**Supplementary Table 1**. Descriptive characteristics of the Medicare cohort, by stroke status.

| Characteristic[1] | No Stroke | Hemorrhagic Stroke | *P* Value[2] | Ischemic Stroke | *P* Value[3] |
|---|---|---|---|---|---|
| No. of persons | 254,937 (96.0) | 1802 (0.7) | — | 8946 (3.4) | — |
| Age, yr | 64.6 ± 15.2 | 62.2 ± 14.5 | <0.001 | 68.8 ± 12.2 | <0.001 |
| Male | 136,046 (53.4) | 904 (50.2) | 0.007 | 3649 (40.8) | <0.001 |
| Race/Ethnicity | | | <0.001 | | <0.001 |
| African-Amer. | 76,037 (29.8) | 683 (37.9) | | 2973 (33.2) | |
| Caucasian | 140,295 (55.0) | 715 (39.7) | | 4758 (53.2) | |
| Hispanic | 26,551 (10.4) | 287 (15.9) | | 900 (10.1) | |
| Other | 12,054 (4.7) | 117 (6.5) | | 315 (3.5) | |
| BMI category | | | <0.001 | | 0.036 |
| <20 kg/m$^2$ | 24,598 (9.7) | 242 (13.4) | | 870 (9.7) | |
| 20-24.9 kg/m$^2$ | 80,613 (31.6) | 650 (36.1) | | 2891 (32.3) | |
| 25-29.9 kg/m$^2$ | 72,923 (28.6) | 508 (28.2) | | 2612 (29.2) | |
| ≥30 kg/m$^2$ | 76,803 (30.1) | 402 (22.3) | | 2573 (28.8) | |
| Smoker | 14,335 (5.6) | 118 (6.6) | 0.090 | 440 (4.9) | 0.004 |
| Substance abuser | 5530 (2.2) | 90 (5.0) | <0.001 | 110 (1.23) | <0.001 |
| Unemployed | 242,245 (95.0) | 1743 (96.7) | <0.001 | 8761 (97.9) | <0.001 |
| Unable to ambulate | 10,986 (4.3) | 52 (2.9) | 0.003 | 405 (4.5) | 0.32 |
| Unable to transfer | 3921 (1.5) | 10 (0.6) | <0.001 | 144 (1.6) | 0.59 |
| In-center HD[4] | 237,545 (93.2) | 1723 (95.6) | <0.001 | 8316 (93.0) | 0.42 |
| Hb < 11.0 g/dL | 169,312 (72.8) | 1305 (78.7) | <0.001 | 6047 (74.6) | <0.001 |
| Comorbidities | | | | | |
| HTN | 214,792 (84.3) | 1573 (87.3) | <0.001 | 7794 (87.1) | <0.001 |
| DM | 134,561 (52.8) | 992 (55.1) | 0.055 | 5514 (61.6) | <0.001 |

| | | | | | |
|---|---|---|---|---|---|
| CHF | 84,244 (33.1) | 537 (29.8) | 0.004 | 3345 (38.5) | <0.001 |
| CAD | 71,447 (28.0) | 396 (22.0) | <0.001 | 2891 (32.3) | <0.001 |
| PVD | 38,852 (15.2) | 222 (12.3) | <0.001 | 1588 (17.8) | <0.001 |
| Prior CVA | 25,893 (10.2) | 235 (13.0) | <0.001 | 1419 (15.9) | <0.001 |
| Permanent AF | 36,345 (14.3) | 219 (12.2) | 0.011 | 2055 (23.0) | <0.001 |
| Comorbidity Score[5] | 5.0 ± 2.8 | 4.8 ± 2.6 | <0.001 | 5.6 ± 2.7 | <0.001 |
| Cause of ESRD | | | <0.001 | | <0.001 |
| DM | 119,334 (46.8) | 921 (51.1) | | 4965 (55.5) | |
| HTN | 67,290 (26.4) | 504 (28.0) | | 2323 (26.0) | |
| GN | 22,233 (8.7) | 132 (7.3) | | 452 (5.1) | |
| Other | 46,080 (18.1) | 245 (13.6) | | 1206 (13.5) | |

[1]Characteristics shown as *n* (%), except for continuous variables, which are shown as mean ± 1 standard deviation.

[2]Compares individuals with hemorrhagic stroke to those with no stroke.

[3]Compares individuals with ischemic stroke to those with no stroke.

[4]In-center HD is contrasted to self-care dialysis, which consists of home HD plus peritoneal dialysis.

[5]Comorbidity score is derived from an adapted form of the Liu Comorbidity Index.

Abbreviations: African-Amer., African-American; BMI, body mass index; HD, hemodialysis; Hb, hemoglobin; HTN, hypertension; DM, diabetes mellitus; CHF, congestive heart failure; CAD, coronary artery disease; PVD, peripheral vascular disease; CVA, cerebrovascular event; AF, atrial fibrillation; ESRD, end stage renal disease; GN, glomerulonephritis

**Supplemental Material**

## Appendix 1: Data sources and covariates

*Details on data sources and linking strategy*

We performed a retrospective cohort analysis of incident, Medicare and Medicaid ("dually eligible") chronic dialysis patients. Medicare is a federally-funded program for which nearly all adults with end stage renal disease are entitled, regardless of age; while not all individuals receiving chronic dialysis are Medicare enrollees, the vast majority is (1, 2).

Data for these analyses were assembled from two primary sources. First, we utilized the USRDS, a national system that collects data on virtually all patients undergoing chronic dialysis in the U.S. From the USRDS, we received standard patient records that included demographics, comorbidites, functional status, and dialysis modality (from the Medical Evidence Form, known as "CMS 2728") at the time of dialysis commencement. The USRDS also incorporates data on inpatient and outpatient medical claims paid by Medicare, which provides insurance coverage for the vast majority of dialysis patients. The Medicare claims files contain International Classification of Diseases – 9$^{th}$ Revision (ICD-9) codes for each date of service.

To make possible the study of dually-eligible individuals, the USRDS performed a deterministic match of these Medicaid beneficiaries against the core USRDS files to identify dually-eligible individuals on chronic dialysis. This permitted us to link USRDS data with Centers for Medicare & Medicaid Services (CMS) Medicaid prescription drug billing claims, in the form of the Medicaid Analytic eXtract Personal Summary Files and the final action prescription drug claims files. Medicaid files were used to determine prescription records for

methimazole or propothiouracil, which were elements used in our algorithms to identify nontransient, nonvalvular atrial fibrillation, as described below. These sources were linked using previously-described methodology to permit identification of dually-eligible dialysis patients in 2000-05.

*Details on covariates and descriptive variables*

Demographic and clinical variables, drawn from the CMS 2728 dialysis intake form, included age, sex, race by ethnicity, body mass index, employment status, smoking, substance abuse (alcohol or illicit drugs), ability to ambulate and to transfer, cause of ESRD, and dialysis modality. Comorbidities consisted of diabetes, congestive heart failure, coronary artery disease, cerebrovascular disease, and peripheral vascular disease. Ethnicity was categorized into one of four mutually-exclusive groups: non-Hispanic Caucasians, non-Hispanic African-Americans, Hispanics, and Others. Body mass index (BMI) was classified into 4 categories: $< 20$ kg/ $m^2$, 20-24.99 kg/$m^2$, 25-29.99 kg/$m^2$, $\geq 30$ kg/$m^2$. Cause of ESRD was categorized as diabetes, hypertension, glomerulonephritis, or other. Because the CMS 2728 form is structured such that diabetes and hypertension may be considered as both a cause of ESRD and/or a "freestanding" comorbidity, for the purposes of the present analysis, these two covariates were considered a comorbidity if they were listed as either the cause of ESRD or as a "freestanding" comorbidity on the CMS 2728 form (3, 4). Dialysis modality at time of dialysis initiation was categorized as in-center hemodialysis or self-care dialysis (home hemodialysis or peritoneal dialysis). We used a modified form (5) of the Liu Comorbidity Index (6). This index is a summary measure of disease burden which also includes cause of ESRD; therefore, cause of ESRD was not modeled separately. However, our form of this index used only 90 (rather than 180) days in which to

acquire claims since we have previously found little difference in indices generated using 90 or 180 days of claims data (5) and because we had required that patients have Medicaid and Medicare coverage throughout the first 90 days.

*Details on the determination of permanent atrial fibrillation*

The ICD-9 code 427.31 was used to identify AF claims using a well-established algorithm designed to determine the presence of nontransient, nonvalvular AF (4, 7). Individuals who had hyperthyroidism or thyrotoxicosis were eliminated, based on the presence of relevant ICD-9 and/or CPT (Common Procedural Technology) and/or HCPCS (Healthcare Common Procedure Coding System) codes, or by a prescription at any time for methimazole or propothiouracil. We next eliminated patients with evidence of valvular heart disease (using ICD-9 codes). Finally, to minimize potential misclassification from perioperative sources of AF (e.g., coronary artery bypass surgery), claims (rather than individuals) were eliminated unless there was a preexisting (> 30 d) AF claim. This resulted in the elimination of individuals in whom AF claims were only proximally related to cardiac surgery, but allowed inclusion of individuals in whom there was evidence of preexisting AF. To classify individuals as having permanent AF, we initially required a total of 2 (or more) AF claims, separated by ≥30 days, of which no more than 1 was an inpatient claim. Additionally, we expunged all outpatient AF claims within 7 days of a subsequent AF claim-containing admission and within 30 days after an AF claim-containing admission, retaining only the original inpatient claim.

*Details on the determination of stroke events*

A specific approach, in which only the codes with higher specificities for ischemic or hemorrhagic strokes were used, was the primary approach. For purposes of sensitivity analyses, we also created more inclusive approaches that allowed for a broader range of ischemic and hemorrhagic stroke codes. For identifying ischemic strokes, we used the strategy of Go et al (8, 9). The patient was considered to have an ischemic stroke if the principal diagnosis ICD-9 code at the time of hospital discharge was 434 or 436 and one of the following occurred: (a) the patient expired during the hospitalization; (b) the hospitalization lasted ≥ 48 hours; or (c) the hospitalization lasted < 48 hours and the patient did not have a carotid endarterectomy (ICD-9 code 381.2). In the absence of 434 or 436, ICD-9 code 362.3 was sufficient to diagnose an ischemic stroke. The sensitive approach differed only by treating code 433 analogously to 434 and 436.

For hemorrhagic strokes, a specific approach, using codes with higher specificity for hemorrhagic stroke, comprised the primary analysis. As above, codes were hospital discharge codes residing in the first position. A sensitivity analysis, in which codes with greater sensitivity for hemorrhagic stroke were used, was also performed. The primary (specific) approach utilized codes 430 and 431, while the more sensitive approach added codes 432, 852.0, 852.2, 852.4, and 853.

## Appendix 2: Detailed Statistical Methods of the Survival Model

*Assessment of dialysis time on survival following stroke*

We initiated our analysis of time to death by any cause by plotting separate Kaplan Meier (KM) curves among the subset of individuals who experienced a hemorrhagic stroke, and again for the subset who experienced an ischemic stroke. The start time ("time zero") for these analyses were the date the stroke occurred, thus survival time was time from stroke until death from any cause. For each stroke type subgroup (hemorrhagic or ischemic), subjects were stratified by their time on dialysis prior to the stroke occurrence. The log-rank test for the hemorrhagic stroke subgroup did not find evidence of a significant effect of time on dialysis on survival from the stroke until death from any cause ($p = 0.53$). Similarly for the ischemic stroke subgroup, time on dialysis did not affect survival following stroke ($p = 0.40$). Visual inspection of the corresponding KM curves of survival time from stroke occurrence through death from any cause was consistent with these log-rank test results, indicating that the survival experience following either type of stroke event was not significantly altered by how long a subject had been on dialysis before the stroke occurred.

*Survival function when strokes do not occur*

For subjects that did not experience a stroke during the observation window or, for those that did, for their pre-stroke time (i.e., from time of cohort entry [dialysis plus 90 days] through death or the end of their follow-up time), initial modeling of survival times was obtained using a Cox proportional hazards (PH) model. For this analysis, the start time ("time zero") was cohort entry (date of dialysis initiation plus 90 days) and the outcome of interest was death from any

cause. Our Cox model of this outcome allowed us to identify risk factors associated with mortality. These risk factors included baseline demographic and clinical measures, as well as atrial fibrillation, which was developed as a time-dependent covariate following the algorithm described in Appendix 1. The baseline clinical measures included the occurrence of a stroke prior to cohort entry, but the impact of a stroke event on all-cause mortality that occurred during follow-up was addressed in a different manner, as described below. The validity of the PH assumptions for this model were ascertained using log-log survival plots for the categorical predictors and plots of Schoenfeld residuals versus time for the continuous variables.

Next, to be able to estimate years of life lost due to stroke after adjusting for the effect of all risk factors, we needed to identify a fully parametric form of the survival model. We first did this by fitting a baseline hazard function to our semi-parametric Cox PH model, and then incorporating the impact of a hemorrhagic or ischemic stroke event using an additive hazard extension, which is described in detail below. We generated the KM survival curve of individuals who did not have a stroke (or, for those who had a stroke, their pre-stroke time) to use for validation of the fully parametric models we selected (i.e., for the observed vs. expected plot comparisons). Of the various candidate parametric distributions (e.g., exponential, Weibull, lognormal, loglogistic and gamma) that were considered for assessing the model fit of this baseline survival function, a Weibull distribution with a shape parameter of 0.91 and scale parameter of 80.21 was found to be the best candidate based on Akaike's information criterion (AIC = 133727.8). Exploiting the fact that a Weibull distribution can also be expressed by means of a PH model, this baseline survival function was then adjusted to account for the effect of other risk factors − for categorical variables, this represented the proportion of subjects with the various characteristics − using the mean value for each explanatory variable to provide a

population-wide risk adjusted survival curve (i.e., the expected curve) for comparison to the KM curve (i.e., the observed curve) mentioned above to facilitate model assessment by visual inspection of the observed vs. expected plots.

*Modeling changes to the survival function as a result of stroke*

Since ischemic and hemorrhagic strokes are two very different life-changing events, it is reasonable to assume that the survival function will change very differently − even in a manner that is no longer proportional in terms of hazards − upon incidence of one of the two types of strokes. As the overall aim was to be able to quantify residual longevity between those who experienced a stroke compared to those who did not, it was required to model changes in hazard of death upon the occurrence of each type of stroke to facilitate this comparison. To this end, we used a semi-Markov model with additive hazard extension, which enabled us to utilize the fully parametric model of survival from cohort entry to death described above, and also to incorporate the dramatic changes in the survival functions due to the occurrence of a stroke while on dialysis using additive hazard extensions.

As a brief explanation, a semi-Markov model assumes that the mortality experience of patients at any given time *t* after their entry into cohort (dialysis initiation plus 90 days) is affected by a single transitive stroke event in two ways: by the occurrence of stroke, and, by the time since the stroke occurred. As suggested by the KM curves (above), the amount of time on dialysis before stroke occurs does not influence this mortality experience following the stroke, so that layer of complexity beyond the semi-Markov assumption is not required. The additive hazard extension allows us the flexibility to model incremental hazard of death for those experiencing stroke in addition to the hazard influences of risk factors other than strokes.

Notably, both semi-Markov models (10) and additive hazards models have been used in clinical and biostatistical research (11). More information about semi-Markov models with an application on bone marrow transplant study is available (12).

To model survival time from a stroke event (designated as "time zero") to death from any cause, we selected from candidate parametric models. This was done to facilitate our aim to estimate years of life lost due to stroke after adjusting for the effect of all mortality risk factors: a subject's survival would follow that of the Weibull model initially, and in the event the subject then experienced a stroke, his or her remaining survival time (until death from any cause) would subsequently follow the survival function identified here. Of the various candidate parametric distributions that we considered (such as exponential, Weibull, lognormal, gamma, loglogistic), the generalized gamma distribution for individuals who experienced a hemorrhagic stroke (shape parameter = -2.281, scale parameter = 1.258, AIC = 1907.07) and the lognormal distribution for individuals who experienced an ischemic stroke (scale parameter = 1.883; AIC = 7411.36) were found to provide the best fit based on their AIC values. Additionally, we also performed visual inspection of the observed (KM) vs. expected plots for time since stroke ("time zero" for these models) until death from any cause to further validate our selected models. It should be noted here that the choice of generalized gamma distribution is a very flexible option that includes both Weibull (shape parameter = 1) and lognormal distributions (shape parameter = $\infty$) as special cases, and can be used to model a variety of hazard shapes (both increasing and decreasing) that cannot be modeled by some popular parametric distributions (13). Additionally, a complete taxonomy of hazard functions related to the generalized gamma distribution has been described (14).

*Estimates of median residual lifetime*

Using this flexible modeling approach we were able to obtain, at any given survival time on dialysis *t*, say, and time spent on stroke $t_\Delta$, say, (where $t_\Delta = t - t_s$ , so $t_s$ represents the time on dialysis that the stroke occurred), the following estimates were calculated:

1. median residual lifetime on dialysis absent a stroke, calculated as the number of months for the survival function for no stroke to drop in half from a given point of reference (say, $t_s$), so the point on the time axis where the height of the survival curve drops to $0.5 \cdot S(t_s)$;
2. median residual lifetime on dialysis following a stroke event, calculated as above except with the survival function now following a different path (due to the additive hazard) to find the point where the height of the survival curve drops to $0.5 \cdot S(t_s)$, noting that the curves for hemorrhagic stroke are different than those for ischemic; and
3. median residual life lost due to stroke (hemorrhagic and ischemic) was found by subtracting (2) from (1) above.

*Resulting survival model and further diagnostics*

Overall, this framework allowed us to study risk accentuation in this dialysis cohort, and in particular the heavy influence of ischemic and hemorrhagic stroke events on estimates of remaining survival time. This also facilitated the creation of survival curves to demonstrate these influences of stroke. These curves included adjustment for effects of other modeled risk factors.

As a diagnostic check for our modeling approach, we compared the model-predicted survival profile (adjusted for all risk factors as described above) for patients who experienced a

stroke near the point of cohort entry (as stated above, defined as dialysis initiation plus 90 days) with the unadjusted KM curve for subjects that experienced strokes using survival time after stroke as the time axis for the KM curve (i.e., stroke occurrence as "time zero" for this diagnostic check) and found agreement (that is, no lack-of-fit). We were also able to estimate the relative hazard for those who had strokes compared to those that did not as a function of time after stroke. This varied based on how long the subject had been on dialysis before the stroke occurred due to the fact that the Weibull hazard of non- or pre-stroke (reference group) subjects was found to be a decreasing function of time. Thus, this non-constant hazard ratio is influenced by stroke "vintage", though maintaining the same overall trend over time. Finally, using our approach we were able to quantify a sample-based average estimate of median residual months of life lost due to stroke while simultaneously allowing standard interpretation of hazard ratios for the other risk factors.

*Sensitivity analyses*

We performed multiple sensitivity analyses. First, we eliminated the adapted form (5) of the Liu Comorbidity Index (15) because the index contains some disease conditions which overlap with conditions listed on the CMS 2728 dialysis intake form. We next used more sensitive claims-based definitions of both ischemic and hemorrhagic stroke to perform the analyses. We then eliminated individuals with a history of a previous stroke (either on the CMS 2728 form, which captures predialysis events, on in claims during the first 90 days of dialysis), tand performed a subset analysis on patients who were on HD only (but who also had no additional strokes). Finally, we conducted four additional analyses in the Medicare cohort, first

eliminating those with previous strokes and then confining the analysis to HD patients (who also did not have previous strokes).

## References


1. United States Renal Data System. USRDS 2012 Annual Data Report: Atlas of End-Stage Renal Disease in the United States. In: National Institutes of Health, National Institute of Diabetes and Digestive and Kidney Diseases. 2012
2. Wetmore JB, Rigler SK, Mahnken JD, Mukhopadhyay P, Shireman TI: Considering health insurance: how do dialysis initiates with Medicaid coverage differ from persons without Medicaid coverage? *Nephrol Dial Transplant* 25**:** 198-205, 2010
3. Volkova N, McClellan W, Soucie JM, Schoolwerth A: Racial disparities in the prevalence of cardiovascular disease among incident end-stage renal disease patients. *Nephrol Dial Transplant* 21**:** 2202-2209, 2006
4. Wetmore JB, Mahnken JD, Rigler SK, Ellerbeck EF, Mukhopadhyay P, Spertus JA, Hou Q, Shireman TI: The prevalence of and factors associated with chronic atrial fibrillation in Medicare/Medicaid-eligible dialysis patients. *Kidney Int* 81**:** 469-476, 2012
5. Rigler SK, Wetmore JB, Mahnken JD, Dong L, Ellerbeck EF, Shireman TI: Impact of a modified data capture period on Liu comorbidity index scores in Medicare enrollees initiating chronic dialysis. *BMC nephrology* 14**:** 51, 2013
6. Liu J, Huang Z, Gilbertson DT, Foley RN, Collins AJ: An improved comorbidity index for outcome analyses among dialysis patients. *Kidney Int* 77**:** 141-151, 2010
7. Go AS, Hylek EM, Phillips KA, Chang Y, Henault LE, Selby JV, Singer DE: Prevalence of diagnosed atrial fibrillation in adults: national implications for rhythm management and stroke prevention: the AnTicoagulation and Risk Factors in Atrial Fibrillation (ATRIA) Study. *Jama* 285**:** 2370-2375, 2001
8. Go AS, Hylek EM, Chang Y, Phillips KA, Henault LE, Capra AM, Jensvold NG, Selby JV, Singer DE: Anticoagulation therapy for stroke prevention in atrial fibrillation: how well do randomized trials translate into clinical practice? *Jama* 290**:** 2685-2692, 2003
9. Wetmore JB, Ellerbeck EF, Mahnken JD, Phadnis MA, Rigler SK, Spertus JA, Zhou X, Mukhopadhyay P, Shireman TI: Stroke and the "Stroke Belt" in Dialysis: Contribution of Patient Characteristics to Ischemic Stroke Rate and Its Geographic Variation. *J Am Soc Nephrol*, 2013
10. Putter H, Fiocco M, Geskus RB: Tutorial in biostatistics: competing risks and multi-state models. *Statistics in medicine* 26**:** 2389-2430, 2007

11. Xie X, Strickler HD, Xue X: Additive hazard regression models: an application to the natural history of human papillomavirus. *Computational and mathematical methods in medicine* 2013**:** 796270, 2013
12. Dabrowska M, Sun G, Horowitz MM: Cox regression in a Markov renewal model: An application to the analysis of bone marrow transplant data. *J Am Statistical Asso* 89**:** 867-877, 1994
13. Holtz TH, Sternberg M, Kammerer S, Laserson KF, Riekstina V, Zarovska E, Skripconoka V, Wells CD, Leimane V: Time to sputum culture conversion in multidrug-resistant tuberculosis: predictors and relationship to treatment outcome. *Annals of internal medicine* 144**:** 650-659, 2006
14. Cox C, Chu H, Schneider MF, Munoz A: Parametric survival analysis and taxonomy of hazard functions for the generalized gamma distribution. *Statistics in medicine* 26**:** 4352-4374, 2007
15. Liu J, Huang Z, Gilbertson DT, Foley RN, Collins AJ: An improved comorbidity index for outcome analyses among dialysis patients. *Kidney Int* 77**:** 141-151,