Quantiles of the survival time from Inverse Probability Weighted Kaplan-Meier estimates

Andrea Discacciati

A joint work with Nicola Orsini and Matteo Bottai

Unit of Biostatistics and Unit of Nutritional Epidemiology
The Institute of Environmental Medicine (IMM)
Karolinska Institute

The aim

- Estimate **quantiles of the survival time** for each level of a categorical exposure variable
- Easily done indirectly from Kaplan-Meier estimates

$$\hat{t}_q = \min\left\{t_i \middle| \hat{S}\left(t_i\right) \leq 1-q\right\},\ q \in \left(0,\ 1\right)$$

- Biased estimates in presence of confounding effects
- Inverse Probability Weighted (IPW) Kaplan-Meier estimates
- -`stqkm`- Stata command

Notation

- $i = 1,\ldots,n$ – independent subjects
- T_i – possibly right-censored event time
- δ_i – censoring indicator
- $E_i = 1,\ldots,K$ – exposure groups
- $\mathbf{Z}_i$ – covariate vector

The weights

- To the i-th subject in the k-th group, the weight w_{ik} is assigned

$$w_{ik} = \Pr\left(E_i = k \middle| \mathbf{Z}_i\right)^{-1}$$

- Weights can be estimated
  - Non-parametrically (sample proportions)
  - Parametrically (logistic or multinomial logistic models)

IPW Kaplan-Meier estimates

- Weighted number of events: $d_{jk}^{w} = \sum_{i:T_i = t_j} w_{ik} \delta_i I(E_i = k)$
- Weighted risk set: $r_{jk}^{w} = \sum_{i:T_j \geq t_j} w_{ik} I(E_i = k)$
- IPW Kaplan-Meier for the *k*-th group

$$\hat{S}_k^w(t) = \prod_{j:t_j \leq t} \left(1 - d_{jk}^w / r_{jk}^w\right)$$

- Marginal survival curves

Monte Carlo Simulation

- Variable C (3 categories), which will act as the confounding variable
- Exposure variable (2 categories) conditionally on C
- The survival curves of the two exposure groups are different if the confounding effect of C is not taken into account
- They are the same if controlling for C
- True marginal survival function calculated using a Riemann-Stieltjes integral

Unadjusted analysis I

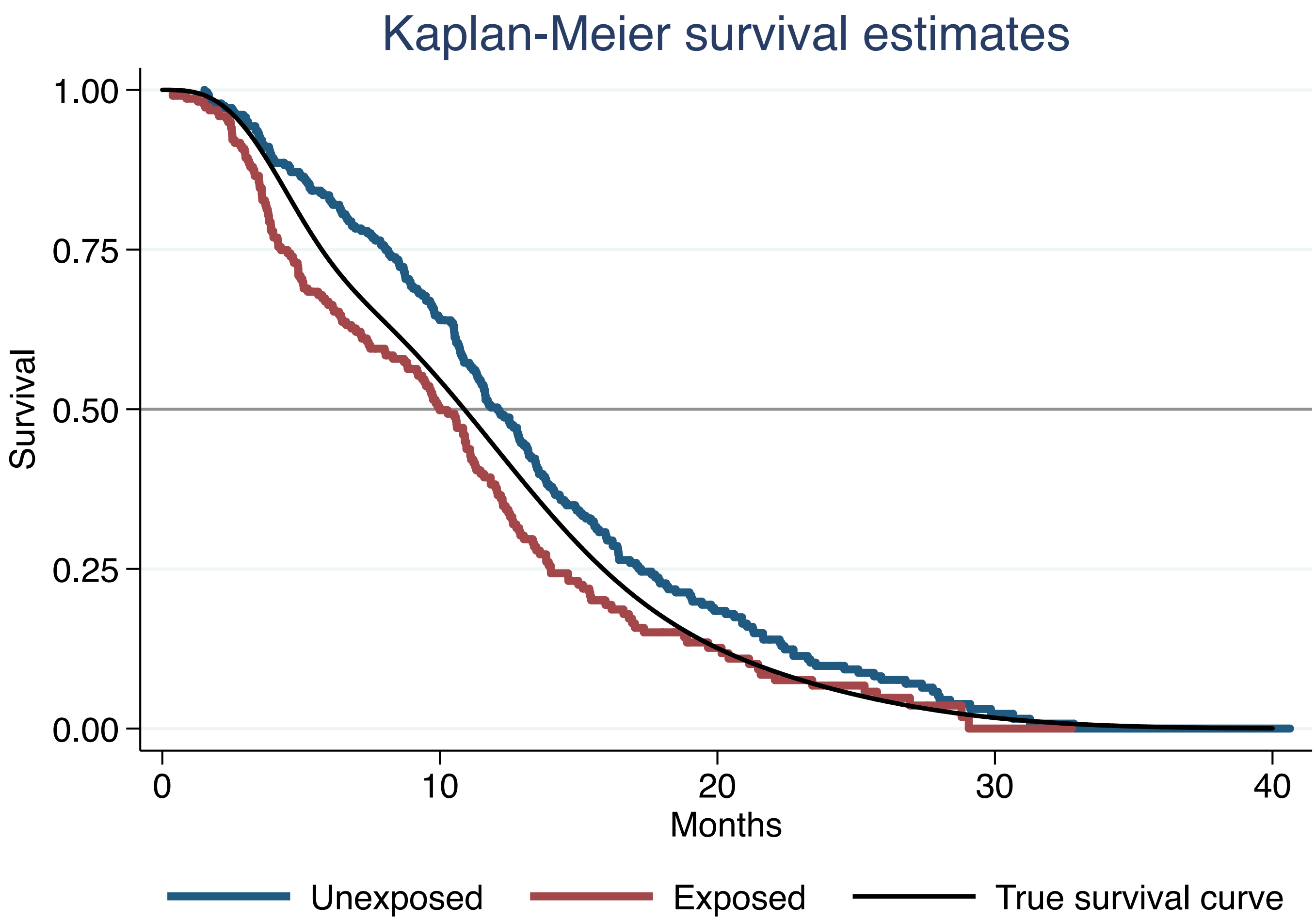
- `stset time, fail(event)`
- `stqkm exposed, q(50)`

```
Quantiles from IPW Kaplan-Meier                       Number of obs  =       500
  bootstrap(20) SEs
```

_t	Observed Coef.	Bootstrap Std. Err.	t	P>|t|	Normal-based [95% Conf.	Interval]
q50						
exposed	-1.935799	.6536007	-2.96	0.003	-3.219954	-.6516438
_cons	12.14248	.4175894	29.08	0.000	11.32202	12.96293

- The difference in the median survival time between exposed and unexposed subjects is –1.9 months (95% CI: –3.2 to –0.6 months)

Unadjusted analysis II



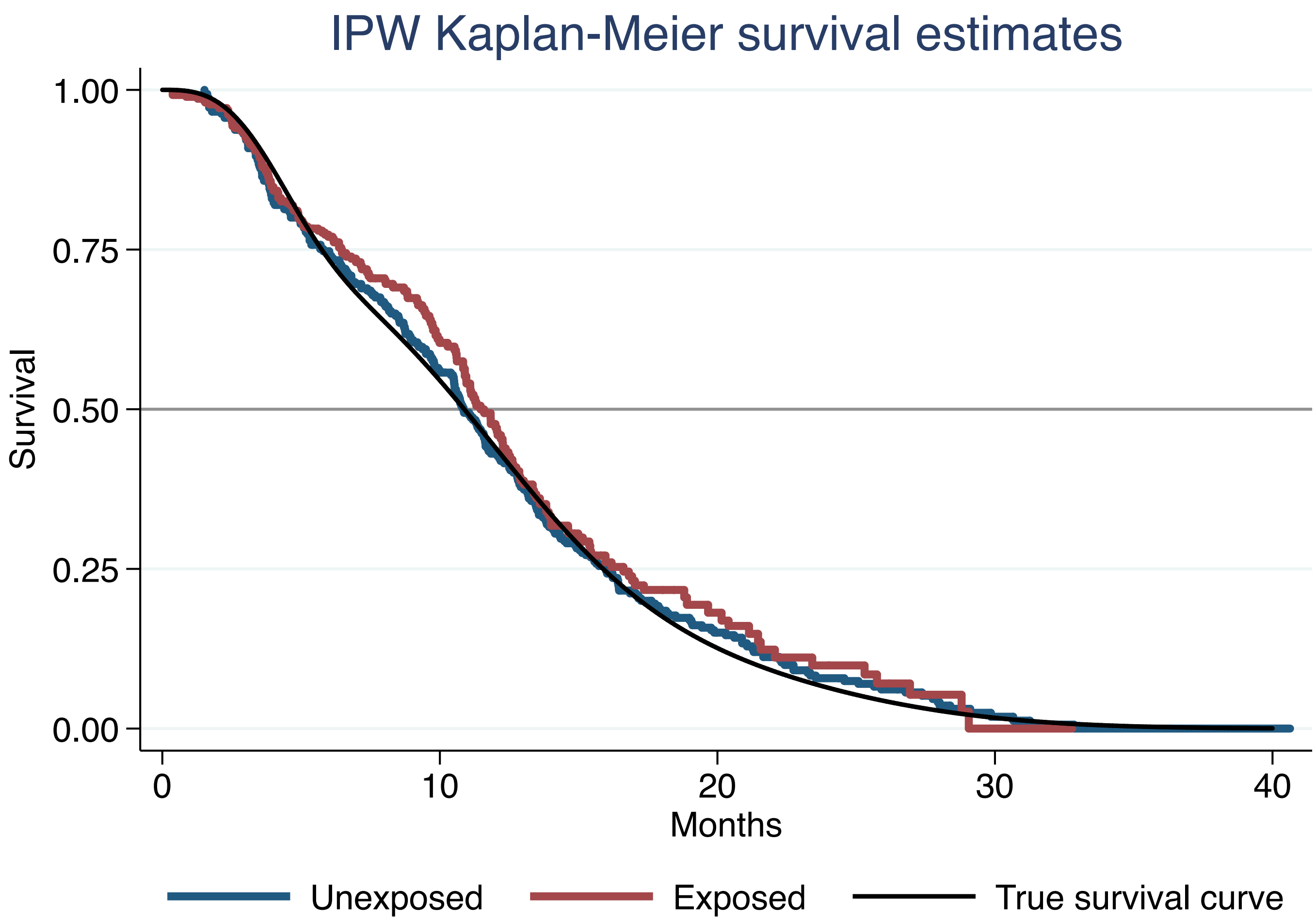
Adjusted analysis I

- `xi: stqkm exposed, q(50) adjustfor(i.c)`

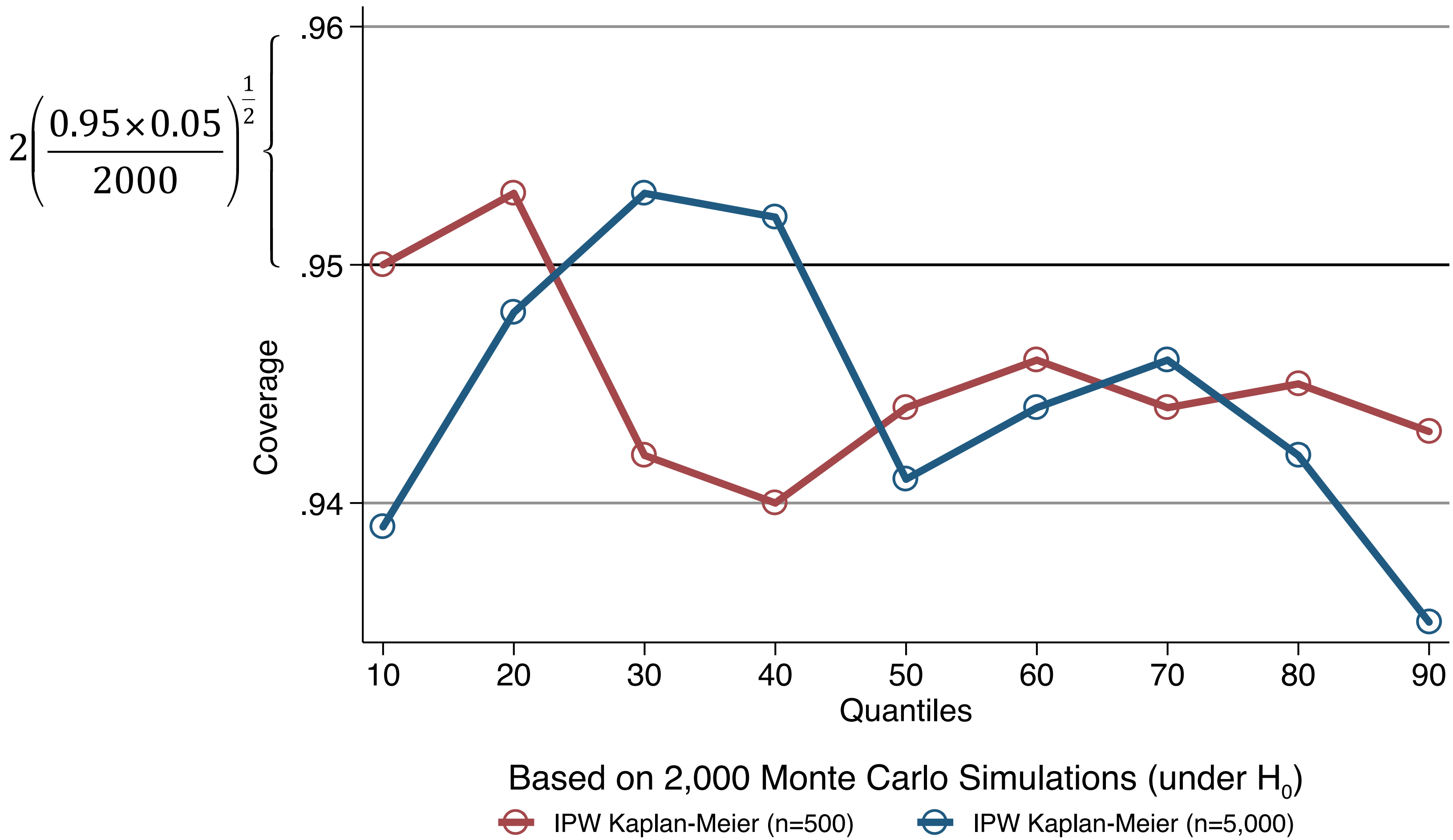
```
Quantiles from IPW Kaplan-Meier                     Number of obs =       500
  bootstrap(20) SEs

------------------------------------------------------------------------------
             |  Observed   Bootstrap                         Normal-based
          _t |     Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
q50          |
     exposed |  .7250223   .7367458     0.98   0.326    -.7224909    2.172536
       _cons |  10.87058   .4079353    26.65   0.000     10.06909    11.67207
------------------------------------------------------------------------------
```

- The adjusted difference in the median survival time between exposed and unexposed subjects is 0.7 month (95% CI: –0.7 to 2.2 months)

Adjusted analysis II



95% Confidence Interval coverage



Additional features of -stqkm-

- It's possible to estimate simultaneously more than one quantile of the survival time
  - `xi: stqkm exposed, q(25 50 75) adjustfor(i.c)`

- It's possible to use post-estimation commands such as -lincom- and -test-
  - `lincom [q50]_cons + [q50]exposed`

_t	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
(1)	11.5956	.5657995	20.49	0.000	10.48395	12.70725

  - `test [q25]exposed = [q75]exposed = 0`

```
( 1)  [q25]exposed - [q75]exposed = 0
( 2)  [q25]exposed = 0

      F(  2,   498) =    0.29
           Prob > F =    0.7459
```

Conclusions

Karolinska Institutet

- Limitations
  - Only categorical exposures
  - Can be computationally slow

- Strengths
  - The idea and the **interpretation** are **straightforward**
  - No assumptions of proportionality of the hazards
  - Good confidence interval coverage under H_0

References

- Cole SR, Hernán MA. Adjusted survival curves with inverse probability weights. *Comput Methods Programs Biomed*. 2004 Jul;75(1):45-9.
- Xie J, Liu C. Adjusted Kaplan-Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Stat Med*. 2005 Oct 30;24(20):3089-110.
- Discacciati A, Orsini N, Bottai M. Quantiles of the survival time from Inverse Probability Weighted Kaplan-Meier estimates. In preparation for the Stata Journal.