APPENDIX A

Analyses of time-updated AF using Marginal Structural Models

Our analyses of time-updated AF utilized marginal structural models (MSM) which applies inverse probability weighting in a discrete time failure model.(1) A substantial body of work has emerged demonstrating the usefulness of statistical tools like marginal structural models in the areas of HIV and CKD.(2, 3) Briefly, MSM is a two-step approach wherein models were first fit to predict the probability of AF during follow-up (i.e., the exposure of interest) and the probability of non-censoring, and second, inverse-probability weighted structural models were fit for the outcomes. In both steps, the data structure was set up so that each individual could contribute multiple records, each half a year in length. Specifically, we defined each of the 6 month periods starting from each participant's baseline visit or half-year anniversaries of the baseline visit. As a result, an individual contributed one record if the follow-up period is <6 months, 2 records if between 6 and 12 month, 3 records if between 12 and 18 months, and so on. The maximum of number of periods in this dataset is 19.

In the first part of the first step (i.e., calculating the exposure weights), the indicator of incident AF at the end of each of the 6-month period was defined. A logistic regression model was fit on the AF incident indicator with adjustment for some covariates from baseline such as age, gender, race/ethnicity, and clinical site, and some concurrent covariates such as proteinuria, eGFR, tobacco use, heart failure, coronary heart disease, hypertension, diabetes, systolic blood pressure, BMI, hemoglobin, diuretic use, and ACE/ARB use. Quadratic spline terms for urine PCR and eGFR were incorporated to account for nonlinear relationships. We also assumed a different intercept for each of the periods and collapsed the last three periods as the numbers of observations were small. For each period, the concurrent covariates were defined using the values at the beginning of the period. We assumed the value of each covariate stayed the same until being updated with information from annual clinic visit or semi-annual phone visit.

We calculated the weights that were used in the first step based on the predicted probability of incident AF. The weight for a particular 6-month period was calculated as one over the cumulative probabilities of the observed AF incident up to that 6-month period, calculated as the product of the probabilities of the observed AF incident up to that 6-month period.

To improve model stability and statistical efficiency, we stabilized the weights by multiplying the estimated probability of observed AF incident conditional on baseline predictors only (i.e., baseline: age, gender, race/ethnicity, clinical site, proteinuria(quadratic spline), eGFR (quadratic spline), tobacco use, heart failure, coronary heart disease, hypertension, diabetes, systolic blood pressure, BMI, hemoglobin, diuretic use, and ACE/ARB use.(1, 3) This was done by fitting a second model for incident AF using the baseline predictors only.

In the second part of the first step, we considered another weight due to censoring. Participants might be censored due to different reasons such as withdrawal (n=125), death (n=314), and

reaching the end of study (n=1886) before having an ESRD event. Those lost to follow-up may be systematically different from those who were adherent to the study schedule. Similarly, subjects who withdrew from the study or died were different from those who were still in the study. To correct the possible bias due to informative censoring, we applied similar modeling strategy described above for AF. Specifically, we fit two logistic regression models for the binary indicator of censoring using time-updated predictors and the baseline predictors only, respectively. We called this method the "two-level model". The only difference between the models in this part and the first part in the first step was that we included the incident of AF indicator in the models. We calculated the censoring weight as the cumulative product of the predicted probability of censoring using baseline only predictors divided by the cumulative product of the predicted probability of censoring using the additional time-updated predictors. The final weight for each observation in the second step is the product of the exposure weight multiplied by the censoring weight.

As a sensitivity analysis, we also tried to model different reasons for censoring via a multinomial logistic regression. Specifically, in this model, the outcome has four levels (censored due to withdrawal, censored due to death, censored due to reaching end of study, and not censored) and we called this the "four-level model". For this model, due to the small number of outcomes in some of the categories, the model does not converge if we assume a different intercept for each of the periods. The model worked when we assumed the same intercept for all of the 6-month periods. The weights derived from the "four-level model" are similar to those derived from the "two-level model" strategy and results are similar as well (data not shown). We decided to report the results from the "two-level model".

The final weights are the product of the weights derived from the first and second part in the first step. We did not truncate the weights as the maximum of the weights is 6.9.

One key assumption for MSM is the so-called positivity assumption. Specific to this study, it requires that the probability of developing incident AF and not developing incident AF conditional on any combination of the predictors has to be strictly positive (i.e., for any stratum corresponding to a particular combination of the predictors, there are observed individuals having developed incident AF or having not developed incident AF). Assuming that no individual was observed in one of the two categories for incident AF for a particular combination of the predictors, we would be unable to make any inference on this particular stratum (i.e., no causal contrast could be made with respect to the other incident AF category). Given the large sample size we had in our study, we felt this assumption was satisfied, confirmed by the distribution of the weights derived in the first part in the first step. The largest weight we observed for a particular observation in that was 3.9. The weight for a particular observation will go to infinity if the positivity assumption is violated.

In the second step, we fit a discrete time failure model for the outcome, i.e., ESRD by applying the final weight derived in the first step using the GENMOD procedure in SAS 9.3 (SAS

Institute, Cary, NC). ESRD was defined using the ESRD status at the end of each 6-month period. Each 6-month period contributed one observation to the data. Since we adjusted for baseline covariates in the numerators for the weight models by assuming at each time, the AF status is randomized within the levels of the included baseline covariates.(4) Cole and Hernán (2008) suggest to adjust for these covariates in the final model. As a result, the causal effect estimated will not be unconditional (marginal) but rather, conditional on the included covariates.

To test for interactions between the incident AF indicator and baseline covariates, as recommended by Robins et al,(1) we added the interactions in the final model without changing the weight model. For stratified analysis, we refit both the weight models in the first and second steps and the final model stratified by baseline covariates.

REFERENCES

1. Robins JM, Hernan MA, Brumback B: Marginal structural models and causal inference in epidemiology. *Epidemiology,* 11**:** 550-560, 2000
2. Brunelli SM, Joffe MM, Israni RK, Yang W, Fishbane S, Berns JS, Feldman HI: History-adjusted marginal structural analysis of the association between hemoglobin variability and mortality among chronic hemodialysis patients. *Clin J Am Soc Nephrol,* 3**:** 777-782, 2008
3. Hernan MA, Brumback B, Robins JM: Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men. *Epidemiology,* 11**:** 561-570, 2000
4. Cole SR, Hernan MA: Constructing inverse probability weights for marginal structural models. *Am J Epidemiol,* 168**:** 656-664, 2008