



Practice of Epidemiology

Comparing Different Strategies for Timing of Dialysis Initiation Through Inverse Probability Weighting

Arvid Sjölander*, Olof Nyrén, Rino Bellocco, and Marie Evans

* Correspondence to Dr. Arvid Sjölander, Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Nobels Väg 12, 171 77 Stockholm, Sweden (e-mail: arvid.sjolander@ki.se).

Initially submitted January 05, 2011; accepted for publication June 28, 2011.

Dialysis has been used in the treatment of patients with end-stage renal disease since the 1960s. Recently, several large observational studies have been conducted to assess whether early initiation of dialysis prolongs survival, as compared with late initiation. However, these studies have used analytic approaches which are likely to suffer from either lead-time bias or immortal-time bias. In this paper, the authors demonstrate that recently developed methods in the causal inference literature can be used to avoid both types of bias and accurately estimate the ideal time for dialysis initiation from observational data. This is illustrated using data from a nationwide population-based cohort of patients with chronic kidney disease in Sweden (1996–2003).

bias (epidemiology); dialysis; dynamic model; kidney failure, chronic; lead time; probability weighting; therapeutics

Abbreviations: CI, confidence interval; eGFR, estimated glomerular filtration rate; FI, from initiation; FT, from threshold; GFR, glomerular filtration rate; IDEAL, Incremental Decrease in Clinical Endpoints Through Aggressive Lipid Lowering; IPW, inverse probability weighting.

Dialysis has been used in the treatment of patients with end-stage renal disease since the 1960s. However, there is still no uniform agreement on the ideal time at which to start dialysis in the progression of disease. Recent results from the randomized IDEAL (Incremental Decrease in Clinical Endpoints Through Aggressive Lipid Lowering) Trial (1) indicated no difference between early start and late start. However, these results are not readily interpreted, since the IDEAL study suffered from compliance problems. Several large observational studies have been conducted to compare different strategies (2–11), with somewhat contradictory results. Korevaar et al. (2), Bonomini et al. (3), Tattershall (4), and Tang et al. (5) observed better survival for patients who started dialysis early, whereas Traynor et al. (6), Sawhney et al. (7), Beddhu et al. (8), Kazmi et al. (9), Lassalle et al. (10), and Stel et al. (11) observed better survival for those who started late.

One possible explanation for this discrepancy is that “baseline” (i.e., the time point from which survival is measured) was defined differently in these studies. For instance, Korevaar et al. (2) defined baseline as the time of dialysis initiation—we refer to this method as the “from initiation” (FI) method—whereas Traynor et al. (6) defined baseline as

the time at which renal function dropped below a fixed threshold (estimated creatinine clearance <20 mL/minute); we refer to this method as the “from threshold” (FT) method. Patients who die or drop out of the study before they start dialysis pose difficulties for both methods. These patients cannot be classified as “early” starters or “late” starters and were excluded from the analyses of both Korevaar et al. (2) and Traynor et al. (6). The FI method gives early starters an artificial survival advantage, since they are, at baseline, earlier in the course of their disease progression than late starters; this bias is usually referred to as “lead-time bias.” The FT method gives every subject an artificial “immortal period” between baseline and dialysis initiation, since patients who die or drop out before the start of dialysis are excluded. However, since late starters are further away from baseline at the start of dialysis, these subjects are given a longer immortal period and thus an artificial survival advantage; this bias is usually referred to as “immortal-time bias.”

To the best of our knowledge, an analysis method which avoids both types of bias has not yet been proposed in the nephrologic literature. This is not surprising, given that traditional statistical methods are developed for static treatment

strategies. In the causal inference literature, a treatment strategy is called “static” if it does not depend on the evolution of patient characteristics over time. Clinically relevant strategies, however, are typically “dynamic”—that is, decisions and interventions are guided by the current clinical status of the individual patient. In a series of papers, Robins et al. (12–14) have shown that dynamic treatment strategies can be evaluated and compared through a statistical method based on expanded risk sets and inverse probability weighting (IPW). Our purpose in this paper is to show how the general methods developed by Robins et al. (12–14) can be applied to tackle the specific problem of comparing dynamic strategies for dialysis initiation.

Recently, we analyzed the importance of early initiation of dialysis versus late initiation using data from a nationwide population-based cohort of patients with chronic kidney disease in Sweden (15). We found a nonsignificant advantage for late starters (15). In this paper, we reanalyze the Swedish chronic kidney disease data with the IPW method. The data are briefly described below; a more detailed description can be found in our previous article (15). For illustrational purposes, we conduct a slightly simplified FT analysis and contrast it with an analysis of the same data using the FI method. We then describe the proposed IPW method and apply it to the same data. Our motivation for the IPW method will be largely heuristic; a more stringent discussion can be found in the article by Robins et al. (12–14).

DATA AND NOTATION

Study population and follow-up

The decision to initiate dialysis is guided by the clinical status of the patient and measures of renal function, most often the estimated glomerular filtration rate (eGFR), which is derived from creatinine levels in serum. A joint effort made between May 20, 1996, and May 31, 1998, by all laboratories that analyzed serum creatinine in Sweden identified all Swedish-born patients aged 18–74 years who were shown for the first time to have a serum creatinine level greater than 3.4 mg/dL (men) or 2.8 mg/dL (women) and whose loss of renal function was considered permanent. Patients with acute renal failure, pre- or postrenal azotemia, or a renal transplant were excluded. At inclusion, patients received a questionnaire on anthropometric measures, socioeconomic background, diet, alcohol consumption, and tobacco use. Information on occupational history, medical history, and use of prescribed and nonprescribed medications was collected thereafter. Medical records were reviewed for information on comorbid conditions and laboratory values until the end of follow-up (June 1, 2003). We also linked the cohort to the Swedish Renal Registry, a national register of all patients starting renal replacement therapy in Sweden that contains information on date of the start of renal replacement therapy and kidney transplantation date.

eGFR interpolation

For each patient, we estimated the glomerular filtration rate (GFR) at several points along the disease progression

continuum during follow-up, using the 4-variable Modification of Diet in Renal Disease equation ($eGFR = 186 \times [\text{serum creatinine}]^{-1.154} \times [\text{age}]^{-0.203} \times [0.742 \text{ if patient is female}] \times [1.212 \text{ if patient is black}]$) (16). The serum creatinine determinations were part of the routine care and varied in number and timing (median number per patient, 5.3; range, 1–6). In the FT and IPW analyses, we started counting person-time and outcome events among patients still at risk at $eGFR = 16 \text{ mL/minute}/1.73 \text{ m}^2$ (“baseline”). Our analyses required more detailed eGFR measures than were available. In particular, the FI analysis requires that eGFR has been measured at dialysis initiation; the FT analysis requires that the time at which eGFR drops below the specified baseline threshold be measured; and the IPW analysis requires that eGFR has been measured at regular intervals. We used regression splines to interpolate between the available eGFR measures, yielding a value for eGFR in each month for each patient. Below, we use “eGFR” as shorthand for “interpolated value of eGFR.”

Exclusions

Before exclusions, 920 patients were eligible. For 72 patients, eGFR was measured on only 1 occasion, which would have hampered interpolation. For 82 additional patients, the interpolation algorithm resulted in a monotonically increasing trajectory for GFR, which is due to either improvements or interpolation errors. For 235 additional patients, the eGFR never dropped below $16 \text{ mL/minute}/1.73 \text{ m}^2$ during follow-up; thus, these patients never passed “baseline” for the FT and IPW analyses. An additional 13 patients started dialysis before their eGFR dropped below $16 \text{ mL/minute}/1.73 \text{ m}^2$. Another 19 patients had missing data for at least 1 variable used in the analyses. These 421 patients ($72 + 82 + 235 + 13 + 19 = 421$) were excluded from all 3 analyses. In the FI and FT analyses, we had to exclude 58 additional patients who did not start dialysis before death or censoring. These patients were not excluded from the IPW analysis. In total, 441 patients were included in the FI and FT analyses, and 499 patients were included in the IPW analysis.

Notation

In our illustration, time t is measured in months, with $t = 0$ for baseline. Thus, $t = 0$ corresponds to different points in disease progression in the FI analysis, as compared with the FT and IPW analyses. Let $C(t)$ be the censoring status ($C(t) = 1$ if the patient is censored at or before t , and 0 otherwise), $Y(t)$ be the event status ($Y(t) = 1$ if the patient dies at or before t , and 0 otherwise), $A(t)$ be the dialysis status ($A(t) = 1$ if the patient initiated dialysis at or before t , and 0 if the patient has not yet initiated dialysis), and $L(t)$ be the eGFR level at month t . We use an overbar to represent a variable’s history—for example, $\bar{A}(t) = \{A(1), A(2), \dots, A(t)\}$. Each patient who initiates dialysis during follow-up is classified as an “early” ($16 \geq eGFR > 12$ at initiation), “intermediate” ($12 \geq eGFR > 7.5$), or “late” ($eGFR \leq 7.5$) starter. Let X denote the timing of initiation (“early,” “intermediate,” or “late”). V is a vector of covariates measured at enrollment in the study, including sex, age (<45, 45–65, or >65 years),

body mass index (weight (kg)/height (m)²; <20, 20–<25, 25–<30, or ≥30), Charlson comorbidity index (17) (categorized into approximate tertiles: <2, 2–4, or >4), primary renal disease (categorized into diabetes, hypertension, glomerulonephritis, or other/unknown), smoking (no smoking, ≤20 pack-years, or >20 pack-years), alcohol consumption (no alcohol, ≤50 g/day, or >50 g/day), and educational level (<10, 10–13, or >13 years). The covariates are thoroughly described elsewhere by Evans et al. (15).

Table 1 shows data for the 441 subjects who were included in all analyses. It displays, for each covariate, the number and fraction of patients at each level of the covariate, together with *P* values from tests for association with timing of dialysis initiation and survival, respectively.

STANDARD METHODS

The FI method

We used Cox proportional hazards regression to assess the association between timing of dialysis initiation and survival, adjusted for covariates *V*. For each patient, baseline was defined as the month in which dialysis was initiated. Let $\lambda(t|X, V)$ denote the conditional hazard at month *t*, given *X* and *V*. We modeled $\lambda(t|X, V)$ as

$$\log\{\lambda(t|X, V)\} = \log\{\lambda_0(t)\} + \beta_1'X + \beta_2'V. \quad (1)$$

Using the “early” starters as the reference group, we obtained estimated hazard ratios of 0.81 (95% confidence interval (CI): 0.51, 1.21) and 0.77 (95% CI: 0.48, 1.25) for the “intermediate” and “late” initiators, respectively. In a test for equal hazards across treatment groups, we obtained a *P* value of 0.57. Thus, the point estimates obtained from this FI analysis suggest a trend across levels of starting time, with better survival for a later start. However, the *P* value shows that this trend is nonsignificant. As we argued in the Introduction, an FI analysis artificially favors the early starters. Thus, we may suspect that this analysis underestimates the “true” effect.

The FT method

We refitted the adjusted Cox model in equation 1, now defining baseline as the month in which eGFR first dropped below 16 mL/minute/1.73 m², for each patient. We obtained estimated hazard ratios of 0.62 (95% CI: 0.39, 0.98) and 0.56 (95% CI: 0.35, 0.91) for the “intermediate” and “late” starters, respectively (*P* = 0.06 in a test for equal hazards). The FT analysis suggests a much stronger trend than the FI analysis. However, as we argued in the Introduction, an FT analysis artificially favors the late initiators. Thus, we may suspect that the “truth” lies somewhere in between the results of the FI and FT analyses.

EXPANDED RISK SETS AND IPW

Expanded risk sets

An appealing feature of the FT method is that survival time is counted from the same point in the progression of

Table 1. Covariate Data for Chronic Kidney Disease Patients Accrued in a Swedish Cohort Between May 20, 1996, and May 31, 1998

Covariate	No.	%	<i>P</i> for Timing ^a	<i>P</i> for Survival ^b
Sex			0.08	0.31
Male	308	70		
Female	133	30		
Age, years			0.07	<0.01
<45	108	25		
45–65	200	45		
>65	133	30		
Body mass index ^c			0.17	0.40
<20	32	7		
20–<25	188	43		
25–<30	161	37		
≥30	60	14		
Charlson comorbidity index			<0.01	<0.01
<2	155	35		
2–4	169	38		
>4	117	27		
Primary renal disease			<0.01	<0.01
Diabetes	139	31		
Hypertension	61	14		
Glomerulonephritis	113	26		
Other/unknown	128	29		
Smoking, pack-years			0.10	<0.01
None	176	40		
≤20	148	34		
>20	117	26		
Alcohol consumption, g/day			0.02	<0.01
None	88	20		
≤50	224	51		
>50	129	29		
Educational level, years			0.30	<0.01
<10	226	51		
10–13	113	26		
>13	102	23		
Initiation of dialysis			NA	<0.01 ^d
Early	37	8		
Intermediate	195	44		
Late	209	48		

Abbreviations: FT, from threshold; NA, not applicable.

^a Chi-square test of association with timing of dialysis initiation, *X*.

^b Log-rank test of association with survival time, counted from the month in which the estimated glomerular filtration rate first dropped below 16 mL/minute/1.73 m² (“baseline” in the FT and inverse probability weighting analyses).

^c Weight (kg)/height (m)².

^d This *P* value corresponds to an unadjusted FT analysis.

disease for all patients. However, this comes at the cost of having to define treatment groups by “peeking into the future.” To elaborate on this issue, consider a fictitious patient

A whose eGFR level falls monotonically by 1 mL/minute/1.73 m² per month after baseline (eGFR = 16 mL/minute/1.73 m²). Suppose further that patient A initiates dialysis at month 7, at eGFR = 9, and dies at month 10. In retrospect, we can say that A is an “intermediate” initiator. However, following A prospectively over time, it is not until month 4, when his eGFR drops to 12, that we can rule him out as an “early” starter. Thus, it is somewhat artificial to label A as already being “intermediate” at baseline.

To overcome this logical deficiency, we note that during months 0–3, patient A had the potential to become an “early,” “intermediate,” or “late” starter, yet A was subject to exactly the same treatment level (no dialysis), regardless of the subsequent course. Thus, we may consider patient A to have followed (or to have been “compatible with”) all 3 strategies up through month 3. Similarly, after having attained eGFR = 12 (at month 4) and thus no longer having the possibility of becoming an “early” starter, A can still be considered to follow both the “intermediate” and “late” strategies during months 4–6. At month 7, A initiates dialysis at eGFR = 9, which identifies him as an “intermediate” initiator. In practice, the time-dependent contribution from each patient can be accommodated by expanding the risk sets in each month, so that each patient who is still at risk contributes with 1 replicate observation for each treatment strategy that the patient is presently compatible with. For instance, during months 0–3, we let A contribute with 3 replicates to each risk set, following strategies X = “early,” X = “intermediate,” and X = “late,” respectively. At month 3, the replicate with X = “early” is artificially censored. During months 4–6, we let A contribute with 2 replicates to each risk set, following strategies X = “intermediate” and X = “late,” respectively. At month 6, the replicate with X = “late” is artificially censored. During months 7–10, we let A contribute with 1 replicate to each risk set, following strategy X = “intermediate.” At month 10, this replicate contributes with an event and is excluded from the risk sets for all later months.

A survival analysis (e.g., Cox proportional hazards regression) based on these expanded risk sets does not suffer from lead-time bias or immortal-time bias. The former is avoided by counting survival time from the same point in the progression of disease for each patient, and the latter is avoided by not excluding patients who die or become censored before they start dialysis; hence, no patient is given an artificial “immortal period.” As an example of the latter, consider a patient B whose eGFR level falls monotonically by 1 mL/minute/1.73 m² per month after baseline and who dies at month 7, without having initiated dialysis. The FI or FT analysis requires that we exclude patient B, since her timing of initiation cannot be determined. However, using the modified risk sets, we let B contribute in the same manner as A during months 0–6. At month 7, she contributes with an event and is excluded from the risk sets for all later months.

Inverse probability weighting

Though it is free from lead-time and immortal-time bias, an analysis based on expanded risk sets suffers from a third kind of bias, induced by the artificial censoring of replicates. Heuristically, this bias arises because replicates who initiate

dialysis at *t* will have, on average, lower eGFR levels than those who do not initiate dialysis and remain uncensored at *t*. Thus, replicates censored at *t* will have a worse prognosis than replicates who are uncensored; the artificial censoring is nonrandom (12). We followed the method of Robins et al. (12–14) and used IPW to adjust for the artificial censoring. Towards this end, we constructed a “pseudo-expanded risk set” at each month, by assigning an individual time-varying weight to each patient in the original expanded risk set. The weight for patient *i* at month *t* is calculated as

$$w_i(t) = \prod_{k=0}^t \frac{\text{Pr}_{\text{EXP}}\{C_i(k) = 0|Y_i(k) = C_i(k - 1) = 0, X_i, V_i\}}{\text{Pr}\{A_i(k)|Y_i(k) = C_i(k) = 0, \bar{A}_i(k - 1), \bar{L}_i(k), V_i\}}, \tag{2}$$

where Pr_{EXP}(·) refers to the probability of (·) in the expanded data. The denominator of *w_i(t)* is the part which adjusts for the artificial censoring and is analogous to the denominator used in IPW for static treatments (18–20). Each factor in the denominator is the probability of the patient’s observed dialysis level, given past dialysis status, past and present eGFR levels, and baseline covariates, and given that the patient is still at risk at *t* = *k*. We note that a patient who has started dialysis is considered to remain in dialysis until death or censoring, that is, *A*(*t* + 1) = 1 if *A*(*t*) = 1. The numerator of *w_i(t)* is not required for censoring adjustment, but it is included to increase efficiency (13). Each factor in the numerator is the probability of remaining uncensored, given treatment strategy, given baseline covariates, and given that the replicate is still at risk at *t* = *k*. The intuition behind these weights has been discussed by Cain et al. (13).

In practice, the weights are not known. We estimated them from the data, fitting the pooled logistic models

$$\text{logit}[\text{Pr}_{\text{EXP}}\{C(t) = 0|Y(t) = C(t - 1) = 0, X, V\}] = \eta_0 + \eta_1 t + \eta_2 t^2 + \eta_3' V + \eta_4 X$$

and

$$\begin{aligned} \text{logit}[\text{Pr}\{A(t) = 1|Y(t) = C(t) = \bar{A}(t - 1) = 0, \bar{L}(t), V\}] \\ = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \alpha_3 L(t) + \alpha_4 L^2(t) \\ + \alpha_5 L(t - 1) + \alpha_6 L^2(t - 1) + \alpha_7 L(t - 2) \\ + \alpha_8 L^2(t - 2) + \alpha_8' V \end{aligned}$$

to the expanded data and the original data, respectively. For each replicate *i* and month *t* in the expanded data, we replaced the true probabilities in equation 2 by the predicted probabilities from the fitted models, to obtain an estimate of *w_i(t)*. Figure 1 displays the distribution of the estimated weights. The distribution is skewed, with a strong peak at 1, and an upper tail. A well-known problem with IPW methods is that observations with extreme weights become very influential, which may result in unstable estimates (18–20). To

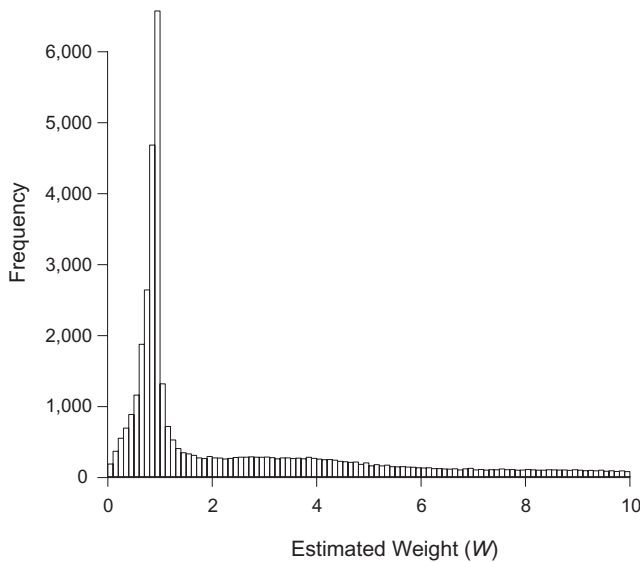


Figure 1. Distribution of the estimated weights (W) for the inverse probability weighting (IPW) method. The patients ($n = 499$ for the IPW analysis and $n = 441$ for the “from initiation” and “from threshold” analyses) represent all members of a Swedish nationwide inception cohort of patients with severe chronic kidney disease, accrued between May 20, 1996, and May 31, 1998, who passed below the baseline level of estimated glomerular filtration rate for the present study (16 mL/minute/1.73 m²) before June 1, 2003.

stabilize estimates, we followed the method of Cain et al. (13) and truncated the weights at 10.00. We tried various cutpoints, up to 100.00, with no noticeable effect on the results (results not shown).

In addition to the artificial censoring of replicates, patients can be “truly” censored, that is, truly lost during follow-up. In principle, IPW can also be used to reduce/eliminate bias due to nonrandom true censoring (20). We did not use IPW to adjust for true censoring, for the following reasons. In our data, only a tiny fraction of patients (4 out of 499) were lost during follow-up. Thus, the risk of bias due to true censoring was negligible, even if this censoring happened to be nonrandom. Furthermore, using IPW to deal with these censorings would have required additional modeling assumptions regarding the censoring mechanism, which if violated could have introduced bias.

MAIN MODEL AND RESULTS

Standard software packages do not allow for time-varying weights in Cox regression models. To overcome this difficulty, we followed the method of Hernán et al. (20) and fitted a weighted pooled logistic regression model to the expanded risk sets. Specifically, using weights $w_i(t)$, we fitted the model

$$\begin{aligned} \text{logit}[\text{Pr}_{\text{EXP}}\{Y(t) = 1 | Y(t-1) = C(t) = 0, X, V\}] \\ = \theta_0 + \theta_1 t + \theta_2 t^2 + \theta_3 t^3 + \theta_4 t^4 + \theta_5' X \\ + \theta_6' X \times t + \theta_7' X \times t^2 + \theta_8' V. \quad (3) \end{aligned}$$

We note that the model in equation 3 differs from the model in equation 1 in two important aspects. Whereas the model in

equation 1 leaves the baseline hazard unspecified, the model in equation 3 assumes that it follows a fourth-degree polynomial. On the other hand, whereas the model in equation 1 assumes proportional hazards, this assumption is relaxed in the model in equation 3 through the terms $\theta_6' X \times t$ and $\theta_7' X \times t^2$. Relaxing the proportional hazards assumption is desirable, since it is unrealistic for dynamic treatment strategies (13).

In the model shown in equation 3, the effect of initiation timing on survival is quantified by the 6-dimensional parameter $\theta = (\theta_5', \theta_6', \theta_7')'$. In particular, there is no effect of timing on survival if $\theta = 0$. However, interpreting a specific estimate $\hat{\theta}$ is difficult because of the high dimensionality of θ . To facilitate interpretation, we estimated standardized survival curves as follows. First, the fitted model in equation 3 was used to estimate conditional survival probabilities $\text{Pr}\{Y(t) = 0 | Y(t-1) = 0, X, V\}$ for each t in $(0, 60)$, for each level of X , and each observed value of V . Next, these were multiplied to yield estimates of $\text{Pr}\{\bar{Y}(t) = 0 | X, V\}$ for each t in $(0, 60)$, for each level of X , and each observed value of V . Finally, these were averaged over the sample distribution of V to yield estimates of $E[\text{Pr}\{\bar{Y}(t) = 0 | X = x, V\}]$ for each t in $(0, 60)$ and for each level $X = x$. Figure 2, part A, displays the result.

We tested whether the timing of dialysis initiation had an effect on survival using the multivariate Wald test statistic $T = \hat{\theta}' \hat{\Sigma}^{-1} \hat{\theta}$, where $\hat{\Sigma}$ is a consistent estimate of the variance matrix Σ . Under $H_0: \theta = 0$, T has an asymptotic χ^2 distribution (6 degrees of freedom). Standard software produces an estimate of Σ . However, this estimate is inconsistent, since it does not take the uncertainty in the estimated weights into account. A consistent estimate can be obtained through a bootstrap procedure (13). To compute a bootstrap estimate for Σ , we drew 10,000 bootstrap samples from the original sample. Each of these bootstrap samples contained the same number of patients as the original sample ($n = 499$), and the patients were drawn with replacement within each bootstrap sample. Each bootstrap sample was analyzed with the IPW method to produce an estimate $\hat{\theta}$. The empirical (over the 10,000 bootstrap samples) variance matrix for $\hat{\theta}$ was used as an estimate of Σ . Using this estimate in the Wald statistic, we obtained $P = 0.12$. We used the same bootstrap procedure to test whether $\theta_6 = \theta_7 = 0$, resulting in $P = 0.05$. Thus, there was borderline-significant evidence against proportional hazards.

The results from the IPW analysis were not directly comparable to the results from the standard analyses, since we used different models. To make results comparable, we reanalyzed the data using the FI and FT methods, now using the model shown in equation 3. Parts B and C of Figure 2 display estimated standardized survival curves. We computed bootstrap P values in the same fashion as for the IPW method, obtaining $P = 1.00$ and $P = 0.30$ for the FI and FT methods, respectively.

The FI method gave almost identical survival curves for the 3 dialysis initiation strategies. On the other hand, the FT method suggested a trend of better survival with later initiation. The IPW method took a position in between; late starters appeared to have better survival than the early and intermediate starters, while the latter two groups appeared to have almost identical survival curves.

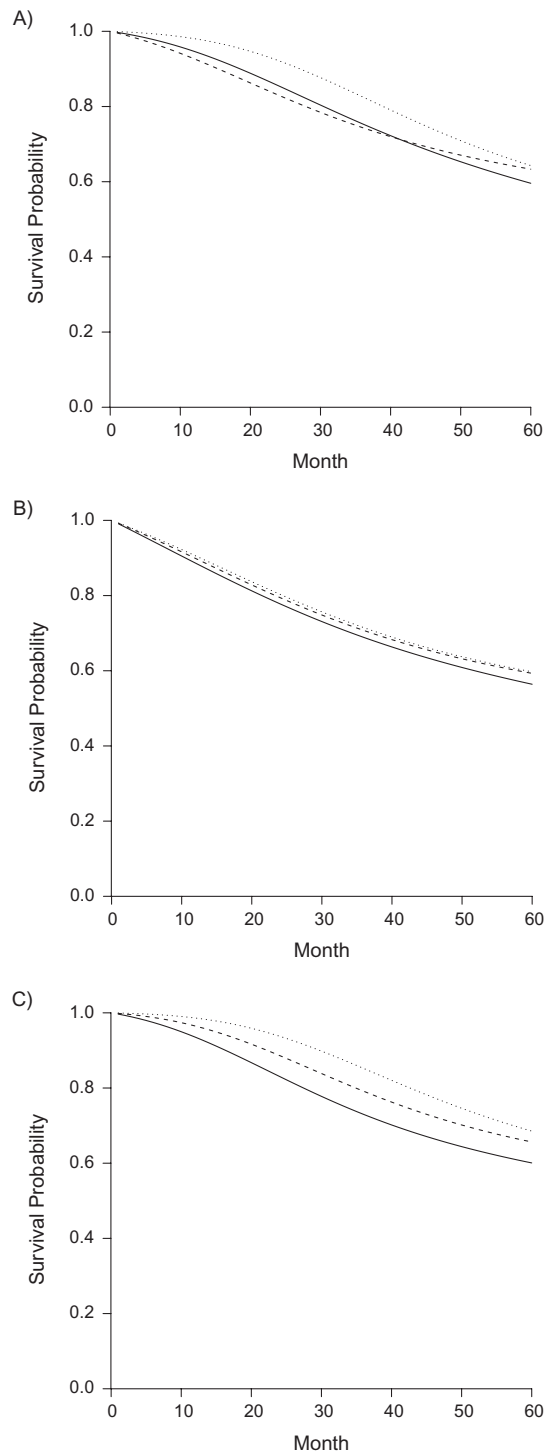


Figure 2. Standardized survival curves for initiation of dialysis (X) derived from regression models using expanded risk sets and the A) inverse probability weighting (IPW), B) “from initiation” (FI), and C) “from threshold” (FT) approaches. Solid lines, “early” initiation of dialysis; dashed lines, “intermediate” initiation; dotted lines, “late” initiation. The patients ($n = 499$ for the IPW analysis and $n = 441$ for the FI and FT analyses) represent all members of a Swedish nationwide inception cohort of patients with severe chronic kidney disease, accrued between May 20, 1996, and May 31, 1998, who passed below the baseline level of estimated glomerular filtration rate for the present study ($16 \text{ mL/minute}/1.73 \text{ m}^2$) before June 1, 2003.

DISCUSSION

In this paper, we have addressed the problem of comparing different dynamic strategies for dialysis initiation. Standard methods are likely to suffer from either lead-time bias (FI) or immortal-time bias (FT). We have shown that both types of bias can be avoided through an appropriate statistical method based on expanded risk sets and IPW.

We applied the IPW method to data from a Swedish study of chronic kidney disease patients and compared it with the standard methods. Using Cox regression, both the FI method and the FT method suggested a trend of better survival with later initiation of dialysis treatment. Using a pooled logistic regression model with nonproportional hazards, this trend persisted only for the FT method. The IPW method suggested a weaker trend than the FT method, with roughly equal survival curves for early and intermediate starters and better survival for late starters. The discrepancy between the FT method and the IPW method indicates that the strong trend suggested by the FT method may be partly explained by immortal-time bias and/or time-varying confounding by GFR. However, none of the results were significant (at the 5% significance level).

One limitation of our study is the possibility of unadjusted confounding. Potential confounders which were not measured include levels of calcium, phosphate, and parathyroid hormone. We also lacked data on more precise levels of albuminuria and updated measures of comorbidity and edema-free body weight. Another important limitation is caused by the inverse relation between serum creatinine levels and protein-energy wasting. Patients with decreased creatinine turnover due to protein-energy wasting have a false-high eGFR and a greater probability of starting dialysis early. Since there is a strong relation between protein-energy wasting and mortality, this may have inflated the observed positive trend for late start of dialysis.

Another limitation is that eGFR was measured at only a few time points and an irregular number of time points for each patient. Thus, we had to interpolate to obtain sufficiently dense and regular GFR trajectories. The choice of interpolation algorithm is somewhat subjective, and the interpolated trajectories can always be questioned to some extent.

Yet another limitation is the large number of exclusions. We had to exclude 8% (72/920) of all patients because no eGFR interpolation could be made (only 1 observed eGFR value). An additional 26% (235/920) of patients had to be excluded because they did not “pass through” baseline ($\text{eGFR} = 16 \text{ mL/minute}/1.73 \text{ m}^2$) during follow-up, although this cannot be attributed to the interpolation algorithm, since any patient with a “true” eGFR trajectory below/above $16 \text{ mL/minute}/1.73 \text{ m}^2$ throughout follow-up would have to be excluded as well. The large number of exclusions induces a larger sampling variability (e.g., larger standard errors). It could also introduce bias if the excluded patients are not exchangeable with the patients who have not been excluded. This could happen, for instance, if the outcome (death) was the main reason for having only 1 observed eGFR value.

We categorized the timing of dialysis initiation into 3 levels. To find the optimal strategy, one would ultimately want to use more categories, or even treat the timing as

a continuous variable. Doing so, however, would require larger data sets and/or stronger modeling assumptions in order to obtain stable estimates.

In conclusion, the question of when to initiate dialysis is important and to a large extent still unsettled, although recent studies such as the IDEAL Trial (1) have given support for a later initiation than is standard clinical practice in the United States today. We believe that the method proposed in this paper will be highly useful in the future search for a more definitive answer.

ACKNOWLEDGMENTS

Author affiliations: Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden (Arvid Sjölander, Olof Nyrén, Rino Bellocco); Department of Medicine, School of Medicine, Vanderbilt University, Nashville, Tennessee (Olof Nyrén); Department of Statistics, University of Milano-Bicocca, Milan, Italy (Rino Bellocco); Department of Clinical Sciences, Intervention and Technology, Karolinska Institutet, Stockholm, Sweden (Marie Evans); and Renal Department, Karolinska University Hospital, Stockholm, Sweden (Marie Evans).

This work was supported by the Swedish Research Council (grant 2008-5375 to A. S.), the Strategic Research Program in Epidemiology at the Karolinska Institutet (O. N.), Fondo Ateneo Ricerca 2009 (R. B.), and Njurfonden (M. E.).

Conflict of interest: none declared.

REFERENCES

- Cooper BA, Branley P, Bulfone L, et al. A randomized, controlled trial of early versus late initiation of dialysis. *N Engl J Med*. 2010;363(7):609–619.
- Korevaar JC, Jansen MA, Dekker FW, et al. When to initiate dialysis: effect of proposed US guidelines on survival. *Lancet*. 2001;358(9287):1046–1050.
- Bonomini V, Vangelista A, Stefoni S. Early dialysis in renal substitutive programs. *Kidney Int*. 1978;13(suppl 8):S112–S116.
- Tattersall J, Greenwood R, Farrington K. Urea kinetics and when to commence dialysis. *Am J Nephrol*. 1995;15(4):283–289.
- Tang SC, Ho YW, Tang AW, et al. Delaying initiation of dialysis till symptomatic uraemia—is it too late? For the Hong Kong Peritoneal Dialysis Study Group. *Nephrol Dial Transplant*. 2007;22(7):1926–1932.
- Traynor JP, Simpson K, Geddes CC, et al. Early initiation of dialysis fails to prolong survival in patients with end-stage renal failure. *J Am Soc Nephrol*. 2002;13(8):2125–2132.
- Sawhney S, Djurdjev O, Simpson K, et al. Survival and dialysis initiation: comparing British Columbia and Scotland registries. *Nephrol Dial Transplant*. 2009;24(10):3186–3192.
- Beddhu S, Samore MH, Roberts MS, et al. Impact of timing of initiation of dialysis on mortality. *J Am Soc Nephrol*. 2003;14(9):2305–2312.
- Kazmi WH, Gilbertson DT, Obrador GT, et al. Effect of comorbidity on the increased mortality associated with early initiation of dialysis. *Am J Kidney Dis*. 2005;46(5):887–896.
- Lassalle M, Labeeuw M, Frimat L, et al. Age and comorbidity may explain the paradoxical association of an early dialysis start with poor survival. *Kidney Int*. 2010;77(8):700–707.
- Stel VS, Dekker FW, Ansell D, et al. Residual renal function at the start of dialysis and clinical outcomes. *Nephrol Dial Transplant*. 2009;24(10):3175–3182.
- Hernán MA, Lanoy E, Costagliola D, et al. Comparison of dynamic treatment regimes via inverse probability weighting. *Basic Clin Pharmacol Toxicol*. 2006;98(3):237–242.
- Cain LE, Robins JM, Lanoy E, et al. When to start treatment? A systematic approach to the comparison of dynamic regimes using observational data. *Int J Biostat*. 2010;6(2):Article 18. (doi:10.2202/1557-4679.1212).
- Orellana L, Rotnitzky A, Robins JM. Dynamic regime marginal structural mean models for estimation of optimal dynamic treatment regimes, part I: main content. *Int J Biostat*. 2010;6(2):Article 8. (doi:10.2202/1557-4679.1200).
- Evans M, Tettamanti G, Nyrén O, et al. No survival benefit from early-start dialysis in a population-based, inception cohort study of Swedish patients with chronic kidney disease. *J Intern Med*. 2011;269(3):289–298.
- Levey AS, Bosch JP, Lewis JB, et al. A more accurate method to estimate glomerular filtration rate from serum creatinine: a new prediction equation. Modification of Diet in Renal Disease Study Group. *Ann Intern Med*. 1999;130(6):461–470.
- Charlson ME, Pompei P, Ales KL, et al. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chronic Dis*. 1987;40(5):373–383.
- Robins JM. Marginal structural models. In: *1997 Proceedings of the American Statistical Association, Section on Bayesian Statistical Science*. Alexandria, VA: American Statistical Association; 1998:1–10.
- Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology*. 2000;11(5):550–560.
- Hernán MA, Brumback B, Robins JM. Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology*. 2000;11(5):561–570.