | % deaths per row | Median person-years at risk | Hazard rate |
|---|---|---|---|---|---|
| RRT cohort | 19,531 (100) | 12,073 (100) | 61.8 | 4.1 | 0.10 |
| PKT group | 1,097 (5.6) | 196 (1.6) | 17.9 | 7.6 | 0.02 |
| Dialysis first group | 18,434 (94.4) | 11,877 (98.4) | 64.4 | 3.9 | 0.11 |

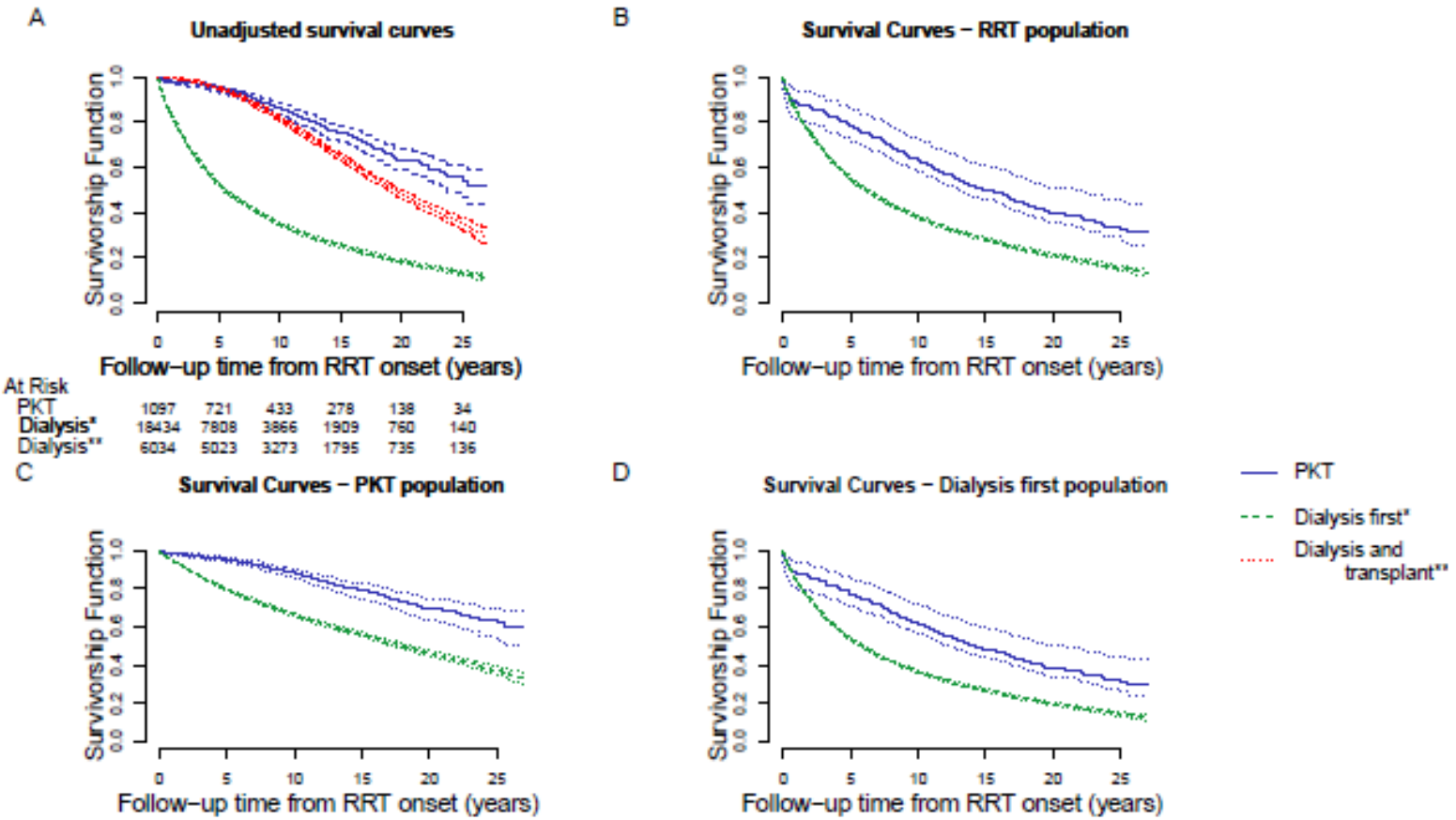
# Confounding and missing covariates

- Baseline confounders: Age, sex, region, primary kidney disease, calender year of RRT onset, diabetes, hypertension, ischemic heart disease, cerebrovascular disease and peripheral artery disease.
- Comorbidities were missing before 1998.
- If we exclude cases before 1998, would imply excluding 36% of events. Multiple imputation assuming missingness at random, multiple imputation.
- Covariate adjustment through Cox-models fitted for the treatment groups separately. With fitted values, calculate covariate specific survival curves under each possible treatment. The average of these two curves over the relevant study population are then contrasted.

# ITT analysis

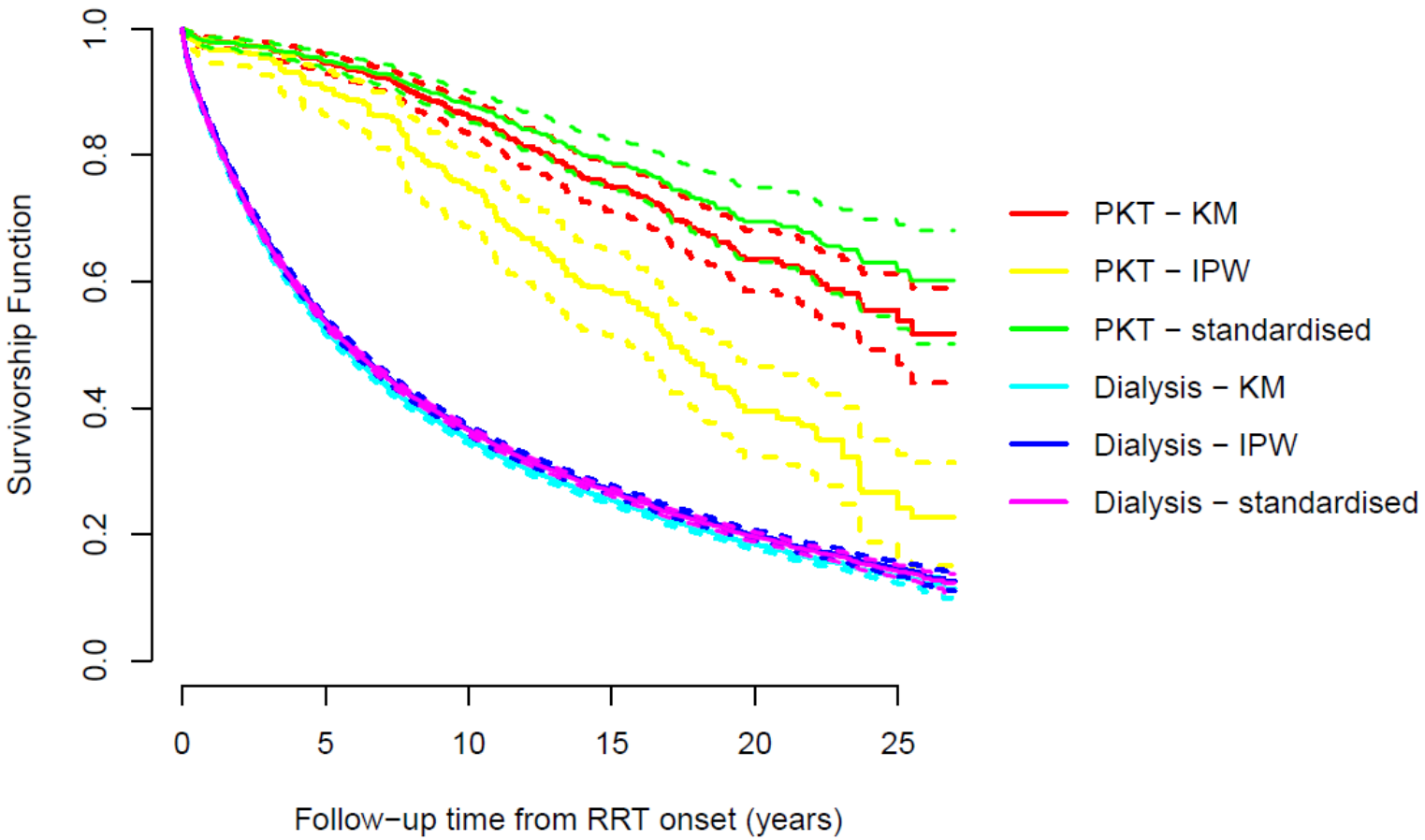
| Year(s) after RRT onset | Survival under PKT (95% CI) | Survival under dialysis first (95% CI) | Difference in survival (95% CI) |
|---|---|---|---|
| RRT population | | | |
| 1 | 0.88 (0.81, 0.94) | 0.85 (0.84, 0.85) | 0.03 (-0.04, 0.09) |
| 5 | 0.78 (0.72, 0.86) | 0.55 (0.54, 0.56) | 0.23 (0.17, 0.31) |
| 10 | 0.63 (0.58, 0.73) | 0.38 (0.38, 0.39) | 0.25 (0.20, 0.35) |
| 15 | 0.50 (0.45, 0.60) | 0.28 (0.28, 0.29) | 0.22 (0.17, 0.32) |
| 20 | 0.40 (0.35, 0.51) | 0.21 (0.20, 0.22) | 0.19 (0.14, 0.30) |
| 25 | 0.33 (0.28, 0.45) | 0.15 (0.14, 0.16) | 0.18 (0.12, 0.30) |
| PKT subpopulation | | | |
| 1 | 0.98 (0.97, 0.99) | 0.94 (0.94, 0.95) | 0.03 (0.03, 0.04) |
| 5 | 0.95 (0.94, 0.96) | 0.79 (0.79, 0.80) | 0.16 (0.14, 0.17) |
| 10 | 0.88 (0.86, 0.90) | 0.66 (0.65, 0.67) | 0.22 (0.19, 0.24) |
| 15 | 0.79 (0.74, 0.82) | 0.56 (0.55, 0.57) | 0.23 (0.18, 0.27) |
| 20 | 0.70 (0.63, 0.75) | 0.46 (0.45, 0.48) | 0.23 (0.17, 0.29) |
| 25 | 0.62 (0.53, 0.69) | 0.37 (0.35, 0.39) | 0.25 (0.16, 0.33) |
| Dialysis subpopulation | | | |
| 1 | 0.88 (0.80, 0.94) | 0.84 (0.84, 0.85) | 0.03 (-0.04, 0.09) |
| 5 | 0.77 (0.71, 0.85) | 0.54 (0.53, 0.54) | 0.24 (0.17, 0.32) |
| 10 | 0.62 (0.57, 0.72) | 0.36 (0.36, 0.37) | 0.25 (0.20, 0.35) |
| 15 | 0.48 (0.44, 0.59) | 0.27 (0.26, 0.27) | 0.21 (0.17, 0.32) |
| 20 | 0.38 (0.34, 0.50) | 0.20 (0.19, 0.20) | 0.19 (0.14, 0.30) |
| 25 | 0.31 (0.26, 0.44) | 0.14 (0.13, 0.15) | 0.17 (0.12, 0.30) |

# ITT analysis

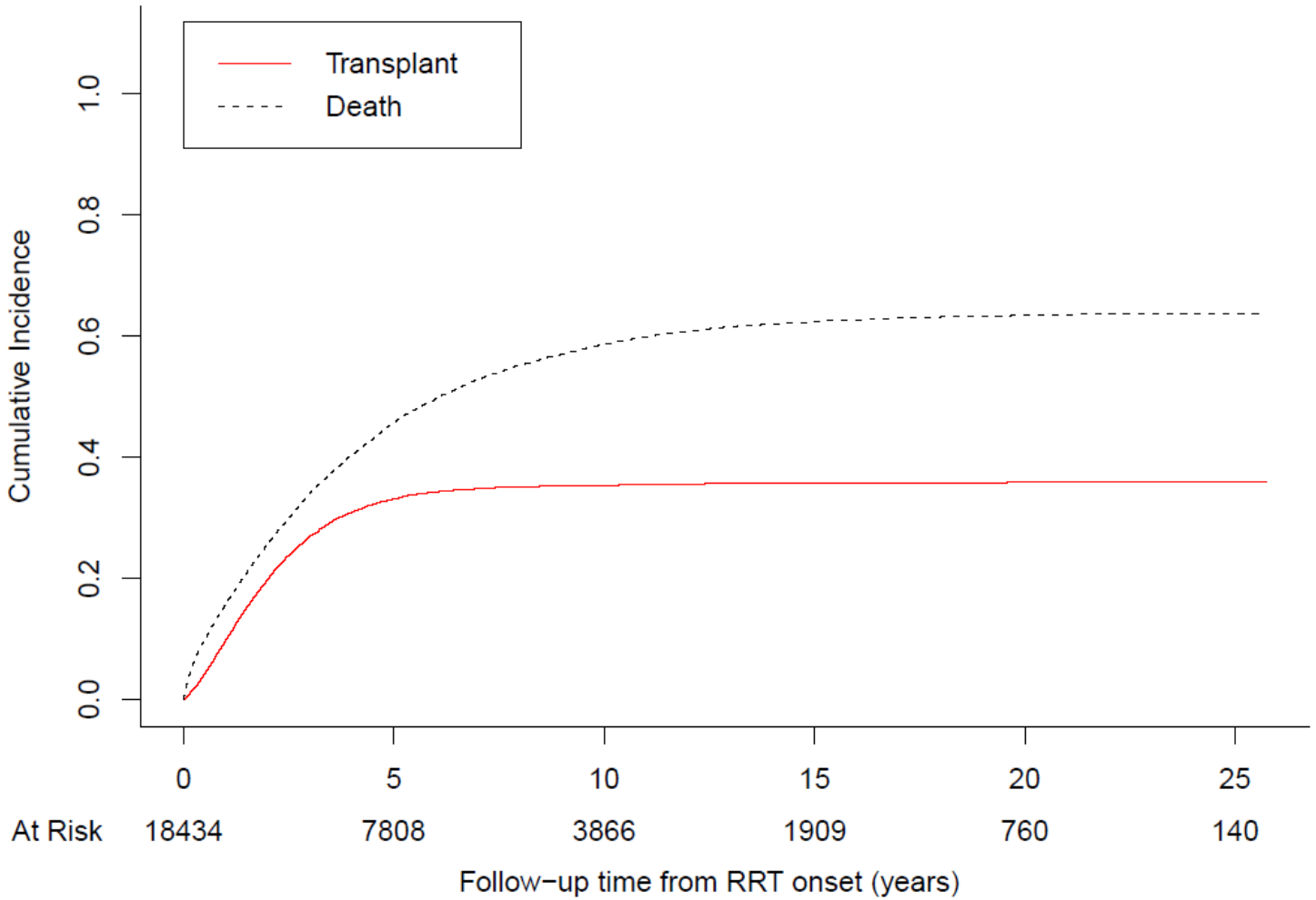


# Censoring, regression vs ipw

- If the baseline confounders include sufficient adjustment set $Z$ for both confounding and censoring the regession model will also adjust for the covariate effect on the censoring.
- If instead we would have adjusted the survival curves through IPW with the propensity score it would not remedy the covariate effect on censoring. Instead we then would need time-varying covariates for a censoring model.


Survivorship Function
0.0
0.2
0.4
0.6
0.8
1.0
0
5
10
15
20
25
Follow−up time from RRT onset (years)
PKT − KM
PKT − IPW
PKT − standardised
Dialysis − KM
Dialysis − IPW
Dialysis − standardised

# Dialysis first group; Cumulative Incidence of Transplant and Death



# As treated: Accelerated failure time model, delayed transplant

Relating survival time on dialysis to what it might have been following PKT.
$T_w$ = the survival time spent without initial transplant
$T_r$ = residual survival time following the delayed transplant (if any).

Structural model (one parameter):

$$T_{1_i}(\psi) \stackrel{d}{=} T_{w_i} \exp(-\psi) + T_{r_i},$$

$\exp(-\hat{\psi}) = 4.8$ 95% CI $(3.9, 5.8)$

On average, every year of survival under initial dialysis is equivalent to surviving 4.8 years under PKT.

# As treated: Accelerated failure time model, delayed transplant

Structural model (two parameters):

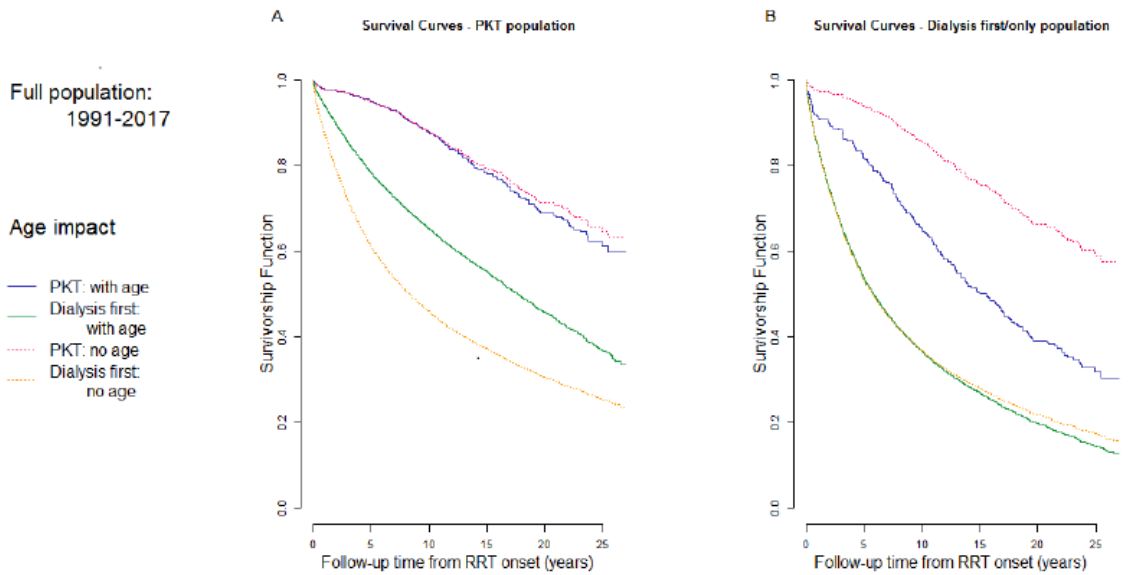
$$T_{1_i}(\boldsymbol{\psi}) \stackrel{d}{=} T_{w_i} \exp(-\psi_w) + T_{r_i} \exp(-\psi_r),$$

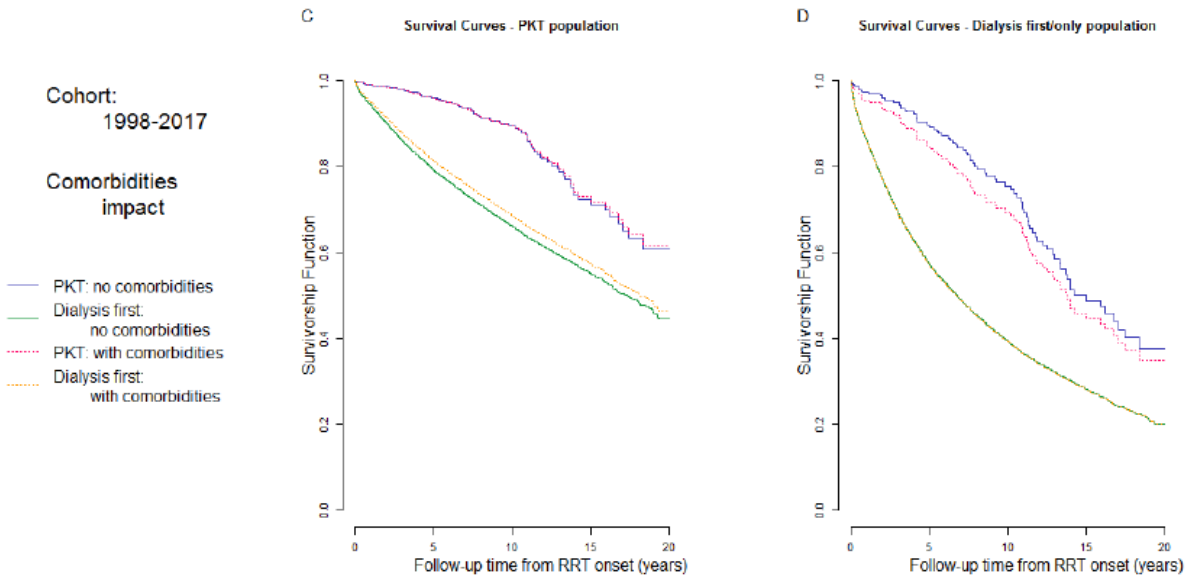
$\exp(-\hat{\psi}_w) = 5.6$ 95% CI $(3.9, 5.8)$
$\exp(-\hat{\psi}_r) = 0.7$ 95% CI $(0.5, 0.9)$

On average, a year of survival under initial dialysis is equivalent of surviving 5.6 years under PKT, and then a year of delayed transplant would be equivalent to living 0.7 years under PKT.

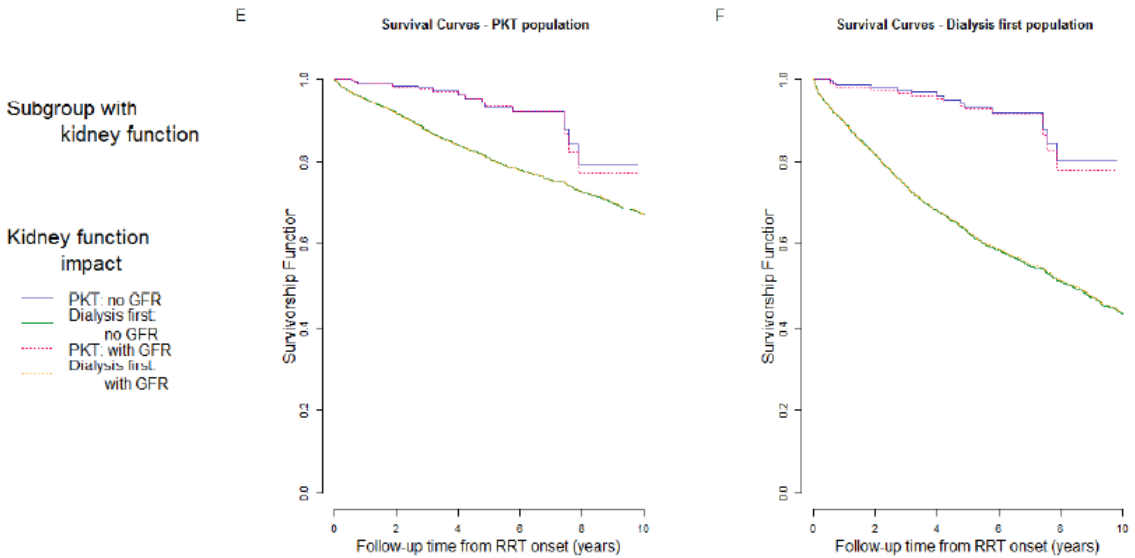
# Sensitivity analyses

We display results under different scenarious related to our assumptions: e.g., survival estimates with and without imputing missing covariates, omitting confounders.




Cohort:
1998-2017
Comorbidities
impact
PKT: no comorbidities
Dialysis first:
no comorbidities
PKT: with comorbidities
Dialysis first:
with comorbidities
C
Survival Curves - PKT population
Survivorship Function
Follow-up time from RRT onset (years)
D
Survival Curves - Dialysis first/only population
Survivorship Function
Follow-up time from RRT onset (years)

# No unmeasured confounding:



# Summary and conclusions

- We investigate the effect of PKT versus dialysis first for patients with ESRD in an observational study with data from the Swedish Kidney Register, setting $t_0$ at the time of disease onset.
- Challenges: missing comorbidities in beginning of registry, no time varying confounders.
- We estimate ITT effects in form of survival differences and as treated effects as survival time lost while on dialysis.
- The findings support previous knowledge and extends the knowledge to effects in different target populations (e.g. dialysis first population).