Comparing different strategies for timing of dialysis initiation through inverse probability weighting

Arvid Sjölander
Olof Nyrén
Rino Bellocco
Marie Evans
Background

- Dialysis used in treatment of kidney diseases since 1960s.
  - No uniform agreement on when to initiate, in the course of disease

- One large randomized trial (Cooper et al, 2010).
  - No significant difference between early or late initiation
  - Potential bias due to severe compliance problems

- Several large observational studies (e.g. Korevaar et al, 2001; Traynor et al, 2002).
  - Contradictory results, possibly due to different analysis methods
  - Trad analysis methods biased, even in absence of unmeasured confounding
Static vs dynamic treatment strategies

- Trad stats methods developed for static treatment strategies
  → comparing fixed levels of treatment. e.g.
    ▪ initiate dialysis at diagnosis, vs
    ▪ do not initiate dialysis

- Our problem concerns dynamic strategies
  → depend on time varying patient characteristics e.g.
    ▪ initiate dialysis when eGFR drops below 16 ml/min/1.73m², vs
    ▪ initiate dialysis when eGFR drops below 12 ml/min/1.73m²

- Recent methods in causal inference developed for dynamic strategies (Cain et al, 2010; Orellana et al, 2010)
  → Inverse Probability Weighting (IPW)
Before we start

- The problem is both clinically important, statistically challenging, and "generic".

  → Similar research questions and statistical problems arise often in the analysis of cohort studies.

- The focus here is on the methodological issues.

  → Non-medical, non-technical exposition
Outline

- Preliminaries
  - Data
  - Exclusions
  - Categorization of "timing"
  - Table 1

- Traditional methods
  - Why they are biased

- IPW method

- Results
Data

- **Cohort:**
  - Swedish-born patients, 18-74 years old
  - diagnosed with permanent loss of renal function during May 1996 – May 1998
  - not acute renal failure or previous renal transplant.
  - 10 years follow-up

- **Baseline measures:** factors associated with survival and/or timing of dialysis initiation (details later)

- **Follow-up measures:**
  - date of dialysis initiation
  - date kidney transplantation
  - date of death
  - eGFR (details later)
estimated Glomelural Filtration Rate (eGFR)

- Measure of renal function (ml/min/1.73m²)

  → 186 *[S-Cr]-1.154 *[Age]-0.203 *[0.742 if patient is female]*[1.212 if patient is black] (Levey et al, 1999)
  → Continuous measure; high values are “good”
  → Irregular measures; 1-6 for each patient

- Our statistical method(s) require regular measures

  → Spline interpolation
  → From now: ”eGFR” as shorthand for ”interpolated value of eGFR”
Exclusions

- Before exclusions: n = 920
  - 72: only one eGFR measure
  - 82: increasing eGFR curve
  - 235: eGFR >16 (baseline) during follow up.
  - 13: started dialysis at eGFR >16.
  - 19: missing data on at least one variable.
  - 58: died or were censored before dialysis start (only a problem for trad analyses)

- After exclusions: n=441(trad), n=499 (IPW)
Timing of dialysis initiation

- Three level categorical:
  - "Early" if $16 \geq \text{eGFR} > 12$ at initiation
  - "Intermediate" if $12 \geq \text{eGFR} > 7.5$ at initiation
  - "Late" if $7.5 \geq \text{eGFR}$ at initiation

- Research question: which of the three levels of initiation is most beneficial, wrt to survival?
<table>
<thead>
<tr>
<th>Covariate</th>
<th>Level</th>
<th>n</th>
<th>%</th>
<th>P timing†</th>
<th>P survival††</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td>male</td>
<td>308</td>
<td>70</td>
<td>0.08</td>
<td>0.31</td>
</tr>
<tr>
<td></td>
<td>female</td>
<td>133</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>&lt;45</td>
<td>108</td>
<td>25</td>
<td>0.07</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>45-65</td>
<td>200</td>
<td>45</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&gt;65</td>
<td>133</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BMI</td>
<td>&lt;20</td>
<td>32</td>
<td>7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>20-25</td>
<td>188</td>
<td>43</td>
<td>0.17</td>
<td>0.40</td>
</tr>
<tr>
<td></td>
<td>25-30</td>
<td>161</td>
<td>37</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&gt;30</td>
<td>60</td>
<td>14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Charlson comorbidity index</td>
<td>&lt;2</td>
<td>155</td>
<td>35</td>
<td></td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>2-4</td>
<td>169</td>
<td>38</td>
<td></td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>&gt;4</td>
<td>117</td>
<td>27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Primary renal disease</td>
<td>diabetes</td>
<td>139</td>
<td>31</td>
<td></td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>hypertension</td>
<td>61</td>
<td>14</td>
<td></td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>glomerulonephritis</td>
<td>113</td>
<td>26</td>
<td></td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>other/unknown</td>
<td>128</td>
<td>29</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
# Table 1, cont’d

<table>
<thead>
<tr>
<th>Covariate</th>
<th>Level</th>
<th>n</th>
<th>%</th>
<th>P timing†</th>
<th>P survival††</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smoking</td>
<td>no smoking</td>
<td>176</td>
<td>40</td>
<td>0.10</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>≤ 20 packyears</td>
<td>148</td>
<td>34</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&gt; 20 packyears</td>
<td>117</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alcohol</td>
<td>no alcohol</td>
<td>88</td>
<td>20</td>
<td></td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>≤ 50 gram/day</td>
<td>224</td>
<td>51</td>
<td>0.02</td>
<td></td>
</tr>
<tr>
<td></td>
<td>&gt; 50 gram/day</td>
<td>129</td>
<td>29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Education level</td>
<td>&lt; 10 years</td>
<td>226</td>
<td>51</td>
<td>0.30</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>10-13 years</td>
<td>113</td>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>&gt; 13 years</td>
<td>102</td>
<td>23</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Remark on censoring

- After kidney transplantation, the eGFR is no longer a meaningful measure of the renal function.
  
  → Thus we consider kidney transplantation as a censoring event.

- The methods that we present are easily adapted to accommodate (random) censoring.

- However, only a minor fraction were factually censored.
  
  → For simplicity, we thus ignore censoring in the exposition.
Trad analysis 1: from initiation (FI)

- Counting survival time from start of dialysis

- Adjusted hazards ratios:
  - Early: reference
  - Intermediate: 0.81 (0.51, 1.21)
  - Late: 0.77 (0.48, 1.25)
Problems

- Unmeasured confounding

- Lead time bias
  - Early starters are earlier in the progression of disease, at initiation
  - Favorizes early starters
  - May underestimate a true protective effect of late initiation
Trad analysis 2: from threshold (FT)

- Counting survival time from fixed point in the course of disease
  → eGFR = 16 ml/min/1.73m²

- Adjusted hazards ratios, counting from eGFR=16:
  → Early: reference
  → Intermediate: 0.62(0.39,0.98)
  → Late: 0.56(0.35,0.91)
Problems

- **Unmeasured confounding**

- **Immortal time bias**
  - Artificial immortal period between baseline and initiation
  - Favorizes late starters
  - May overestimate a true protective effect of late initiation
IPW - outline

- Target parameter - a causal effect
- Conditions for unbiased estimation of causal effects in observational studies
- Analysis
- Results
Target parameter

- The causal effect of each dynamic dialysis strategy on survival
  - Survival/hazard function if everybody in the cohort would start early/intermediate/late

- Same cohort under different strategies
  - comparing ”like with like”.

- Practically impossible
  - Each patient can only be assigned to one strategy.
Ideal randomized trials

- Each subject is randomly assigned to one of the three strategies.

- By randomization, subjects are "comparable" across strategies.

\[ \rightarrow \text{The survival for those randomly assigned to "late initiation" is identical to the survival had everybody been assigned to "late initiation"} \]
Observational studies

- The IPW analysis produces causal effects in the absence of unmeasured confounding
  → Untestable

- Unmeasured potential confounders in our data:
  → Levels of
    - calcium
    - phosphate
    - parathyroid hormone
    - Albuminuria

  → Longitudinal information on co-morbid diseases
Expanded risk sets

- **FT**
  - dialysis strategies are determined by "peaking into the future"

- **Expanded risk sets**
  - each subject contributes to each strategy s(he) is presently compatible with
    - one replicate for each compatible strategy
    - artificial censoring of replicate when no longer compatible

Arvid Sjölander
Problem

- The artificial censoring of replicates is non-random.
  
  → Replicates who are censored early have, on average lower eGFR than subjects who are censored late.

  → Early censoring is associated with worse prognosis.

- This problem can be solved with IPW
Notation

- Time $t$ in months
- $V = \text{baseline covariates}$
- $C(t) = \text{censoring status at } t \ (0/1)$
- $Y(t) = \text{death status at } t \ (0/1)$
- $A(t) = \text{dialysis status at } t \ (0/1)$
- "Overbar" for variable history, e.g. $\overline{A}(t) = \{A(1), A(2), \ldots, A(t)\}$
- $X = \text{timing}$
  - Early
  - Intermediate
  - Late
IPW

- Construct a "pseudo risk set" by assigning a time varying weight to each subject in the expanded risk set.

\[ w_i(t) = \prod_{k=0}^{t} \frac{\Pr_{\text{EXP}}\{C_i(k) = 0 | Y_i(k) = C_i(k-1) = 0, X_i, V_i\}}{\Pr\{A_i(k) | Y_i(k) = C_i(k) = 0, \bar{A}_i(k-1), \bar{L}_i(k), V_i\}} , \]

- The denominator adjusts for non-random artificial censoring.

- The numerator stabilizes the resulting estimates.
Estimation of weights

\[ w_i(t) = \prod_{k=0}^{t} \frac{\Pr_{\text{EXP}} \{ C_i(k) = 0 \mid Y_i(k) = C_i(k-1) = 0, X_i, V_i \}}{\Pr \{ A_i(k) \mid Y_i(k) = C_i(k) = 0, \bar{A}_i(k-1), \bar{L}_i(k), V_i \} }, \]

- Pooled logistic regression:

\[ \logit \left[ \Pr_{\text{EXP}} \{ C(t) = 0 \mid Y(t) = C(t-1) = 0, X, V \} \right] = \eta_0 + \eta_1 t + \eta_2 t^2 + \eta_3 'V + \eta_X X \]

\[ \logit \left[ \Pr \{ A(t) = 1 \mid Y(t) = C(t) = \bar{A}(t-1) = 0, \bar{L}(t), V \} \right] = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 L(t) + \alpha_4 L^2(t) + \alpha_5 L(t-1) + \alpha_6 L^2(t-1) + \alpha_7 L(t-2) + \alpha_8 L^2(t-2) + \alpha_8 'V \]
Main model

- Standard surv analysis softwares do not allow time varying weights

- Pooled log reg model of discrete time hazards:

\[
\text{logit} \left[ \text{Pr}_{\text{EXP}} \{ Y(t) = 1 \mid Y(t-1) = C(t) = 0, X, V \} \right] \\
= \theta_0 + \theta_1 t + \theta_2 t^2 + \theta_3 t^3 + \theta_4 t^4 + \theta_5' X + \theta_6' X \cdot t + \theta_7' X \cdot t^2 + \theta_8' V
\]

- Note 1: baseline hazard parametrically specified.

- Note 2: no proportional hazards assumption
  \( \rightarrow \) prop hazards is particularly unrealistic for dynamic treatment regimes
Hypothesis tests

\[
\logit \left[ \Pr_{\text{EXP}} \{ Y(t) = 1 \mid Y(t-1) = C(t) = 0, X, V \} \right] = \theta_0 + \theta_1 t + \theta_2 t^2 + \theta_3 t^3 + \theta_4 t^4 + \theta_5 'X + \theta_6 'X \cdot t + \theta_7 'X \cdot t^2 + \theta_8 'V
\]

- Proportional hazards if \( \theta_6 = \theta_7 = 0 \)
  \rightarrow Bootstrap p-value: 0.05

- No effect of \( X \) on \( Y \) if \( \theta_5 = \theta_6 = \theta_7 = 0 \)
  \rightarrow Bootstrap p-value: 0.12
Interpretation of the model

\[ \text{logit} \left[ \Pr_{\text{EXP}} \{ Y(t) = 1 \mid Y(t - 1) = C(t) = 0, X, V \} \right] \]

\[ = \theta_0 + \theta_1 t + \theta_2 t^2 + \theta_3 t^3 + \theta_4 t^4 + \theta_5' X + \theta_6' X \cdot t + \theta_7' X \cdot t^2 + \theta_8' V \]

- No simple interpretation of specific parameter value \( \theta \)

- We use standardization to obtain survival function if everybody in the cohort would start early/intermediate/late

- To facilitate comparison with the FT and FI methods, we re-analyzed the data with these methods, using the model above.
Standardization

- Fit main model to estimate discrete time hazards
  \[ \Pr\{Y(t) = 1 \mid Y(t-1) = 0, X, V\} \]

- Accumulate 1-hazards to obtain survival probabilities
  \[ \Pr\{\overline{Y}(t) = 0 \mid X, V\} \]

- Average over marginal (over \(X\)) sample distribution for \(V\) to obtain
  \[ E[\Pr\{\overline{Y}(t) = 0 \mid X, V\}] \]
Standardized survival functions, FI

- $p = 1.00$
- Favorizes early starters
Standardized survival functions, FT

- $p = 0.30$
- Favorizes late starters
Standardized survival functions, IPW

- $p = 0.12$
- Unbiased
Conclusions

- Estimating the optimal timing of dialysis initiation is
  - Clinically important
  - Statistically challenging
  - Generic

- Traditional methods are biased, even in the absence of unmeasured confounding

- Unbiased estimates can be obtained through IPW.
References

- **Our work:**

- **Dynamic treatment strategies:**