Dialysis modality, vascular access and mortality: an application of marginal structural models to data from a clinical registry

Jessica Kasza[1,2]
Rory Wolfe[1,2]
Kevan Polkinghorne[1]

1. Department of Epidemiology and Preventive Medicine, Monash University
2. ViCBiostat: Victorian Centre for Biostatistics

ISCB 2014

Outline

1. Dialysis and the ANZDATA Registry

2. Unmeasured confounding

3. Clustering by dialysis centre

Outline

1. Dialysis and the ANZDATA Registry
2. Unmeasured confounding
3. Clustering by dialysis centre

What happens when kidneys fail?

End-stage renal disease treatment:

- Kidney transplantation;
- Dialysis: blood is filtered artificially to remove waste products.

Two forms of dialysis (dialysis modalities):

- Haemodialysis (HD)
  - **Home HD**: performed by the patient at home;
  - **Facility HD**: performed in a hospital/dialysis centre.
  - Vascular access types:
    - Arterio-venous fistula or graft: **AVF**/**AVG**
    - Central venous catheter: **CVC**
- **Peritoneal dialysis** (PD)

Which modality and VA combination is best?

Aim: determine which modality and VA combination is the best for patient survival.

Treatments of interest:

$$\mathcal{A} = \left\{ \begin{array}{cc} \text{Home HD AVF/AVG} & \text{Facility HD AVF/AVG} \\ \text{Facility HD CVC} & \text{PD} \end{array} \right\}$$

RCTs difficult in this context $\Rightarrow$ turn to observational data.

Which modality and VA combination is best?

Aim: determine which modality and VA combination is the best for patient survival.

Treatments of interest:

$$\mathcal{A} = \left\{ \begin{array}{cc} \text{Home HD AVF/AVG} & \text{Facility HD AVF/AVG} \\ \text{Facility HD CVC} & \text{PD} \end{array} \right\}$$

RCTs difficult in this context $\Rightarrow$ turn to observational data.

ANZDATA: Australian and New Zealand Dialysis and Transplant Registry

- Collects data from all dialysis patients in Australia and NZ.
- Changes between PD, home HD, facility HD recorded as they occur.
- Data (including comorbidities, vascular access) collected at dialysis start and at yearly surveys.

The ANZDATA dataset used for analysis

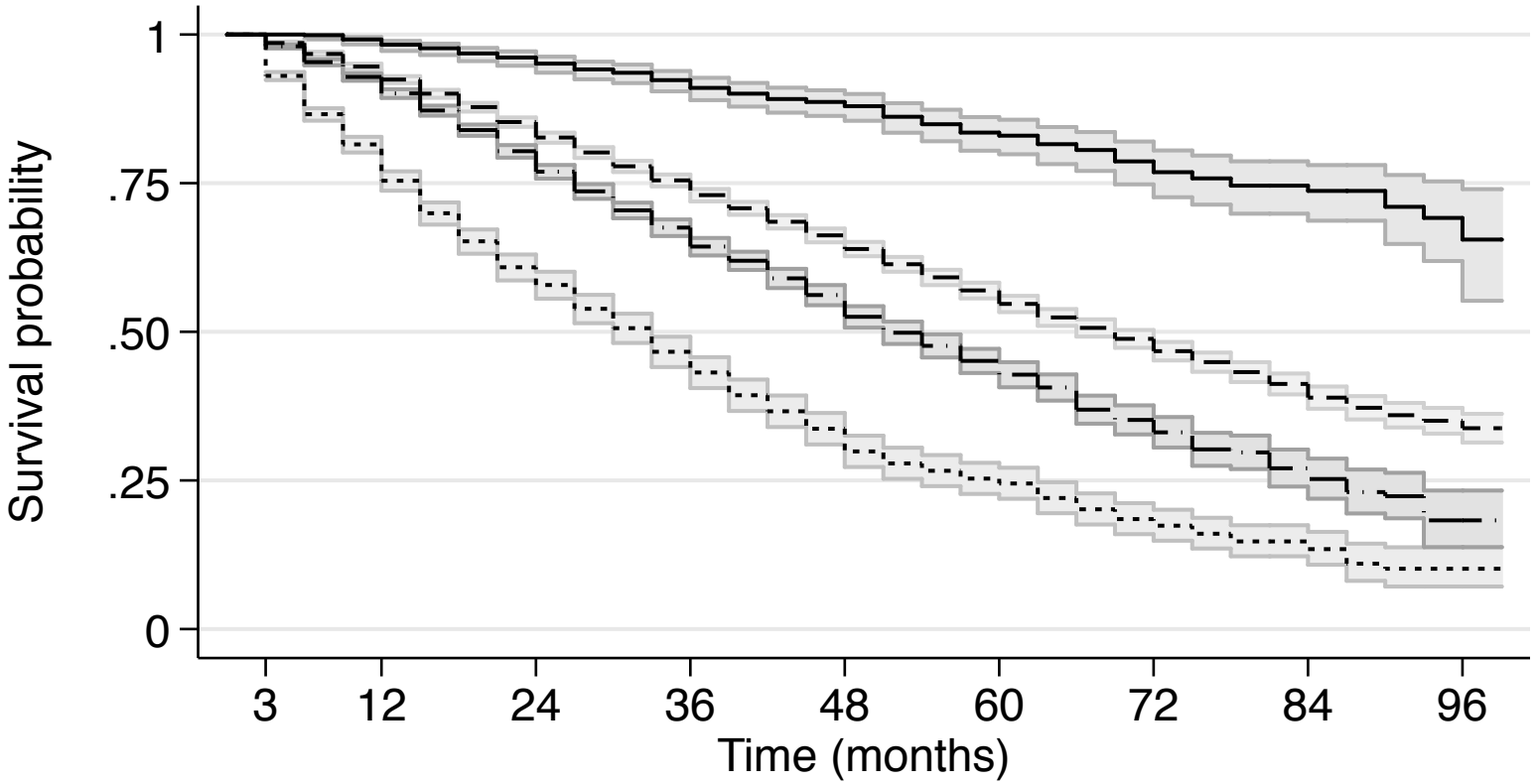
All patients commencing dialysis between October 1 2003 and December 31 2011, undergoing at least 90 days of dialysis.
20,191 patients:

- 210,741 90-day periods of follow-up
- 6,971 deaths
- 2,966 kidney transplants
- 267 recovered kidney function

Over their treatment course, **30% of all patients had changes** in dialysis modality/VA

- Modality/VA choice thought to be affected by, and affect, comorbidities (e.g. coronary artery disease).
- We use MSMs to estimate the effect of modality/VA on mortality.

Unadjusted Kaplan-Meier survival curves



Number at risk									
HD home AVF/AVG	712	1007	933	703	494	320	173	85	19
HD facility AVF/AVG	6936	7745	6105	4465	3058	1888	1037	466	112
HD facility CVC	5409	1534	775	481	309	185	100	34	3
PD	7051	5512	3389	1906	1029	506	202	60	10

HD home AVF/AVG
HD facility AVF/AVG
HD facility CVC
PD

Applying MSMs to ANZDATA

Problems:

1. ANZDATA is a registry, so set of **measured confounders is limited**.
2. Patients are **clustered** within dialysis centres.

Applying MSMs to ANZDATA

Problems:

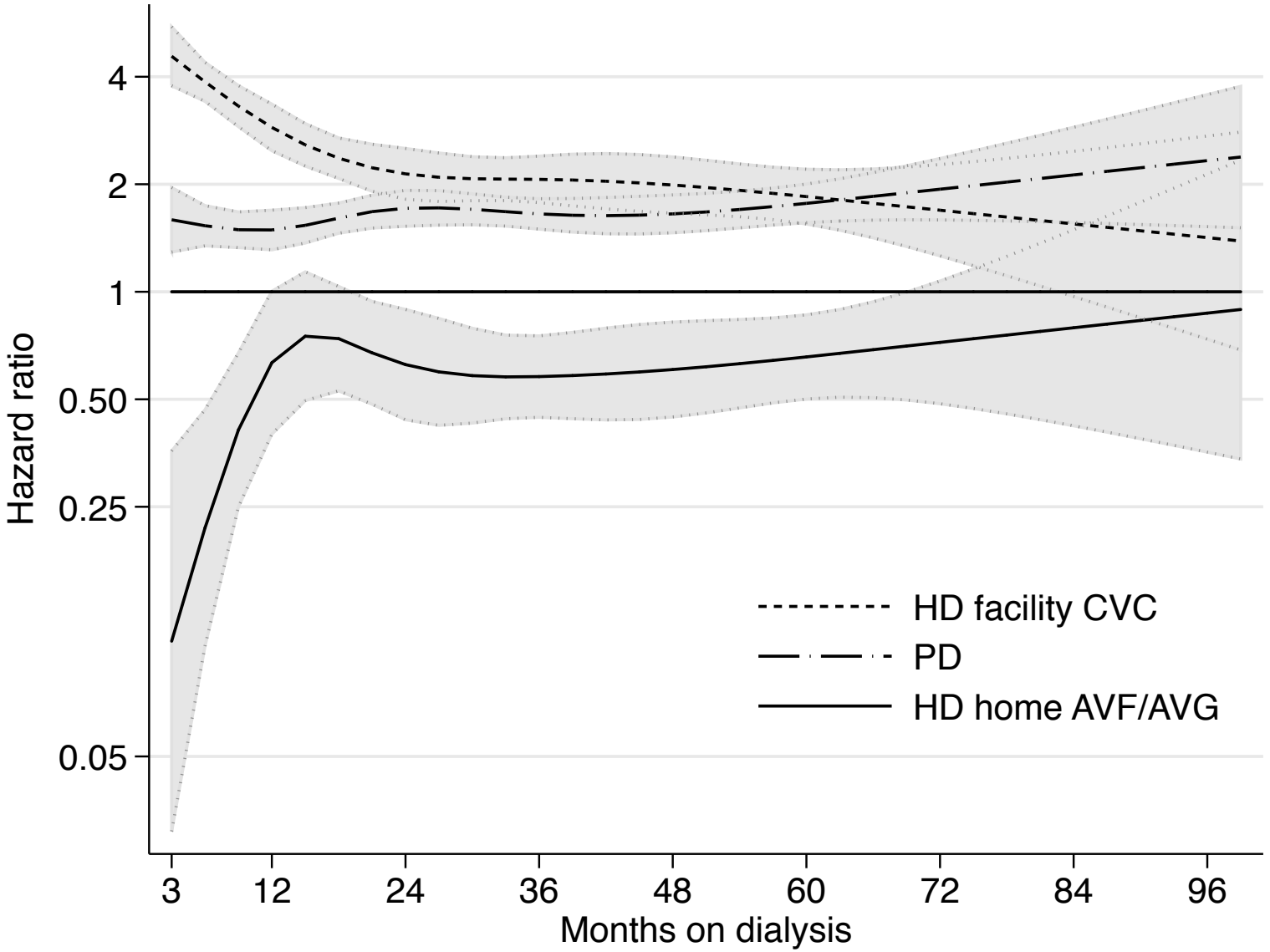
1. ANZDATA is a registry, so set of **measured confounders is limited**.
2. Patients are **clustered** within dialysis centres.

Let's ignore these problems for the moment, and fit a pooled logistic regression model:

$$\text{logit}\left[P(D_i(t) = 1 | D_i(t-1) = 0, Rx_i(t), V_i)\right] = \beta_0(t) + \beta_1(t) Rx_i(t) + \beta_2 V_i,$$

where the observation of each patient at each period is weighted by the stabilised inverse probability of treatment and censoring weight.

Estimated HRs, relative to facility HD AVF/AVG



Outline

1. Dialysis and the ANZDATA Registry

2. Unmeasured confounding

3. Clustering by dialysis centre

How sensitive are estimated HRs to unmeasured confounding?

- Following Brumback et al. (2004), develop a confounding function for each treatment $a \in \mathcal{A}$, $c(a)$.

$$c(a) = \frac{P(D_a(t) = 1 | A(t) = a, V = v)}{\frac{1}{\sum_{a^* \in \mathcal{A} \setminus \{a\}} P(a^*)} \sum_{a^* \in \mathcal{A} \setminus \{a\}} P(a^*) P(D_a(t) = 1 | A(t) = a^*, V = v)},$$

$$P(a^*) = P(A(t) = a^* | V = v).$$

- Informal interpretation of $c(a)$:
  HR of death comparing patients on a to those not on a, had those patients been (contrary to the fact!) on a.

How sensitive are estimated HRs to unmeasured confounding?

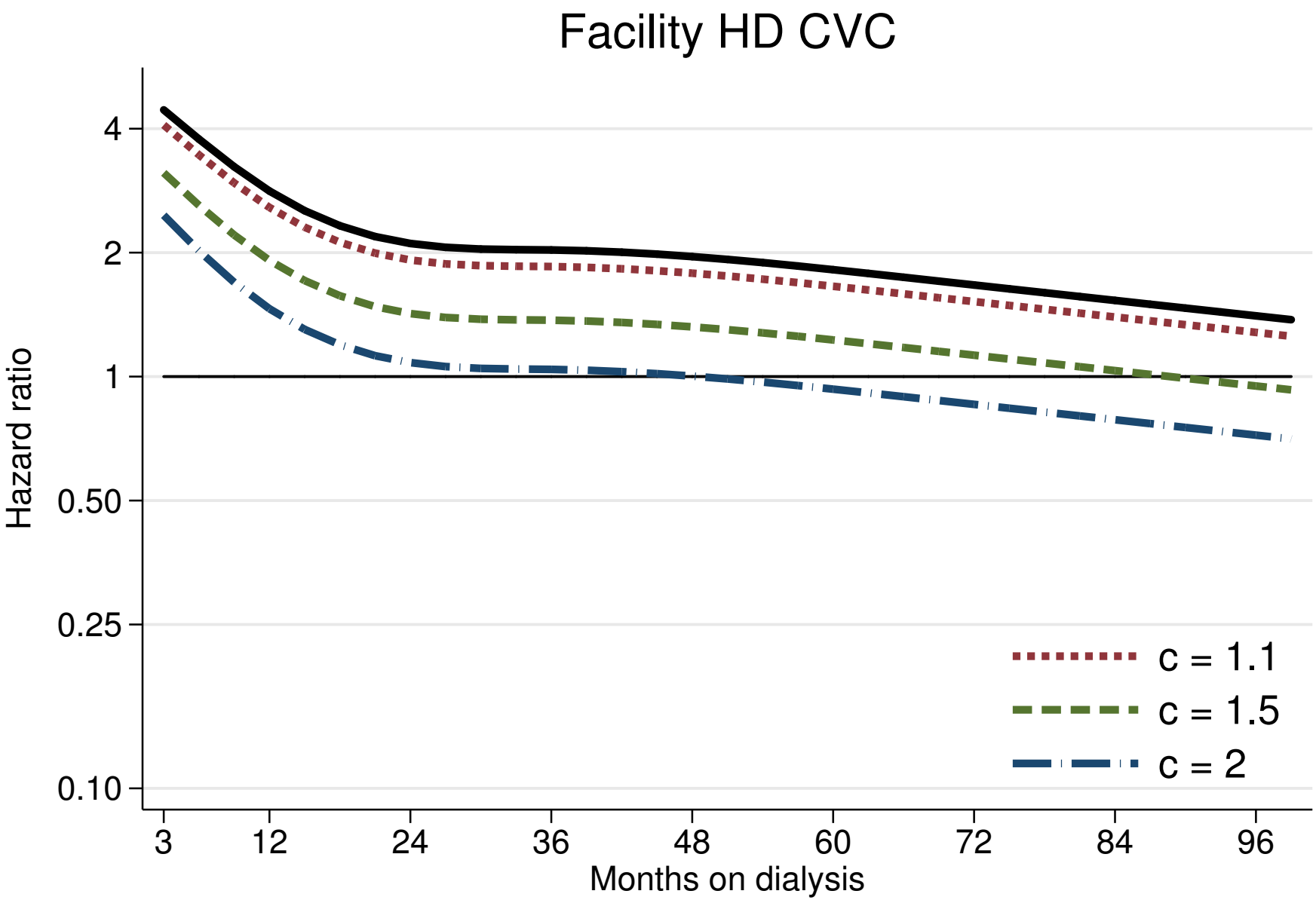
- Following Brumback et al. (2004), develop a confounding function for each treatment $a \in \mathcal{A}$, $c(a)$.

$$c(a) = \frac{P(D_a(t) = 1 | A(t) = a, V = v)}{\frac{1}{\sum_{a^* \in \mathcal{A} \setminus \{a\}} P(a^*)} \sum_{a^* \in \mathcal{A} \setminus \{a\}} P(a^*) P(D_a(t) = 1 | A(t) = a^*, V = v)},$$

$$P(a^*) = P(A(t) = a^* | V = v).$$

- Informal interpretation of $c(a)$:
  HR of death comparing patients on a to those not on a, had those patients been (contrary to the fact!) on a.
  - $c(a) = 1$: no difference in the risk of death of patients on a and those not on a.
  - **c(Facility HD CVC) > 1**: Facility HD CVC patients have a greater risk of death than those patients on PD/ Home HD/ Facility HD AVF/AVG (had those patients been on Facility HD CVC).

HRs accounting for unmeasured confounding



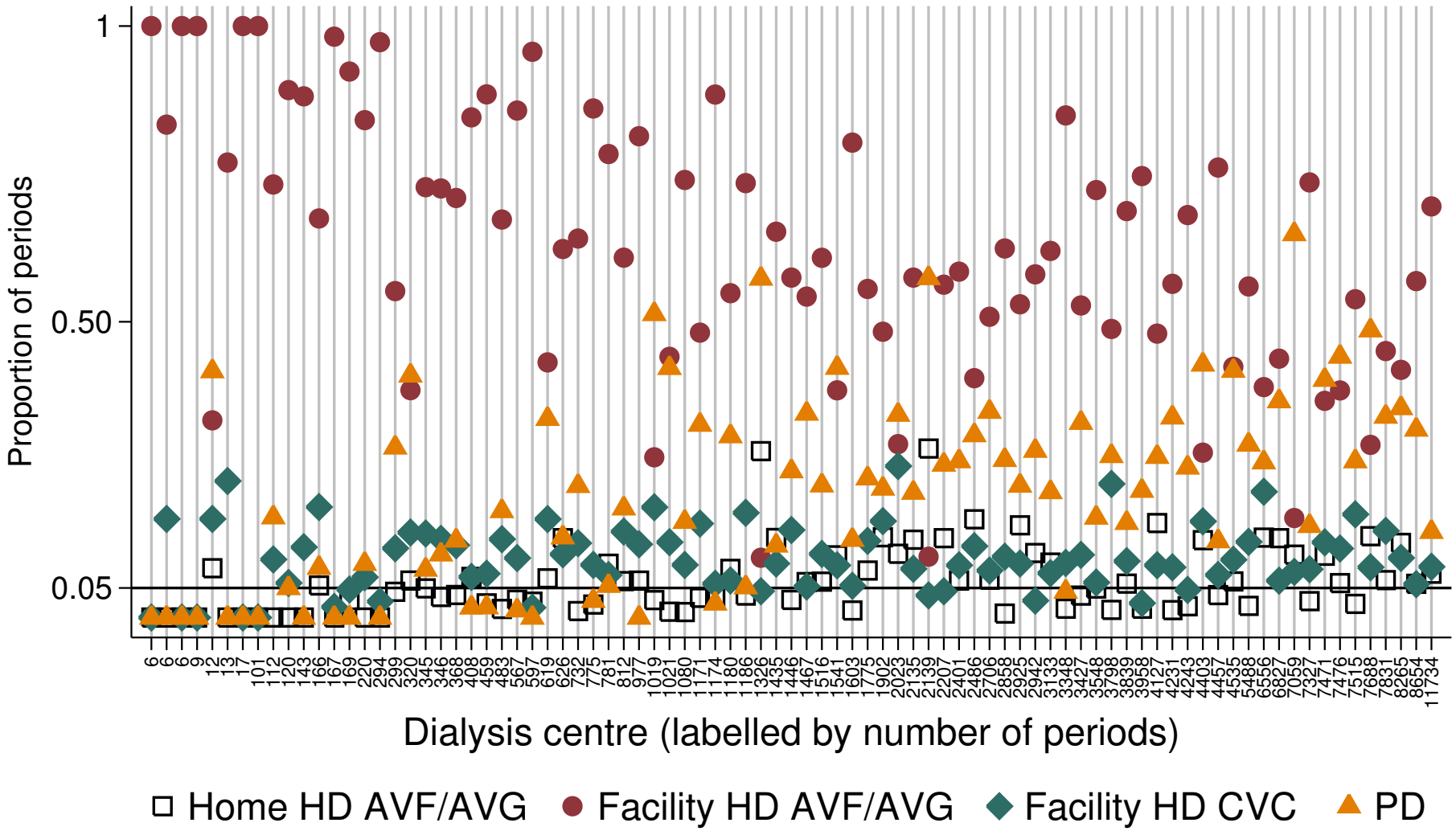
Outline

1. Dialysis and the ANZDATA Registry
2. Unmeasured confounding
3. Clustering by dialysis centre

Clustering by dialysis centre

- All patients, even those undergoing home-based treatment, have a dialysis centre which is responsible for administering their treatment.
  - 85 dialysis centres are represented in our dataset.
  - There are differences in practice and survival across centres.
- An extreme difference: **not all dialysis types are available/represented in all centres** (or only occur rarely within a centre).
  - In violation of the positivity assumption...

Clustering of treatments within the 85 centres



Dealing with this violation of the positivity assumption:

1. Modify the set of centres:
   - Include only those centres in which all treatments are possible (or probable - occurring at least 5% of the time).

Dealing with this violation of the positivity assumption:

1. Modify the set of centres:
   - Include only those centres in which all treatments are possible (or probable - occurring at least 5% of the time).
2. Modify the set of treatments within each centre:
   - Limit the sets of treatments at each centre to those possible/probable.
   - To prevent unavailable treatments being assigned non-zero probabilities, use an alternative-specific conditional logit (McFadden's choice) model to estimate propensity scores.

Dealing with this violation of the positivity assumption:

1. Modify the set of centres:
   - Include only those centres in which all treatments are possible (or probable - occurring at least 5% of the time).
2. Modify the set of treatments within each centre:
   - Limit the sets of treatments at each centre to those possible/probable.
   - To prevent unavailable treatments being assigned non-zero probabilities, use an alternative-specific conditional logit (McFadden's choice) model to estimate propensity scores.

Both approaches: to account for unexplained variation between centres, include fixed effects for centres in treatment, censoring and survival models.

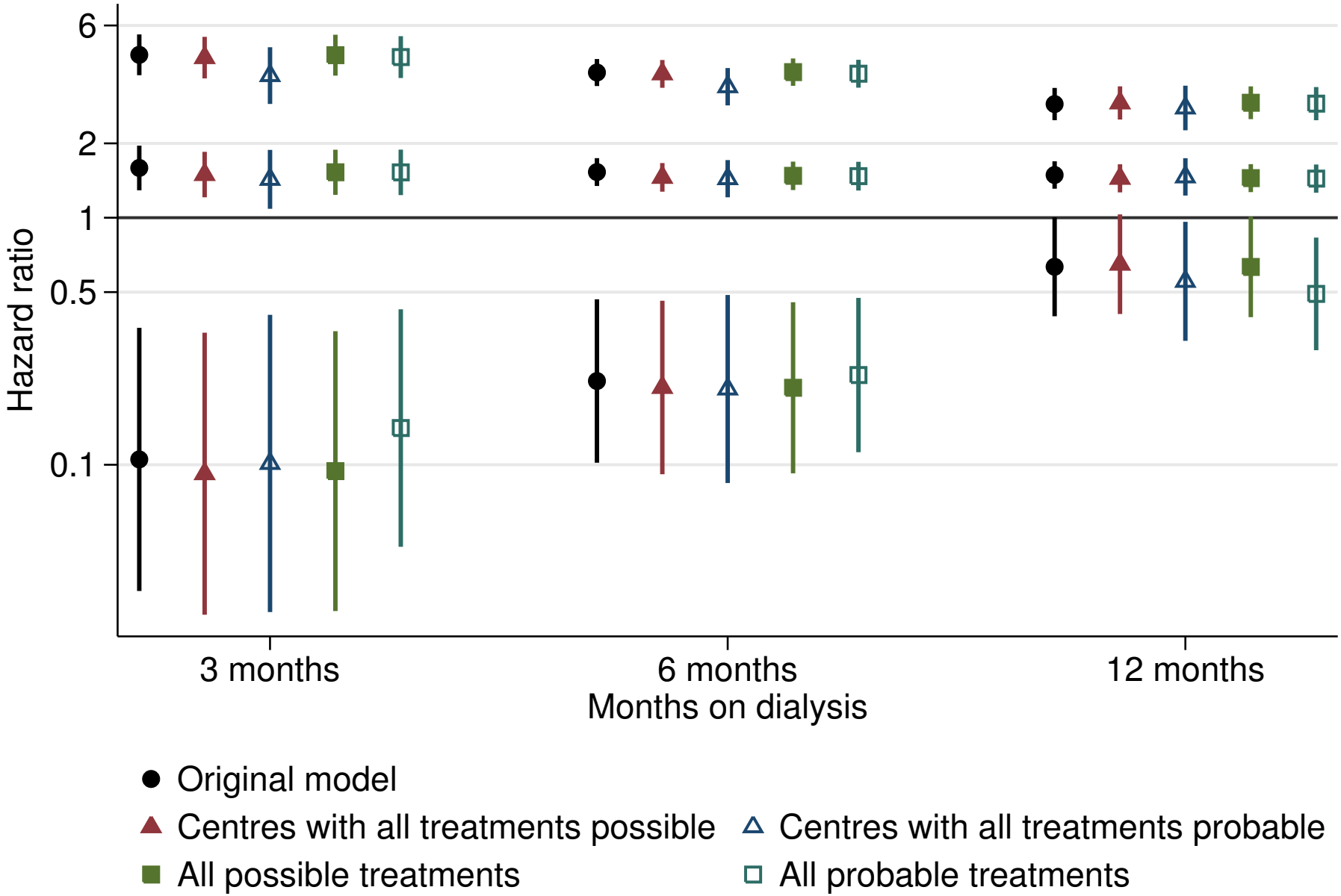
Analyses accounting for clustering by centre C_j, treatments $a \in \mathcal{A}$

Centres with only one treatment possible/probable must be excluded from all analyses.

- Exclude 11 C_j with < 150 periods (545 periods excluded in total)
- Leaves **74 centres, 208132 periods**

	Restriction	Included centres	Total no. of periods
1	Centres w/ $P(A = a \mid C_j) > 0, \forall a \in \mathcal{A}$	68	192166
2	Centres w/ $P(A = a \mid C_j) > 0.05, \forall a \in \mathcal{A}$	34	127888
3	Treatments w/ $P(A = a \mid C_j) > 0$	74	208132
4	Treatments w/ $P(A = a \mid C_j) > 0.05$	70	206905

HRs accounting for clustering by centre



Discussion

- Effect of unmeasured confounding supposed to be constant over time:
  - Possible that groups start off as quite different, but become more similar as time spent on dialysis increases.
  - Time-varying confounding should be corrected for, but choice of appropriate time-varying confounding function is difficult.
- Clustering is not often accounted for in the application of MSMs:
  - If treatment options are restricted (instead of centres): HRs defined only for those centres in which the treatment is available.
  - Accounting for clustering did not markedly change conclusions.
- Research into accounting for differential amounts of unmeasured confounding across clusters ongoing.

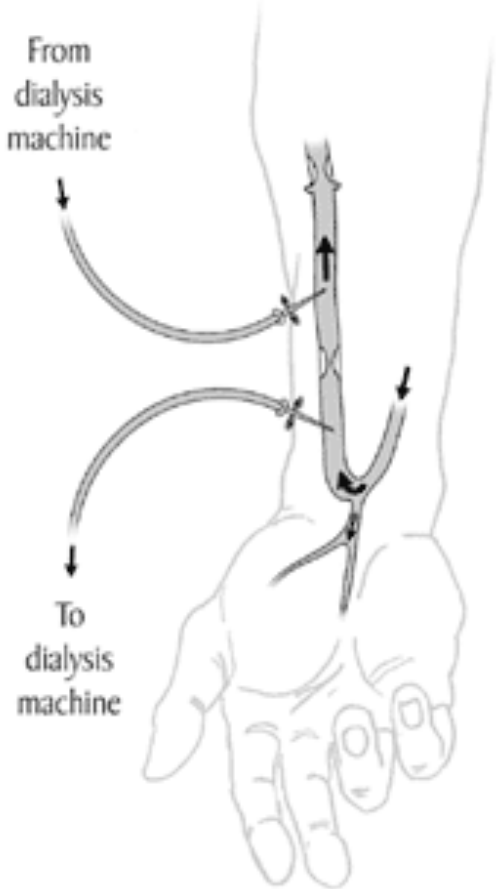
Acknowledgements

- ANZDATA Registry
- Monash University, School of Public Health and Preventive Medicine Travel Grant
- Victorian Centre for Biostatistics

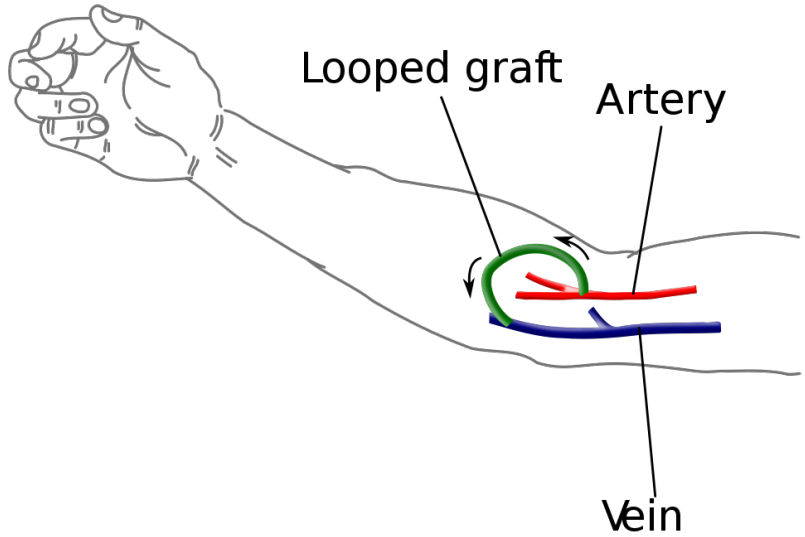
VICBiostat

HD vascular access types

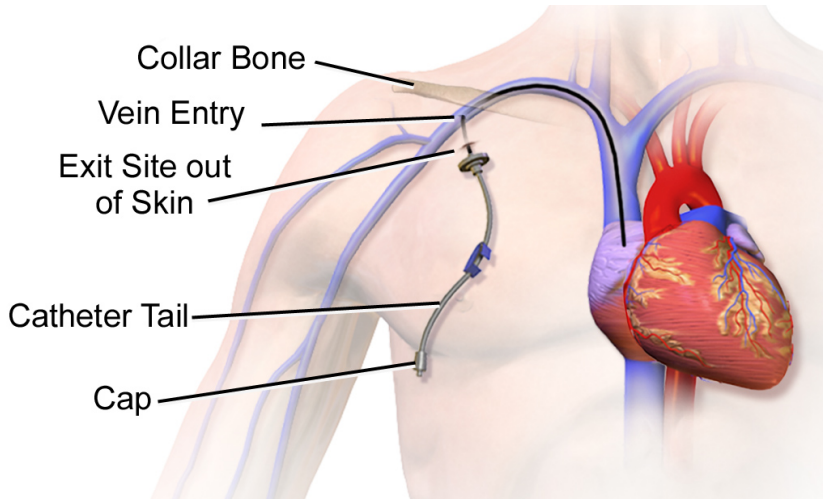
Arterio-venous fistula (AVF):



Arterio-venous graft (AVG):



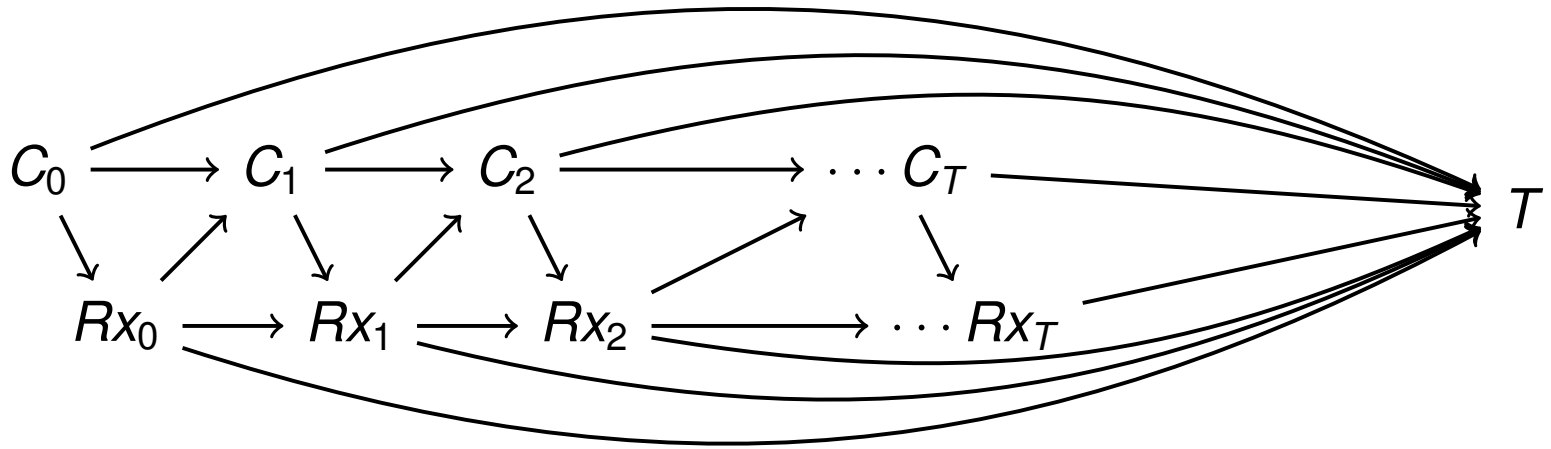
Central venous catheter (CVC):



Patients may change dialysis modality/VA

Comorbidities and BMI are time-varying confounders that are affected by previous dialysis modality and VA:

- on the causal pathway between dialysis type and death.



- If comorbidity history is conditioned on, effect of dialysis modality acting through comorbidities is blocked.
- We used marginal structural models to estimate the causal effect of dialysis modality on survival.

Imputation of VA change times

Dates of change between PD/home HD/facility HD are recorded:

- **Problem:** VA change times are not recorded!
  - We impute these stochastically, using a distribution estimated from the data.
  - 50 sets of VA change times imputed, and Rubin's rules used to combine estimates.

Table: Number of periods, deaths, transplants/regain function for each exposure category, averaged over the 50 simulations: mean, (sd).

	90-day periods	Deaths	Transplants/ regain function
Home HD AVF/AVG	16,073 (73)	152 (2)	474 (3)
Facility HD AVF/AVG	109,968 (68)	3,107 (9)	1,316 (5)
Facility HD CVC	21,517 (62)	1,493 (8)	321 (5)
PD	61,134 (1)	2190 (0)	1082 (0)

How sensitive are our conclusions to unmeasured confounding?

Modifying Brumback et al. SiM (2004):

- Dialysis type at time t denoted by $A(t)$, taking values $a \in \mathcal{A}$, baseline variables V
- $D(t) = 1$ if death at time t
- $D_a(t)$: **counterfactual** outcome had this patient received dialysis type a.

For each $a \in \mathcal{A}$, confounding function:

$$c(a, v, t) = \frac{P(D_a(t) = 1 | A(t) = a, V = v)}{\frac{1}{\sum_{a^* \in \mathcal{A} \setminus \{a\}} P(a^*)} \sum_{a^* \in \mathcal{A} \setminus \{a\}} P(a^*) P(D_a(t) = 1 | A(t) = a^*, V = v)}$$

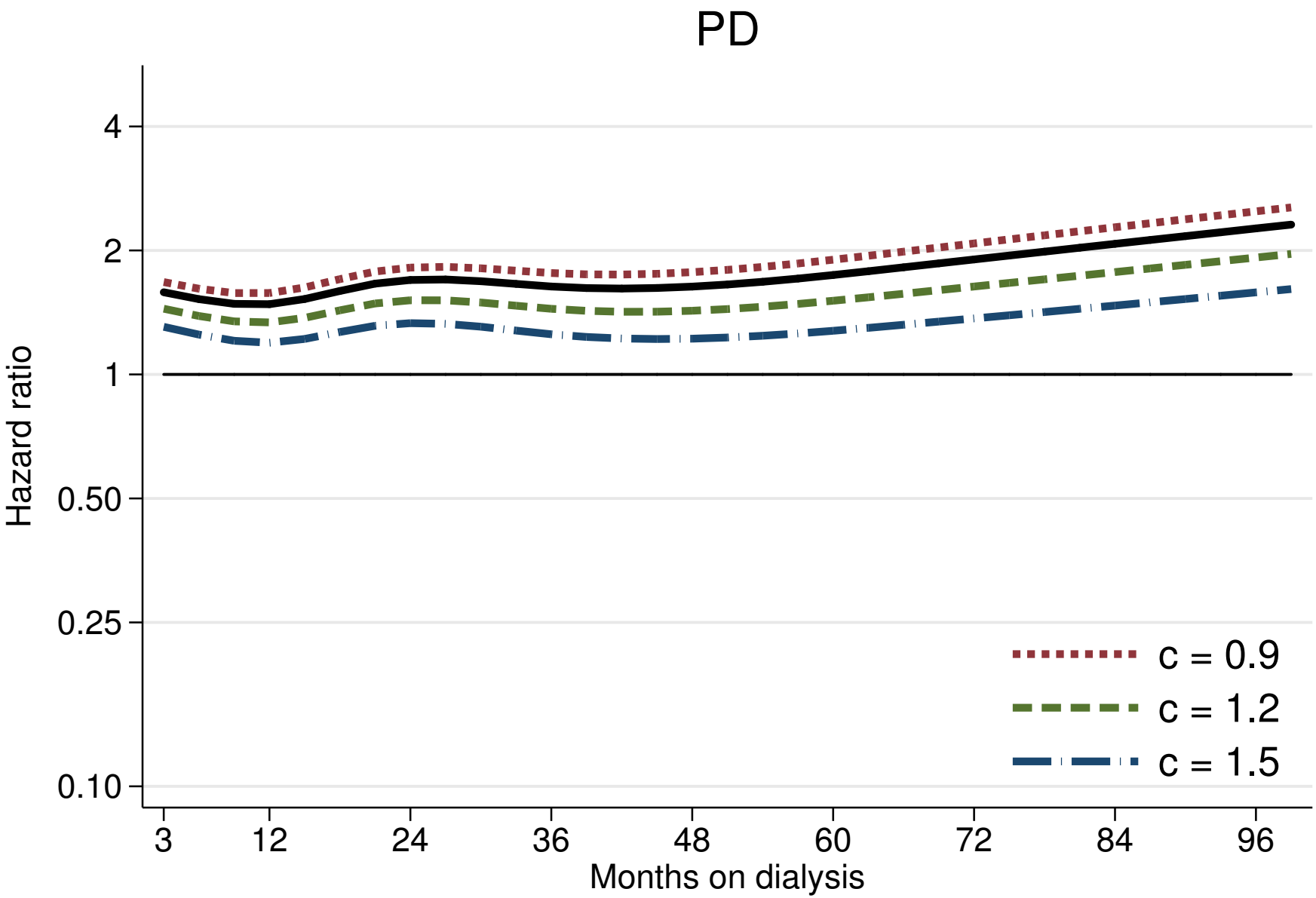
$$P(a^*) = P(A(t) = a^* | V = v).$$

Informal interpretation as an odds ratio.

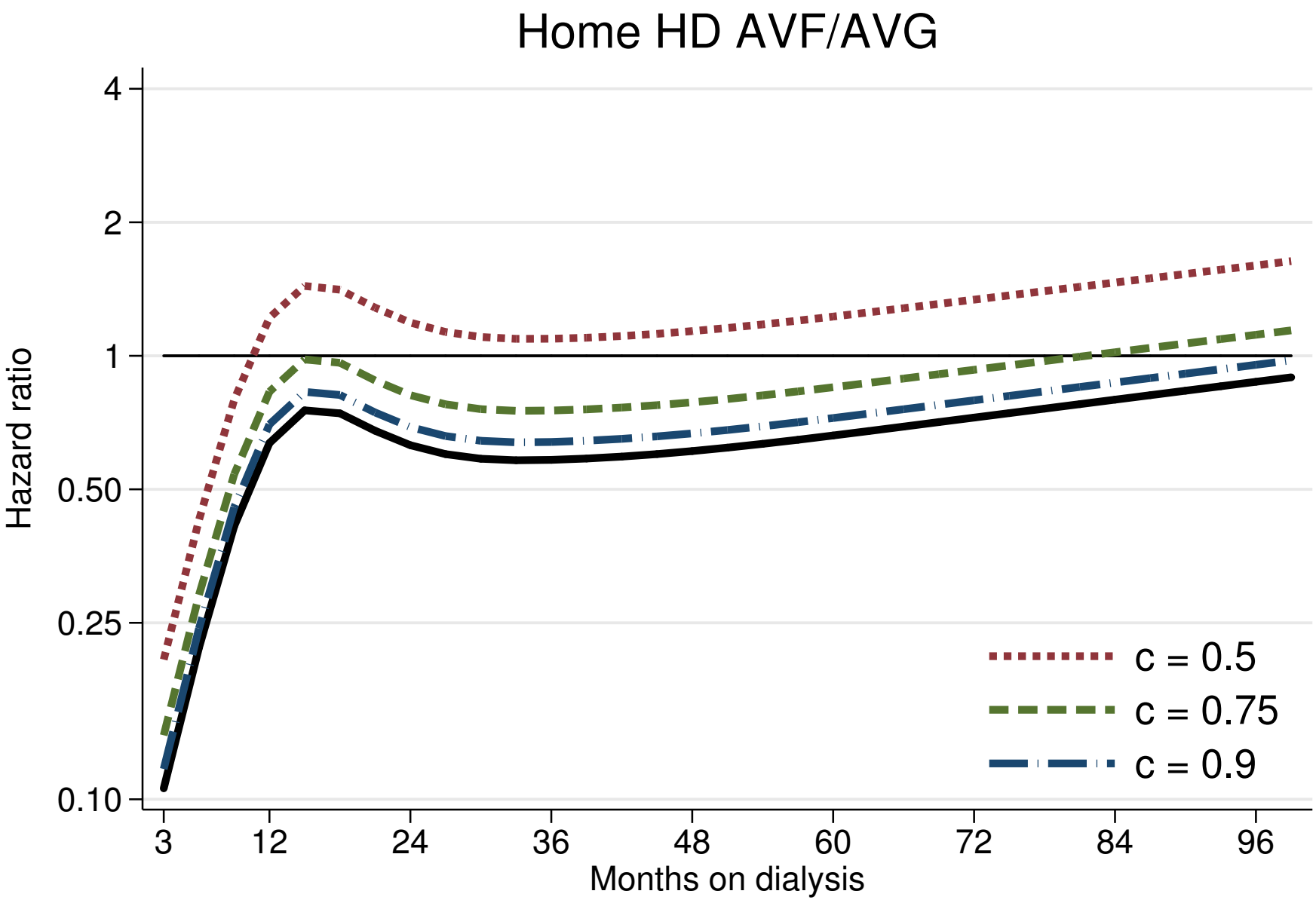
Assessing the impact of unmeasured confounding

- $c(a, v, t) = 1$: no difference in the risk of death of patients on a and those not on a.
- e.g. Facility HD CVC patients thought to be less healthy than other patients on average (controlling for what is already measured)
  - $\mathbf{c}(\textbf{Facility HD CVC}, \mathbf{v}, \mathbf{t}) > \mathbf{1}$: Facility HD CVC patients have a greater risk of death than those patients on PD/ Home HD/ Facility HD AVF/AVG (had those patients been on Facility HD CVC).
- Can then obtain an expression for the amount of bias due to unmeasured confouding.

HRs accounting for unmeasured confounding



HRs accounting for unmeasured confounding



Positivity assumption in the presence of clustering

- Usual positivity assumption, patient i:

$$\frac{P(A_i(t) = a_i|\bar{A}_i(t), V_i)}{P(A_i(t) = a_i|\bar{A}_i(t), \bar{L}_i(t), V_i)} < \infty, \quad \forall a_i \in \mathcal{A}$$

- Positivity assumption in the presence of clustering:
  patient i in centre C_j, $\mathcal{A}_j$ = set of treatments available in C_j:

$$\frac{P(A_{ij}(t) = a_{ij}|\bar{A}_{ij}(t), V_{ij}, C_j)}{P(A_{ij}(t) = a_{ij}|\bar{A}_{ij}(t), \bar{L}_{ij}(t), V_{ij}, C_j)} < \infty, \quad \forall a_{ij} \in \mathcal{A}_j$$

Including laboratory measurements

- Calcium (mmol/l);
- Phosphate (mmol/l);
- Haemoglobin (g/l);
- EPO agent (yes or no);
- Ferritin (ug/l);
- % saturation iron.

Lab measurements recorded at surveys:

- **not at dialysis start**.
- Don't necessarily correspond to labs at treatment change times.

Idea: consider only those 4905 patients starting dialysis within 90 days of a survey.

- Maybe these measurements are highly correlated with measurements at dialysis start...
  - **No.** Labs are quite variable during the initial months of analysis.

Solution:
Start observation time from the first survey occurring $\geq$ 90 days after dialysis start.

References I


B. A. Brumback, M. A. Hérnan, S. J. P. A. Haneuse & J. M. Robins
Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures.
Statistics in Medicine, 23:749–767, 2008.


R. M. Daniel, S. N. Cousens, B. L. De Stavola, M. G. Kenward & J. A. C. Sterne.
Methods for dealing with time-dependent confounding.
Statistics in Medicine, 32:1584–1618, 2013.


Z. Fewell, M. A. Hérnan, F. Wolfe, K. Tilling, H. Choi & J. A. C. Sterne.
Controlling for time-dependent confounding using marginal structural models.
The Stata Journal, 4(4): 402-420, 2004.

References II


M. A. Hérnan, B. Brumback & J. M. Robins .
Marginal structural models to estimate the causal effects of Zidovudine on the survival of HIV-positive men.
Epidemiology, 11(5):561-570, 2000.


M. A. Hérnan, E. Lanoy, D. Costagliola & J. M. Robins .
Comparison of dynamic treatment regimes via inverse probability weighting.
Basic & Clinical Pharmacology & Toxicology, 98:237-242, 2006.