Blood pressure and the risk of Chronic Kidney Disease progression using multistate marginal structural models in the CRIC Study

Alisa J. Stephens-Shields[*,†], **Andrew J. Spieker**[†], **Amanda Anderson**[†], **Paul Drawz**[§], **Michael Fischer**[¶], **Stephen M. Sozio**[‖], **Harold Feldman**[†], **Marshall Joffe**[†], **Wei Yang**[†,††], **Tom Greene**[**,††], and **the CRIC Study Investigators**[‡]

[†]Department of Biostatistics and Epidemiology, Perelman School of Medicine, University of Pennsylvania

[§]Department of Medicine, University of Minnesota

[¶]Department of Medicine, University of Illinois College of Medicine

[‖]Department of Medicine, Johns Hopkins University School of Medicine

[**]Department of Population Health Sciences, University of Utah School of Medicine

Abstract

In patients with chronic kidney disease (CKD), clinical interest often centers on determining treatments and exposures that are causally related to renal progression. Analyses of longitudinal clinical data in this population are often complicated by clinical competing events, such as end-stage renal disease (ESRD) and death, and time-dependent confounding, where patient factors that are predictive of later exposures and outcomes are affected by past exposures. We developed multistate marginal structural models (MS-MSM) to assess the effect of time-varying systolic blood pressure on disease progression in subjects with CKD. The multistate nature of the model allows us to jointly model disease progression characterized by changes in the estimated glomerular filtration rate (eGFR), the onset of ESRD, and death, and thereby avoid unnatural assumptions of death and ESRD as non-informative censoring events for subsequent changes in eGFR. We model the causal effect of systolic blood pressure on the probability of transitioning into one of six disease states given the current state. We use inverse probability weights with stabilization to account for potential time-varying confounders, including past eGFR, total protein, serum creatinine, and hemoglobin. We apply the model to data from the Chronic Renal Insufficiency Cohort (CRIC) Study, a multisite observational study of patients with CKD.

[*]Correspondence to: Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadephia, PA, 19104.
[‡]The CRIC Study Investigators: Lawrence J. Appel, MD, MPH, Alan S. Go, MD, Jiang He, MD, PhD, John W. Kusek, PhD, James P. Lash, MD, Akinlolu Ojo, MD, PhD, Mahboob Rahman, MD, Raymond R. Townsend, MD
[††]Wei Yang and Tom Greene contributed equally to this manuscript
John Wiley & Sons, Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

1. Introduction

In studies of chronic kidney disease (CKD), endpoints of various types are often of interest. Time-to-event endpoints such as time to end-stage renal disease (ESRD) are considered as well as longitudinal outcomes such as change in the estimated glomerular filtration rate (eGFR). Studies with a primary interest in longitudinal outcomes are typically challenged by the occurrence of death or ESRD, which censor follow-up during the observation period. When these clinical events, which preclude subsequent measurement of the longitudinal outcome, are interpreted as censoring events, inferences concerning longitudinal changes may be biased due to informative censoring. More fundamentally, estimands dependent on measurements of the longitudinal outcome subsequent to these events can be difficult to interpret, as described in the truncation by death literature [1, 2]. Alternatively, inferences restricted to the time-to-clinical-event outcomes provide limited information concerning effects of treatments or exposures for patients with more modest declines in kidney function. Recent reviews have emphasized likelihood-based shared random effects approaches for joint modeling of longitudinal processes and time-to-event in CKD studies[3, 4]. These methods are challenged by difficulty in implementation, and are subject to the same limitations as standard regression models in that when estimating joint exposure effects across time, neither adjustment nor non-adjustment of time-dependent confounders, which we next describe, result in an unbiased estimate of the causal joint exposure effect.

In the analysis of ESRD, often the primary outcome of interest in CKD studies, changes in longitudinal outcomes, such as eGFR, prior to the absorbing event, e.g., ESRD, may serve as confounding factors for the effects of exposures. When exposures are time-varying, interim longitudinal outcomes can cause time-dependent confounding, where time-dependent confounders are variables that are affected by prior exposures and predictive of later exposures and outcomes. In CKD, several variables could potentially serve as time-dependent confounders given the complex, and often bi-directional, interplay between blood pressure and kidney function, other clinical factors, and treatment decisions. Further, death is a competing risk for ESRD.

To estimate the effect of time-varying exposures on CKD progression defined by both time-to-event and longitudinal outcomes such as ESRD and changes in eGFR, respectively, with proper adjustment of time-dependent confounding, we propose multistate marginal structural models (MS-MSMs). Marginal Structural Models (MSMs) are a class of causal models[5] that may be used to consistently estimate effects of time-varying exposures in the presence of time-dependent confounding. MS-MSMs are particularly suited to this setting for several reasons. First, hypertension is both a risk factor and effect of CKD. As a result, several factors that are associated with CKD are likely to function as time-dependent confounders for the effect of changes in blood pressure on CKD outcomes. As an example, hypertension may lead to decreased eGFR, which in turn can elevate subsequent blood pressure levels and also the risk of ESRD or death. Unbiased estimation of the causal effect of time-varying blood pressure on CKD progression therefore requires statistical methods that can accommodate time-dependent confounders. Standard methods of longitudinal and survival analysis such as Cox proportional hazards regression with time-dependent covariates and

mixed models or semi-parametric regression methods such as generalized estimating equations (GEE) cannot handle time-dependent confounding appropriately [5]. Alternatively, MSMs with inverse probability weighting [6], may be used to consistently estimate causal effects of time-varying exposures on longitudinal and time-to-event outcomes.

Our main objective is to introduce MS-MSMs with estimation by inverse probability weighting to estimate the effect of time-dependent exposures on joint longitudinal and time-to-event outcomes by a Markov multinomial model approach. In formulating MS-MSMs, we define states that are jointly characterized by longitudinal and time-to-event endpoints and model the effect of exposures on transitions among these states. [7] recently used inverse probability weighting to estimate the causal effect of alternative employee benefit programs on transitions among employment states. This analysis considered a time-independent exposure variable and Cox and Aalen models for estimating transition intensities. Our approach is generalized to time-dependent exposures, and we estimate transition intensities using a generalized linear model approach. Our model is structurally similar to the MSM for cause-specific hazards in the analysis of time-to-event data of multiple outcome types developed by [8], but differs substantially in the interpretation since our model accommodates both the absorbing states encountered in the survival literature and the non-absorbing states observed in repeated-measures settings. We can use our model to estimate the probability of patients cycling among multiple states defined by longitudinal outcomes during the progression and regression of a chronic condition while simultaneously estimating the probability of a patient experiencing an absorbing event. As a result, our model is better suited for the analysis of CKD since clinical interest lies in characterizing a patient's trajectory over time [9] in addition to evaluating time to clinical, terminal events. We apply our methodology to a study of the effect of systolic blood pressure (SBP) on CKD progression within the Chronic Renal Insufficiency Cohort (CRIC) Study, a multisite, prospective cohort study of nearly 4,000 subjects with mild to moderate CKD at study entry.

In Section 2, we define notation and formalize identification assumptions for the causal effects of time-varying exposures on multistate outcomes, where MSMs model the causal effect of time-varying exposures on the probability of transitions among states. We then utilize transition probabilities to determine the marginal probability of being in a potential state at the end of observation under a specified joint exposure. In Section 3 we conduct a simulation study to illustrate the properties of inverse probability weighted estimators for multistate outcomes, comparing weights with and without truncation, and in Section 4 we use our MS-MSMs to determine the effect of SBP on CKD progression among mild to severe CKD as defined the level of eGFR and clinical endpoint states including ESRD and death.

2. Methods

2.1. Notation

We first define our notation and then describe our causal model and identifying assumptions. Our notation follows that introduced by [5] for time-varying exposures and scalar outcomes. First considering the observed data, let $Y_{i,j+1}$ denote an outcome taking levels 1,2,…,K,

indicating the ith subject's state at time $j+1$ for j=0,1, 2, ..., J. Similarly, let $A_{i,j}$ denote subject $i's$ exposure level at time j. Covariates measured at time j are represented by $\mathbf{L}_{i,j}$ and may include baseline as well as time-varying covariates. Finally, $C_{i,j}$ indicates whether a subject is censored by the beginning of the jth observation time with $C_{i,j}=1$ denoting that a subject is censored and $C_{i,j}=0$ representing that a subject is still under observation. Within each time interval we assume that variables are observed in the temporal order $(\mathbf{L}_{i,j}, A_{i,j}, C_{i,j+1}, Y_{i,j+1})$. Overbars are used to denote variable history, such that $\overline{Y}_{i,j+1}=(Y_{i,1}, Y_{i,2}, \ldots, Y_{i,j+1})$. Causal effects of exposures are defined using potential outcomes notation [10], where $Y_{i,j+1}^{\overline{a}}$ is the outcome that would have been observed for subject i at time j+1 had he or she received the joint exposure $= (a_0, a_1, \ldots, a_j)$. At the population level, interest often lies in comparing mean potential outcomes under different

joint exposures , where a nonzero average causal effect is indicated by $E[Y_{i,j}^{\overline{a}}] \neq E[Y_{i,j}^{\overline{a}'}]$ for

joint exposures $'$. Our multistate approach targets comparisons of $P(Y_{i,J+1}^{\overline{a}}=k)$, the marginal probability of a potential outcome state at the end of observation under alternative joint exposures $'$.

2.2. Model

Multistate models facilitate estimation of the marginal probabilities of outcome states by first modeling the probability of transitioning to an outcome state at a given time $j+1$ conditional on previous outcome states. We take this same approach using potential outcomes to estimate causal effects on multistate outcomes. First we factor the joint

likelihood of the full history of potential outcome states $f\left(\overline{Y}_{i,J+1}^{\overline{a}}\right) = \left(Y_{i,1}^{\overline{a}}, Y_{i,2}^{\overline{a}}, \ldots, Y_{i,J+1}^{\overline{a}}\right)$ from time 1 to $J+1$ as a product of conditional distributions

$f\left(\overline{Y}_{i,J+1}^{\overline{a}}\right) = \prod_{j=0}^{J} f\left(Y_{i,J+1}^{\overline{a}}/Y_{i,J}^{\overline{a}}\right)$. Under a first order Markov assumption we further simplify

the product of conditional distributions as $\prod_{j=0}^{J} f\left(Y_{i,J+1}^{\overline{a}}|\overline{Y}_{i,J}^{\overline{a}}\right) = \prod_{j=0}^{J} f\left(Y_{i,J+1}^{\overline{a}}|Y_{i,J}^{\overline{a}}\right)$, indicating that the probability of a given outcome state at time j+1 depends solely on the most recent potential outcome state rather than all states from time 1 to time j. The first order Markov assumption is chosen for convenience but may be relaxed if data provide evidence in favor of a higher order Markov assumption. The implications of higher order assumptions for modeling state transitions are discussed in Section 5. For the multistate potential outcomes considered, multinomial models may be used to characterize the probability of outcome state k at time $j+1$ conditional on the state at time j. Our multistate marginal structural model is

$$log\left(\frac{P\left(Y_{i,j+1}^{\overline{a}}=k|Y_{i,j}^{\overline{a}}=k'\right)}{P\left(Y_{i,j+1}^{\overline{a}}=k'|Y_{i,j}^{\overline{a}}=k'\right)}\right) = h_1\left(\overline{a}_{i,j}; \beta^{k',k}\right) + h_2\left(j; \alpha^{k',k}\right) \tag{1}$$

for known functions $h_1(\cdot)$ and $h_2(\cdot)$ up to the unknown parameters $\beta^{k',k}$ and $\alpha^{k',k}$. The probability of transitioning from state k' to k from time j to $j+1$ under joint exposure $_{i,j}$ is

modeled by the log odds of $Y^{\bar{a}}_{i,j+1}=k$ versus the reference outcome state $Y^{\bar{a}}_{i,j+1}=k'$. The parameter $\beta^{k',k}$ measures the causal effect of exposure on the probability of transition to one potential outcome state given the previous potential outcome state, and function $h_2(j;a^{k',k})$ defines the probability of transition for the reference exposure group ($= (0,\ldots, 0)$), accounting for underlying temporal trends in the relative likelihood of transitioning to state k versus reference state k'. The assumption of additivity of $h_1(\,_{i,j};\beta^{k',k})$ and $h_2(j;a^{k',k})$ implies that the effect, of exposure $a_{i,j}$ on transition probabilities does not change

over time. As an example, one might consider the model $h_1\left(\bar{a}_{i,j};\beta^{k,k'}\right)=\beta^{k',k}a_{i,j}$ for a binary treatment $A_{i,j}$, where $\beta^{k',k}$ would then denote the log odds ratio of transitioning to state k versus state k' from state k' under binary treatment $a_{i,j}=1$ at time j compared to

under control $a_{i,j}=0$ at time j. Another model might consider $h_1\left(\bar{a}_{i,j};\beta^{k,k'}\right)=\beta^{k',k}\sum_{t=0}^{j}a_{i,t}$ in which the log odds ratio of transitioning to state k versus k' from state k' depends on the

cumulative exposure history, where for binary exposures $\sum_{t=0}^{j}a_{i,t}$ is the number of times a

subject was exposed through time j. The reference function $h_2\left(j;\alpha^{k',k}\right)$ may include polynomial terms or splines to flexibly capture temporal changes in the probability of transitions among states.

The predicted outcomes of model (1) define the elements of the $K\times K$ transition matrix

$$
\lambda^{\bar{a}}_j=\begin{bmatrix}
\lambda^{\bar{a},1,1}_j & \lambda^{\bar{a},1,2}_j & \cdots & \lambda^{\bar{a},1,K}_j \\
\lambda^{\bar{a},2,1}_j & \lambda^{\bar{a},2,2}_j & \cdots & \lambda^{\bar{a},2,K}_j \\
\vdots & \ddots & \ddots & \vdots \\
\lambda^{\bar{a},K,1}_j & \lambda^{\bar{a},K,2}_j & \cdots & \lambda^{\bar{a},K,K}_j
\end{bmatrix}, \tag{2}
$$

where $\lambda^{\bar{a},k',k}_j=P\left(Y^{\bar{a}}_{i,j+1}=k|Y^{\bar{a}}_{i,j}=k'\right)$ for transitions among possible states under joint exposure .

Robins (1998) states that the causal parameter $\beta^{k',k}$ is identified under the following assumptions[6]:

1. Sequential Ignorability/Exchangeability. $Y^{\bar{a}}_{i,j+1} \perp A_{i,j}|\overline{Y}_{i,j}, \overline{A}_{i,j-1}, \overline{L}_{i,j}$

   Given the previous state and exposure and covariate history, potential outcomes under treatment at time $j+1$ are independent of the observed exposure $A_{i,j}$.

2. Consistency. For $\bar{A}_{i,j}=\bar{a}_{i,j}$, $Y_{i,j}=Y^{\bar{a}}_{i,j}$ For a subject who is actually observed to have joint exposure $_{i,j}$ the observed outcome $Y_{i,j}$ equals the potential outcome under that exposure history.

3.

Positivity. For every $\overline{\mathbf{L}}_{i,j}$ such that $f_{\overline{\mathbf{L}}_{i,j}}\left(\overline{l}\right) > 0$, $f_{A_{i,j}|\overline{\mathbf{L}}_{i,j}}\left(a_{i,j}|\overline{\mathbf{l}}\right) > 0$ for all possible treatment levels $a_{i,j}$. This states that for any observed covariate history, where past outcomes $Y_{i,j}$ are included in, $\overline{\mathbf{L}}_{i,j}$, there is a nonzero probability of any possible exposure level at time j.

2.3. Estimation of Transition Probabilities $P(Y_{i,j+1}^{\overline{a}} = k | Y_{i,j}^{\overline{a}} = k')$

Robins (2000) describes that the causal parameter of MSMs $\beta^{k,k'}$ may be estimated semiparametrically by weighted estimating equations, where the observed data are weighted using subject-specific, time-specific inverse probability of treatment weights (IPW) [5]. IPW accounts for time-varying confounding by creating a pseudo-population in which covariates at each time of exposure are not associated with the subsequent exposure, but the transition probabilities $P\left(Y_{i,j+1}^{\overline{a}} = k | Y_{i,j}^{\overline{a}} = k'\right)$ remain the same as in the source population. The observed data association model is

$$log\left(\frac{P\left(Y_{i,j+1} = k | Y_{i,j} = k', \overline{A}_{i,j}\right)}{P\left(Y_{i,j+1} = k' | Y_{i,j} = k', \overline{A}_{i,j}\right)}\right) = h_1\left(\overline{a}_{i,j}; \theta^{k',k}\right) + h_2\left(j; \eta^{k',k}\right), \tag{3}$$

and weighted estimating equations are

$$\sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{k=1}^{k-1}\hat{\omega}_{i,j}\left(Y_{i,j+1,k} - \frac{\exp\left\{h_1(\overline{a}_{i,j}; \theta^{k',k}) + h_2(j; \eta^{k',k})\right\}}{1 + \sum_{k*\neq K}\exp\left\{h_1\left(\overline{a}_{i,j}; \theta^{k',k*}\right) + h_2\left(j; \eta^{k',k*}\right)\right\}}\right) = 0, \tag{4}$$

where $Y_{i,j+1,k} = I(Y_{i,j+1} = k)$, and estimated stabilized weights

$\hat{\omega}_{i,j} = \dfrac{\Pi_{t=0}^{j}\hat{P}\left(A_{i,t} = a | \overline{A}_{i,t-1}\right)}{\Pi_{t=0}^{j}\hat{P}\left(A_{i,t} = a | \overline{A}_{i,t-1}, \overline{L}_{i,t}\right)}$ account for time-dependent confounding. Stabilized weights are estimated by fitting two models. The first one models the likelihood of the observed exposure $A_{i,j}$ given past exposure $_{i,j-1}$ and time-dependent covariates $\overline{\mathbf{L}}_{i,j}$. It informs the denominator of the stabilized weight and is required to be correct to consistently estimate the causal parameter. The second model for the numerator, which is not required to be correctly specified for consistent estimation, is based on a model for exposure $A_{i,j}$ given past exposure $_{i,j-1}$. In the pseudo-population the exposure-outcome relationship is no longer confounded by time-dependent exposure so that the associational parameter $\theta^{k,k'}$ is equivalent to the causal parameter $\beta^{k,k'}$ [5].

For binary exposures, the IPW may be estimated by pooled logistic regression. The probabilities of categorical exposures of greater than two levels may be estimated by pooled multinomial logistic regression. We take this approach in modeling SBP as an exposure, defining clinically relevant levels for normal, moderately elevated, and elevated SBP as used

by prior investigators. For continuous exposures, inverse density weighting is possible by specifying a density function for the exposure distribution and estimating components of that density by a standard modeling approach. An alternative approach for continuous exposures is to use a pooled multinomial model considering quantiles of the continuous exposure [11]. For each subject, predicted probabilities of observed exposure level at each time are calculated, and the product is taken within subject over time to estimate $\hat{\omega}_{i,j}$. In categorizing continuous exposures, investigators should be conscious of a tradeoff between the positivity and consistency assumptions. For continuous exposures, it is less likely that every level of the exposure occurs for each level of observed treatment and confounders at each time, thus raising concerns of potential positivity violations. In reducing treatment levels through categorization, it is more plausible that each treatment level is possible at each level of observed treatment and confounders, but violations of the consistency assumption are possible as different continuous exposure levels that result in the same categorical exposure variable could potentially have different implications for potential outcomes, which in turn would violate consistency as a lack of multiple versions of treatment[12], and is particularly relevant for exposures that are biological characteristics[13]. We therefore caution investigators with these considerations when categorizing continuous exposures.

When subjects leave the study prior to the end of the observation period $\theta^{k',k}$ are subject to selection bias when exposures and time-dependent factors are predictive of study withdrawal [5]. In this case, estimation equations are,

$$\sum_{i=1}^{n}\sum_{j=1}^{J}\sum_{k=1}^{k-1} I\left(C_{i,j+1}=0\right)\omega*i,j \left(Y_{i,j+1,k} - \frac{\exp\left\{h_1(\overline{a}_{i,j};\theta^{k',k})+h_2(j;\eta^{k',k})\right\}}{1+\sum_{k*\neq K}\exp\left(h_1\left(\overline{a}_{i,j};\theta^{k',k*}\right)+h_2\left(j;\eta^{k',k*}\right)\right)} \right)=\mathbf{0},$$

(5)

and weights

$$\hat{\omega}*i,j=\frac{\Pi_{t=0}^{j}\hat{P}\left(A_{i,t}=a|\overline{A}_{i,t-1},C_{i,t}=0\right)}{\Pi_{t=0}^{j}\hat{P}\left(A_{i,t}=a|\overline{A}_{i,t-1},C_{i,t}=0,\overline{L}_{i,t}\right)}\times\frac{\Pi_{t=0}^{j}\hat{P}\left(C_{i,t}=0|\overline{A}_{i,t-1},C_{i,t-1}=0\right)}{\Pi_{t=0}^{j}\hat{P}\left(C_{i,t}=0|\overline{A}_{i,t-1},C_{i,t-1}=0,\overline{L}_{i,t-1}\right)}$$ are
the product of the stabilized inverse probability of the observed joint exposure and stabilized inverse probability of being uncensored. Standard errors and confidence intervals of

parameter estimates $\hat{\beta}^{k',k}$ may be calculated using the robust variance estimator as for Generalized Estimating Equations [14], which is available in standard statistical software packages and provides asymptotically conservative standard error estimates provided that the weights are consistently estimated [15], or by a resampling technique such as the nonparametric bootstrap [16].

2.4. Estimating Marginal Probabilities $P\left(Y^{\overline{a}}_{i,J+1}=k\right)$

The predicted probabilities from the multinomial model yield $P\left(Y^{\overline{a}}_{i,J+1}=k\,|\,Y^{\overline{a}}_{i,j}=k'\right)$, the probability of a potential outcome state at time $j+1$ under a specified joint exposure at time j given the outcome state at time j. Although the significance of the causal effect of exposures on multistate outcomes may be tested by the null hypothesis that $\beta^{k',k}=\mathbf{0}$ and the direction of effect inferred by the suggested odds ratio, the interpretation and magnitude of the effect are more readily understood by comparing the marginal probability of the final

outcome state $P\left(Y^{\overline{a}}_{i,J+1}=k\right)$ under various joint exposures. We show in the supplementary

material that the marginal probabilities $\Lambda^{\overline{a}}_{J+1,k}=P\left(Y^{\overline{a}}_{i,J+1}=k\right)$ of the final outcome state are determined by

$$\Lambda^{\overline{a}}_{J+1}=\gamma_0^{\mathrm{T}}\prod_{j=0}^{J}\lambda^{\overline{a}}_{j+1}, \tag{6}$$

where $\Lambda^{\overline{a}}_{J+1}$ is the vector of length K with the kth element of $\Lambda^{\overline{a}}_{J+1}=P\left(Y^{\overline{a}}_{i,J+1}=k\right)$,

$\gamma_0=\left(\gamma_0^1,\gamma_0^2,\ldots,\gamma_0^K\right)^{\mathrm{T}}=\left(P\left(Y_{i,0}=1\right),P\left(Y_{i,0}=2\right),\ldots,P\left(Y_{i,0}=K\right)\right)^{\mathrm{T}}$, characterizing the

marginal distribution of $Y_{i,0}$, and $\lambda^{\overline{a}}_j$ as defined in (2). We consider joint exposures of the type $= (a*,a*,\ldots,a*)$ for $a*$ a specified level of a categorical exposure, although generalization to other joint exposures is straightforward. For our CKD analysis, marginal probabilities are thus interpreted as the probability of a potential outcome state under the case in which a subject maintains the same SBP category for the duration of follow up.

Confidence intervals for $P\left(Y^{\overline{a}}_{i,J+1}\right)$ may be obtained by the nonparametric bootstrap.

3. Simulation Study

We conducted a simulation study to evaluate the properties of our MS-MSM estimator, including evaluating the impact of truncating stabilized weights. [17] demonstrated how to simulate time-to-event outcomes based on a conditional likelihood that induces time-varying confounding and is consistent with a marginal structural model that marginalizes over all covariates and contains a noncollapsible link function. We adapt this approach to simulate transition states that are consistent with our MS-MSM.

Our simulation study considered transitions among four possible states, with state 4 absorbing, determined by the homogenous transition matrix

$$\lambda^{\overline{a}=0}_j=\lambda^{\overline{a}=0}=\begin{bmatrix}0.6 & 0.3 & 0 & 0.1\\ 0 & 0.6 & 0.25 & 0.15\\ 0 & 0 & 0.8 & 0.2\\ 0 & 0 & 0 & 1\end{bmatrix}$$ for the reference exposure of $= (0,0,\ldots,0)$, and the reference transition matrix and MS-MSM for non-reference joint exposures, where the MS-MSM is given by

$$log\left\{\frac{P\left(Y_{i,j+1}^{\overline{a}}=k|Y_{i,j}^{\overline{a}}=k^{'}\right)}{P\left(Y_{i,j+1}^{\overline{a}}=k^{'}|Y_{i,j}^{\overline{a}}=k^{'},\right)}\right\}=\alpha^{k^{'},k}+\beta^{k^{'},k}\sum_{t=0}^{j}a_{i,t},$$

with $\beta^{k^{'},k}=-0.5$ and $\alpha^{k^{'},k}=log\left(\frac{\lambda^{\overline{a}=0,k^{'},k}}{\lambda^{\overline{a}=0,k^{'},k^{'}}}\right)$. The true values of the marginal probabilities $\Lambda_{J+1}^{\overline{a}}$ assuming all subjects started in state 1 were (0.006,0.030,0.152,0.812) for = (0,0,..., 0) and (0.40, 0.307, 0.083,0.210) for = (1,1,...,1) as determined by the product of the true transition intensity matrices for j+1=1 to 10.

Following [17], the data generating algorithm induced time-varying confounding through uniform random variables U_i that were common causes of an intermediate covariate $L_{i,j}$ and outcomes $Y_{i,j+1}$. For each subject, we first generated a vector $\mathbf{U}_{i,0}=\left(U_{i,0}^{1,1},U_{i,0}^{1,2},U_{i,0}^{2,1},U_{i,0}^{2,2},U_{i,0}^{3}\right)$ of independent uniform (0,1) random variables which we used to define thresholds for transition probabilities as we next describe. $Y_{i,0}$ was set to 1 for all subjects, indicating that all subjects started in state 1.

At baseline ($j = 0$) censoring for each subject was set to $C_{i,0} = 0$, indicating that all subjects were present. The baseline covariate $L_{i,0}$ was generated as $\Phi^{-1}(\mu_{i,0}) + \varepsilon_{i,0}$, where Φ^{-1} is the inverse cumulative distribution function for the standard normal distribution and $\varepsilon_{i,0} \sim N(0,0.2)$ and $\mu_{i,0}$ is the mean of all elements of the vector $\mathbf{U}_{i,0}$. Treatment $A_{i,0}$ was Bernoulli($p_{i,0}$), with $logit(p_{i,0}) = -0.8 - 0.8L_{i,0}$. $Y_{i,1}$ followed a multinomial model with

$$log\left\{\frac{P\left(Y_{i,1}=k|Y_{i,0}=1,A_{i,0}\right)}{P\left(Y_{i,1}=1|Y_{i,0}=1,A_{i,0}\right)}\right\}=\alpha^{1,k}+\beta^{1,k}A_{i,0},$$

with $\beta^{1,k}=-0.5$ and $\alpha^{1,k}=log\left(\frac{\lambda^{\overline{a}=0,1,k}}{\lambda^{\overline{a}=0,1,1}}\right)$.

To generate $Y_{i,1}$ consistent with $\lambda^{\overline{a}=0}$ and the MS-MSM, $U_{i,0}^{1,1}$ was compared to $\lambda^{\cdot,1,1}$ ($\lambda^{\cdot,1,1}=0.6$ for $a_{i,0} = 0$ and 0.71 for $a_{i,0}= 1$). For $1 - \lambda^{\overline{a},1,1}<U_{i,0}^{1,1}$, $Y_{i,1}$ was set to $Y_{i,1} = 1$, indicating that the subject remained in state 1 for j=1. For $1 - \lambda^{\overline{a},1,1} \geq U_{i,0}^{1,1}$, $\frac{P\left(Y_{i,1}=2|Y_{i,0}=1,Y_{i,1} \neq 1,A_{i,0}=a_{i,0}\right)}{P\left(Y_{i,1} \neq 1|Y_{i,0}=1,A_{i,0}=a_{i,0}\right)}$, the conditional probability of state 2 given that the subject did not remain in state 1, was then compared to $U_{i,0}^{1,2}$ such that for $U_{i,0}^{1,2}<1 - \frac{P\left(Y_{i,1}=2|Y_{i,0}=1,Y_{i,1} \neq 1,A_{i,0}=a_{i,0}\right)}{P\left(Y_{i,1} \neq 1|Y_{i,0}=1,A_{i,0}=a_{i,0}\right)}$, $Y_{i,1} = 2$ and $Y_{i,1} = 4$ otherwise, noting that $P(Y_{i,1} = 3| Y_{i,0} = 1,A_{i,0}) = 0$ according to $\lambda^{\overline{a}=0}$.

For subsequent time points $j = 1,\ldots,9$, $C_{i,j+1}$ was generated according to the model

$$log \left\{ \frac{P\left(C_{i,j+1}=1 | C_{i,j}=0, \overline{A}_{i,j}, \overline{L}_{i,j}\right)}{1 - P\left(C_{i,j+1}=1 | C_{i,j}=0, \overline{A}_{i,j}, \overline{L}_{i,j}\right)} \right\} = -5.0 - 0.2A_{i,j} + 0.6L_{i,j}$$

. Then for $C_{i,j+1} = 0$ and $Y_{i,j}$ 4, the time-varying confounder $L_{i,j+1}$ was generated as $0.7L_{i,j} - 0.5A_{i,j-1} + \Phi^{-1}(\ _{i0})$, and treatment $A_{i,j+1}$ was Bernoulli($p_{i,j+1}$), with $logit(p_{i,j+1}) = -0.8 - 0.8L_{i,j+1} - 0.4L_{i,j} + 0.3A_{i,j}$. $Y_{i,j+1}$ was generated according to $Y_{i,j}$ as follows based on the reference matrix $\lambda_j^{\overline{a}=0}$ and the multinomial model

$$log \left(\frac{P\left(Y_{i,j+1}=k | Y_{i,j}=k', \overline{A}_i\right)}{P\left(Y_{i,j+1}=1 | Y_{i,j}=1, \overline{A}_{i,j}\right)} \right) = \alpha^{k',k} + \beta^{k',k} \sum_{t=0}^{j} A_{i,t},$$

with $\beta^{k',k} = -0.5$ and $\alpha^{k',k} = log\left(\frac{\lambda^{\overline{a}=0,k',k}}{\lambda^{\overline{a}=0,k',k'}} \right)$.

- For $Y_{i,j}=1$, $Y_{i,j+1}$ was generated as described for the transition from Y_{i0} to Y_{i1} based on λ for

- For $Y_{i,j}=2$, $U_{i,0_{2,1}}$ was compared to $\lambda^{\cdot,2,2}$. For $1 - \lambda^{\overline{a},2,2} < U_{i,0}^{2,1}$, $Y_{i,j+1}=2$, indicating that the subject remained in state 2. For $1 - \lambda^{\overline{a},2,2} \geq U_{i,0}^{2,1}$,

  $$\frac{P\left(Y_{i,j+1}=3 | Y_{i,j}=2, Y_{i,j+1} \neq 2, \overline{A}_{i,j}\right)}{P\left(Y_{i,j+1} \neq 2 | Y_{i,j}=2, \overline{A}_{i,j}\right)} \text{ was then compared to } U_{i,0}^{2,2}, \text{ such that for}$$

  $$U_{i,0}^{2,2} < 1 - \frac{P\left(Y_{i,j+1}=3 | Y_{i,j}=2, Y_{i,j+1} \neq 2, \overline{A}_{i,j}\right)}{P\left(Y_{i,j+1} \neq 2 | Y_{i,j}=2, \overline{A}_{i,j}\right)}, \quad Y_{i,j+1}=3 \text{ and } Y_{i,j+1}=4$$

  otherwise

- For $Y_{i,j}=3$, $U_{i,03}$ was compared to $\lambda^{\cdot,3,3}$ such that for $1 - \lambda^{\overline{a},3,3} < U_{i,0}^3$, $Y_{i,j+1} = 3$, indicating that the subject remained in state 3. For $1 - \lambda^{\overline{a},3,3} \leq U_{i,0}^3$, $Y_{i,j+1} = 4$.

Each replicated dataset had 1000 patients, and the maximum number of time intervals was set to $J = 10$ for subjects who remained uncensored. Results are based on 1000 replicated datasets.

In applying MS-MSMs to the simulated data, time-varying treatment and censoring weights were determined according to the correct models in the denominator, and numerator terms for stabilizing weights included the most recent treatment level $A_{i,j}$ for modeling $C_{i,j+1}$ and $A_{i,j+1}$. Results were generated for unweighted estimates, estimates based on stabilized weights without truncation, and estimates based on weights that were truncated at 98th percentile. Model fitting was stratified by current state, resulting in 5 estimates of the causal log odds ratio parameter for the effect of exposure on transitioning from: 1) State 1 to 2 ($\beta^{1,2}$), 2) State 1 to 4 ($\beta^{1,4}$), 3) State 2 to 3 ($\beta^{2,3}$), 4) State 2 to 4 ($\beta^{2,4}$), and 5) State 3 to 4

($\beta^{3,4}$). For each stratified regression, staying in the current state was the reference transition, hence no model parameters were estimated for that transition.

Simulation results are shown in Tables 1 and 2. Bias of unweighted coefficient estimates ranged from 4.6–35.3%, whereas bias was small for IPW estimates with and without truncation (bias 0.2–7.2%). The IPW estimates, however, were less efficient than unweighted estimates, with Monte Carlo standard errors of estimators based on weights without truncation double to triple the Monte Carlo standard errors of unweighted estimates. The efficiency of weighted estimation was improved by truncation but still had larger standard errors than the unweighted estimates. The robust variance estimator seemed to underestimate the asymptotic variance when weights without truncation were used. For the marginal probability vector $\Lambda^{\overline{a}}_{J+1}$, unweighted estimation resulted in underestimation of the probability of being in the most severe outcome state 4 at the end of follow up for = (0,0, …,0) (0.714 estimated versus 0.812 true) compared to weighted estimation (0.814 without truncation, 0.793 with truncation) and overestimation of the probability of being in outcome states 2 or 3. The opposite occurred for (= (1,1,…,1)), with underestimation of the probability of ending in milder states 1 and 2 without weighting (0.365 estimated versus 0.40 true and 0.286 estimated versus 0.307 true for states 1 and 2, respectively) and overestimation of the probability of ending in more severe states. Bias in the estimation of marginal probabilities was small for the estimated MS-MSM. Compared to weighted estimation without truncation, the marginal probability estimates resulting from weight truncation at the 98th percentile show greater bias but less variability despite similar bias between the coefficient estimates.

4. Analysis of the Chronic Renal Insufficiency Cohort Study

Our study of SBP and CKD progression includes 3,708 participants from the CRIC Study previously described by Anderson et al.(2015), who examined the effect of SBP on time to ESRD (defined as receipt of maintenance dialysis or a kidney transplant) or halving of eGFR[11]. Patients were assessed annually for 0 to 7 years with variable follow up per patient. Over half of patients (2133 (57.5%)) had 5 or more visits; 442 (11.9%) only contributed 1 visit, 317 (8.5%) contributed 8 visits, and median follow up was 5.7 years. Systolic blood pressure was defined categorically at each year of follow up based on the mean of 3 seated measurements as < 120 mmHg ($A_{i,j} = 1$), [120,130) mmHg ($A_{i,j} = 2$), [130,140) mmHg($A_{i,j} = 3$), and 140 mmHg($A_{i,j} = 4$). We defined the CKD outcome state according to levels of eGFR at each year, ESRD, or death, such that $Y_{i,j+1} = 1$ for eGFR 60, $Y_{i,j+1} = 2$ for eGFR [45,60), $Y_{i,j+1} = 3$ for eGFR of [30,45), $Y_{i,j+1} = 4$ for eGFR of (0,30), $Y_{i,j+1} = 5$ for ESRD, and $Y_{i,j+1} = 6$ for death. For patients observed to develop the clinical events of ESRD or death the event was assigned to the first annual visit after the occurrence of the event. At baseline 658 (17.7%) were in state 1 indicating mild CKD, 1041 (28.0%) were in state 2, 1261(34.0%) were in state 3, and 748(20.2%) were in state 4, the most severe level of CKD. A total of 1481 (39.9%) patients had SBP<120 mmHg, 713 (19.2%) were in the interval [120,130), 559 (15.1%) fell within [130,140), and 955 (25.7%) had SBP 140mmHg. For the purpose of our analysis, both ESRD and death were absorbing states such that death indicated death prior to the occurrence of ESRD. Baseline

demographic and clinical variables were sex, race, education level, self-reported hypertension, and baseline values of all time-dependent covariates, where time-dependent covariates were age, presence of cardiovascular disease, diabetes, BMI, use of angiotensin converting enzyme or angiotensin receptor blockers (ACEs/ARBs), number of antihypertensive medications, proteinuria, and past eGFR. Each of the time-dependent covariates potentially confounds the association between blood pressure at any given time and CKD state changes. Additionally, it is plausible that earlier blood pressure values could affect subsequent level of kidney function (i.e., eGFR) or damage (i.e., level of proteinuria), use of anti-hypertensive medications, and occurrence of cardiovascular disease, justifying the exploration of these variables as potential time-dependent confounders.

Loss to follow up was the only censoring event.

The causal model relating state transitions to the previous SBP history is shown in (7)

$$
\begin{aligned}
log & \left(\frac{P\left(Y_{i,j+1}^{\overline{a}}=k \mid Y_{i,j}=k^{'}, \overline{A}_{i,j}\right)}{P\left(Y_{i,j+1}^{\overline{a}}=k^{'} \mid Y_{i,j}=k^{'}, \overline{A}_{i,j}\right)} \right) \\
&= \beta_0^{k^{'},k} + \beta_2^{k^{'},k} \frac{1}{j+1} \sum_{l=0}^{j} I\left(a_{i,l}=2\right) \\
&+ \beta_3^{k^{'},k} \frac{1}{j+1} \sum_{l=0}^{j} I\left(a_{i,l}=3\right) \\
&+ \beta_4^{k^{'},k} \frac{1}{j+1} \sum_{l=0}^{j} I\left(a_{i,j}=4\right) \\
&+ \alpha^{k^{'},k} f\left(j\right),
\end{aligned}
\tag{7}
$$

where $f(j)$ is a spline function for time. In this model, the odds ratio of transition from state $k^{'}$ at time j to state k versus $k^{'}$ is modeled as a function of blood pressure level through time j and time. Coefficients β_a^k may be interpreted as the log odds ratios of transitioning into state k versus $k^{'}$ for always having an SBP level in category a versus always having an SBP level < 120 mm Hg. Our model did not constrain the direction of transitions at all, therefore allowing subjects to transition into improved states. Transitions that were observed fewer than 5 times were assigned zero probability to ensure stability of modeling. These were transitions from eGFR ≥ 60 to eGFR < 30 or ESRD, eGFR [45,60) to ESRD, and eGFR < 30 to eGFR ≥ 60. Our model included a spline term $f(j)$ to flexibly allow for non-homogeneity in the Markov process. The transition diagram corresponding to our analysis is shown in Figure 1.

Stabilized MSM treatment and censoring weights were determined by pooled multinomial and logistic regression, respectively. The final weights were truncated at the 1st and 99th percentile(0.08 and 8.47, respectively). The complete distribution of weights before and after truncation is shown in the supplementary material. Figures 2–5 show the log odds ratios for the effect of blood pressure on transitions among states as estimated by the MS-MSM and standard, unweighted analysis. MS-MSM estimates generally suggest decreased likelihood

of improving and increased likelihood of progressing with increasing blood pressure. Figure 4 particularly shows a strong, monotone effect of blood pressure on transitions to ESRD, with the ≥140 mm Hg level leading to a significantly greater likelihood of transitioning to ESRD than the reference <120 mm Hg blood pressure level. Coefficient estimates from unweighted, standard regression were farther from the null than respective MS-MSM estimates. The estimated marginal probabilities for being in a given state at the end of 7 years if always in a given SBP category are shown in Figure 6 and Table 3 with respective 95% bootstrap confidence intervals based on 1000 bootstrap replicates. Marginal probabilities based on unweighted estimation are shown in Table 4 for comparison. Several significant differences were observed between the joint exposure of always having low (< 120 mm Hg) versus high(≥140 mm Hg) SBP, with favorable effects estimated for low blood pressure. Monotonic relationships were observed for blood pressure effects on mild CKD and ESRD. For increasing levels of SBP, subjects were less likely to be in a mild CKD state at the end of the observation period and more likely to have experienced ESRD. Under consistently low SBP, 22.9% (95% CI 19.4–26.4) of participants would have mild CKD at the end of follow up, whereas under high SBP, only 4.2%(3.1–5.3) of patients would be expected to have mild CKD. The probability of developing ESRD, however, was significantly higher under high SBP (43.4%, 95% CI 38.4%–48.4%) than low SBP (13.4%, 95% CI 11.0%–15.8%). No significant differences were detected for the probability of death as an effect of blood pressure. Similar to the odds ratio estimates, marginal probability estimates based on unweighted estimation overstated the effect of blood pressure (Table 4), suggesting that 9.3% (7.6–11.0) of subjects with consistently low SBP experience ESRD by the end of follow up as compared to 53.9%(50.2–57.6) with consistently high SBP. At any given time interval, it is possible that participants with controlled blood pressure (< 120 mm Hg) were otherwise healthier, as indicated potentially by a better eGFR history, greater use of blood pressure-controlling medications, lack of cardiovascular disease, and other measured patient factors, and thus more likely to have a favorable prognosis than patients with high blood pressure, which could result in incorrectly attributing favorable outcomes to blood pressure such that the standard analysis overstates the beneficial effect of blood pressure. Our time-dependent weights account for such differences throughout the observation period by including these intermediate indicators of overall health in the weights such that estimated effects reflect only blood pressure differences and not these other factors. In general, however, the impact of time-dependent confounding on effect estimates is difficult to disentangle with the large number of time-varying confounders considered that are complexly related to blood pressure and disease progression.

5. Discussion

We have formulated MSMs to estimate the causal effect of time-varying exposures subject to time-dependent confounding on multistate outcomes. Our model allows the consideration of both longitudinal and time-to-event outcomes that are simultaneously accommodated in a multistate outcome framework. This allows the examination of the effect of exposures on progressing to both moderate and advanced disease states and avoids the treatment of clinical outcomes as informative censoring events for examining longitudinal outcome while also accounting for changes in longitudinal outcomes on assessing progression to clinical

endpoints. As with similar marginal model longitudinal approaches applied to panel data, our method easily accommodates unbalanced data characterized by a different number of observations per participant provided that causes for such differential follow up are measured or it is unrelated to unobserved outcome states.

In specifying our model, we make a first order Markov assumption that the transition probabilities depend only on the most recent state and not previous states. We also illustrate how to construct the transition matrices under this assumption and to estimate the marginal state probabilities under a joint exposure through a simple matrix product. When this assumption is violated, the estimates of the MS-MSMs may still be interpreted causally as odds ratios for the likelihood of transitioning from one outcome state to another given the current state and marginalizing over all previous states. However, the estimated marginal probability of each state under a joint exposure at the end of the observation period may not be consistent under our simplified marginalization approach. The dependence of the transition probabilities on additional past outcome states may be tested empirically by adding additional previous outcomes to the MS-MSMs. Evaluation of the first order Markov assumption in the CRIC dataset did reveal that additional prior states were in some cases predictive of transition probabilities. For example, considering all possible transitions from current state 2, i.e., eGFR of [45 – 60), patients who were in state 2 in the previous year were less likely to transition to state 1, i.e., eGFR 60 in the following year than patients who were in state 1 in the previous year. Further considering transitions from current state 2, those who were in eGFR state 3, i.e., eGFR of [15,30), in the previous year were more likely to die than those who were in state 1 the previous year. This encourages the exploration of alternative model specifications and more generalized marginalization techniques. To improve the plausibility of a first order Markov assumption, one can redefine the state representation by including additional prior state information using a summary measure of previous states. For example, one can summarize a patient's trajectory over the past several time periods using a slope term, which can then be used to define states. Clinical knowledge can inform reasonable state definitions that are more likely to satisfy the first order assumption. If a reasonable dimension reduction of state history is not available, the model may be fit under higher-order Markov assumptions. The calculation of marginal probabilities, however, is complicated by marginalizing over multiple previous states such that transition matrices are larger in dimension. For example, when there are K outcome states and transitions depend on both the current and previous state, consecutive yearly transitions may be represented by a matrix of dimension $K^2 \times K^2$, where a state is now defined jointly by consecutive states. Due to the overlap in definition of consecutive joint states, many elements of joint state transition matrices will be zero, which provides some simplification of subsequent matrix products. An alternative approach may consider stratified models for each outcome time $j + 1$ where each model contains the full prior outcome history. This approach would be less efficient as it does not allow borrowing of information across time, but would result in consistent estimation of transition conditional probabilities that could then be entered into a generalized marginalization formula to obtain marginal probabilities at the end of follow up. Stratified models for each outcome time $j + 1$ might be used to develop additional model checking or sensitivity analysis methods to evaluate the impact of violations of the first-order Markov assumption on estimated marginal

probabilities. Our method also relies upon a correctly specified multinomial MSM for the transition probabilities and consistent estimation of the denominators of the censoring and treatment weights. When any of these assumptions are violated the parameter estimates and resulting marginal probabilities are subject to bias for the true average causal odds ratio for exposure. We also assume an additive model for the effects of treatment and time on the transition intensity matrix. This assumption may be relaxed to include interactions of exposures with exogenous time-varying factors such as time, but not endogenous time-varying covariates potentially affected by exposures such as individual patient characteristics[6, 18].

In addition to the first order Markovian assumption, it is important not to lose sight of the dependence of causal inferences under our MS-MSM model to the assumptions of exchangeability, positivity, and consistency. We note, in particular, that the magnitudes of the effect of BP on progression of CKD given in this report appear to be larger than would be expected from the estimated effects of low BP targets on CKD progression reported in previous randomized trials [19, 20, 21]. Thus unmeasured confounding violating the exchangeability assumption cannot be ruled out. In addition, the application of frameworks for causal inference based on counterfactuals has been criticized for applications where the treatment is a biological parameter due to concerns regarding the validity of the consistency assumption [22]. In our example, the concern may be raised that the difference in the values of eGFR that would be observed at two different levels of blood pressure may differ depending on the mechanism by which blood pressure is modified between the two levels being compared; the "effect" of a given change in blood pressure might differ depending on whether the blood pressure level is modified by changes in diet, physical activity, or by different classes of antihypertensive medication. Thus the marginal structural model described in this paper depends on the assumption that the causal effects of a given change in blood pressure on eGFR are independent of the mechanism used to modify blood pressure. The concept that a given modification of blood pressure level produces similar effects on outcomes irrespective of the mechanism by which blood pressure is modified has been advanced in the cardiovascular disease literature.

In our application we treated ESRD as a terminal event. This follows from the original scientific question of causes of CKD progression, for which ESRD is the final stage. Transitions from ESRD to death may be modeled in MS-MSM; however, patients who develop ESRD are expected to differ substantially from patients at milder CKD states in the behavior of their blood pressure. Modeling of the causal effect of SBP on the transition to death would require a separate pooled logistic regression model for exposure status, which, depending on the level of incidence of ESRD in the cohort, may be fit on a relatively small subset of the cohort. In specifying MSMs for multistate outcomes, it is necessary to consider whether a single pooled logistic model is reasonable across all states, or if interactions of treatment and previous state should be considered, and whether enough variability in states is observed to estimate such interactions.

Our analysis is consistent with previous time-to-event MSM Cox analysis which showed increased risk of ESRD for patients with higher blood pressure [11]. Novel to our analysis are the trends noted in transitions among moderate CKD states denoted by changes in eGFR.

We anticipate that these methods can be used in further studies of CKD and other chronic conditions for which both longitudinal and clinical endpoints are important.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. McConnell S, Stuart E, Devaney B. The truncation by death problem. what do do in an experimental evaluation when the outcome is not always measured. Evaluation Review. 2008; 32:157–186. [PubMed: 18319423]

2. Rubin D. Causal inference through potential outcomes and principal stratification: application to studies with "censoring" due to death. Statistical Science. 2006; 21:299–309.

3. McCrink LM, Marshall AH, Cairns KJ. Advances in joint modelling: A review of recent developments with application to the survival of end stage renal disease patients. International Statistical Review. 2013:249–269.

4. Asar O, Ritchie J, Kalra PA, Diggle PJ. Joint modelling of repeated measurement and time-to-event data: an introductory tutorial. International Journal of Epidemiology. 2015:334–344. [PubMed: 25604450]

5. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology. 2000; 11(5):550–560. [PubMed: 10955408]

6. Robins, JM. Marginal structural models. Proceedings of the American Statistical Association Section on Bayesian Statistical Science; 1998. p. 1-10.

7. Gran JM, Lie SA, Oyeflaten I, Borgan O, Aalen OO. Causal inference in multi-state models-sickness absence and work for 1145 participants after work rehabilitation. BMC Public Health. 2015; 15(1082)

8. Moodie EEM, Stephens DA, Klein MB. A marginal structural model multiple-outcome survival data: assessing the ikmpact of injection drug use on several causes of death in the canadian co-infection cohort. Statistics in Medicine. 2014:1409–1425. [PubMed: 24272681]

9. Li L, Astor BC, Lewis J, Hu B, Appel LJ, Lipkowitz MS, Toto RD, Wang X, Wright JT, Greene TH. Longitudinal progression trajectory of gfr among patients with ckd. American Journal of Kidney Diseases. 2012:504–512. [PubMed: 22284441]

10. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology. 1974; 6:688–701.

11. Anderson AH, Yang W, Townsend RR, Pan Q, Chertow GM, Kusek JW, Charleston J, He J, Kallem R, Lash JP, et al. Time-updated systolic blood pressure and the progression of chronic kidney disease. Annals of Internal Medicine. 2015:258–265. [PubMed: 25686166]

12. Rubin DB. Which ifs have causal answers. Journal of the American Statistical Association. Dec; 1986 81(396):961–962.

13. Cole SR, Frangakis CE. the consistency statement in causal inference: A definition or an assumption. Epidemiology. 2009; 20:3–5. [PubMed: 19234395]

14. Liang KY, Zeger S. Longitudinal data analysis using generalized linear models. Biometrika. 1986:13–22.

15. MA H, Brumback B, Robins JM. Marginal structural models to estimate the joint causal effect of nonrandomized treatments. Journal of the American Statistical Association. 2001; 96:440–448.

16. Efron, B., Tibshirani, RJ. An Introduction To The Bootstrap. Chapman & Hall/CRC; Boca Raton: 1993.

17. Havercroft WG, Didelez V. Simulating from marginal structural models with time-dependent confounding. Statistics in Medicine. 2012; 31(30):4190–4206. [PubMed: 22826156]

18. Robins, JM. Marginal structural models versus structural nested models as tools for causal inference. In: E, H., Berry, D., editors. Statistical Methods in Epidemiology: The Environment and Clinical Trials. Vol. 95. Springer-Verlag; New York, NY: 1999. p. 134

19. Peterson JC, Adler S, Burkart JM, Greene T, Herbert LA, Hunsicker LG, Seifter JL. Blood pressure control, proteinuria, and the progression of renal disease: the modification of diet in renal disease study. Annals of Internal Medicine. 1995; 123(10):754–762. [PubMed: 7574193]

20. Wright JT Jr, Bakris G, Greene T, Agodoa LY, Appel LJ, Charleston J, Cheek D. Effect of blood pressure lowering and antihypertensive drug class on progression of hypertensive kidney disease: results from the aask trial. JAMA. 2002; 288(19):2421–2431. [PubMed: 12435255]

21. Wright JT Jr, Williamson JD, Whelton PK, Snyder JK, Sink KM, Rocco MV, Reboussin DM. A randomized trial of intensive versus stnadard blood-pressure control. New England Journal of Medicine. 2015; 373(22):2103–2116. [PubMed: 26551272]

22. Greenland S. Epidemiologic measures and policy formulation: lessons from potentional outcomes. Emerging Themes in Epidemiology. 2005
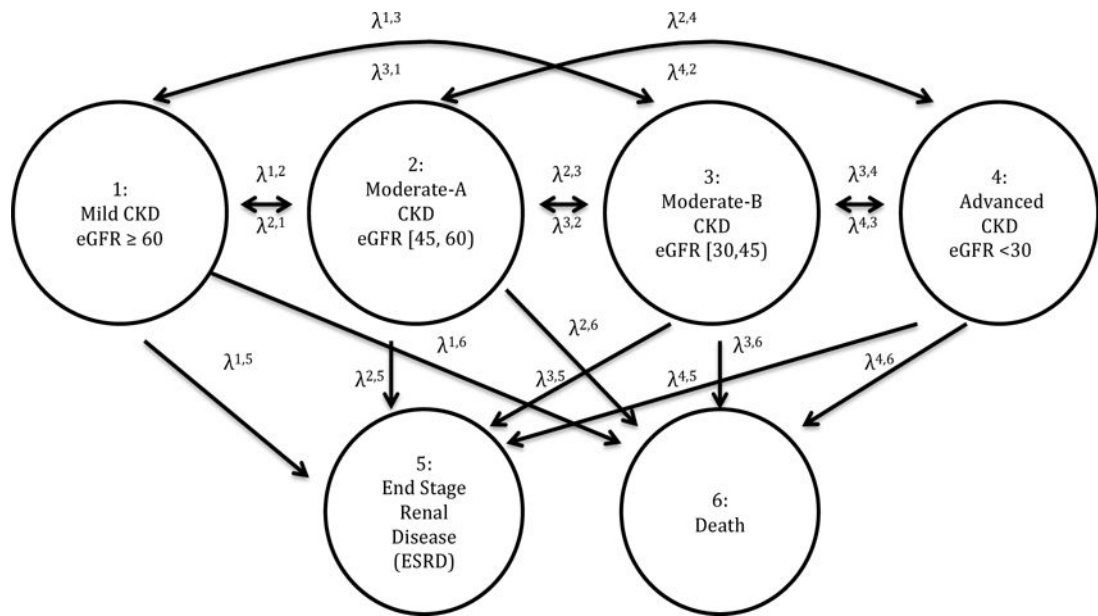
Figure 1. Transition Diagram for Chronic Kidney Disease Progression

Parameters $\lambda^{k',k}$ denotes the probability of transition from state k' to state k. The model allowed for transition to more or less severe states defined by eGFR. Death and ESRD were absorbing states.
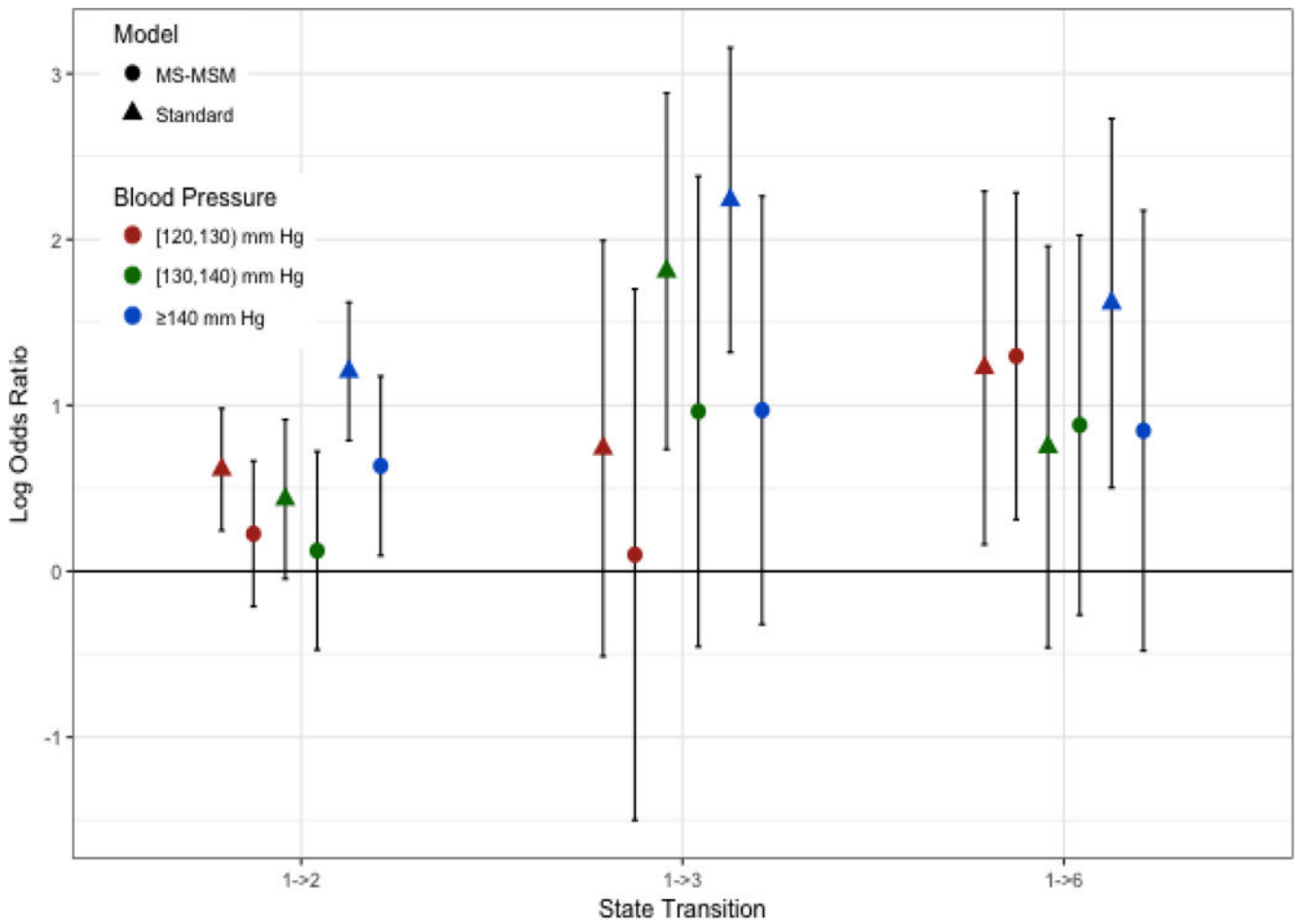
Figure 2. Log Odds Ratios of Blood Pressure Effects for Transitions from State 1
Log odds ratios shown are for the odds of transitioning into the indicated state versus staying
in state 1 for always being in the indicated blood pressure category versus always being in
the reference state of < 120 mm Hg

Figure 3.
Log Odds Ratios of Blood Pressure Effects for Transitions from State 2. Log odds ratios shown are for the odds of transitioning into the indicated state versus staying in state 2 for always being in the indicated blood pressure category versus always being in the reference state of < 120 mm Hg

Figure 4.
Log Odds Ratios of Blood Pressure Effects for Transitions from State 3. Log odds ratios
shown are for the odds of transitioning into the indicated state versus staying in state 3 for
always being in the indicated blood pressure category versus always being in the reference
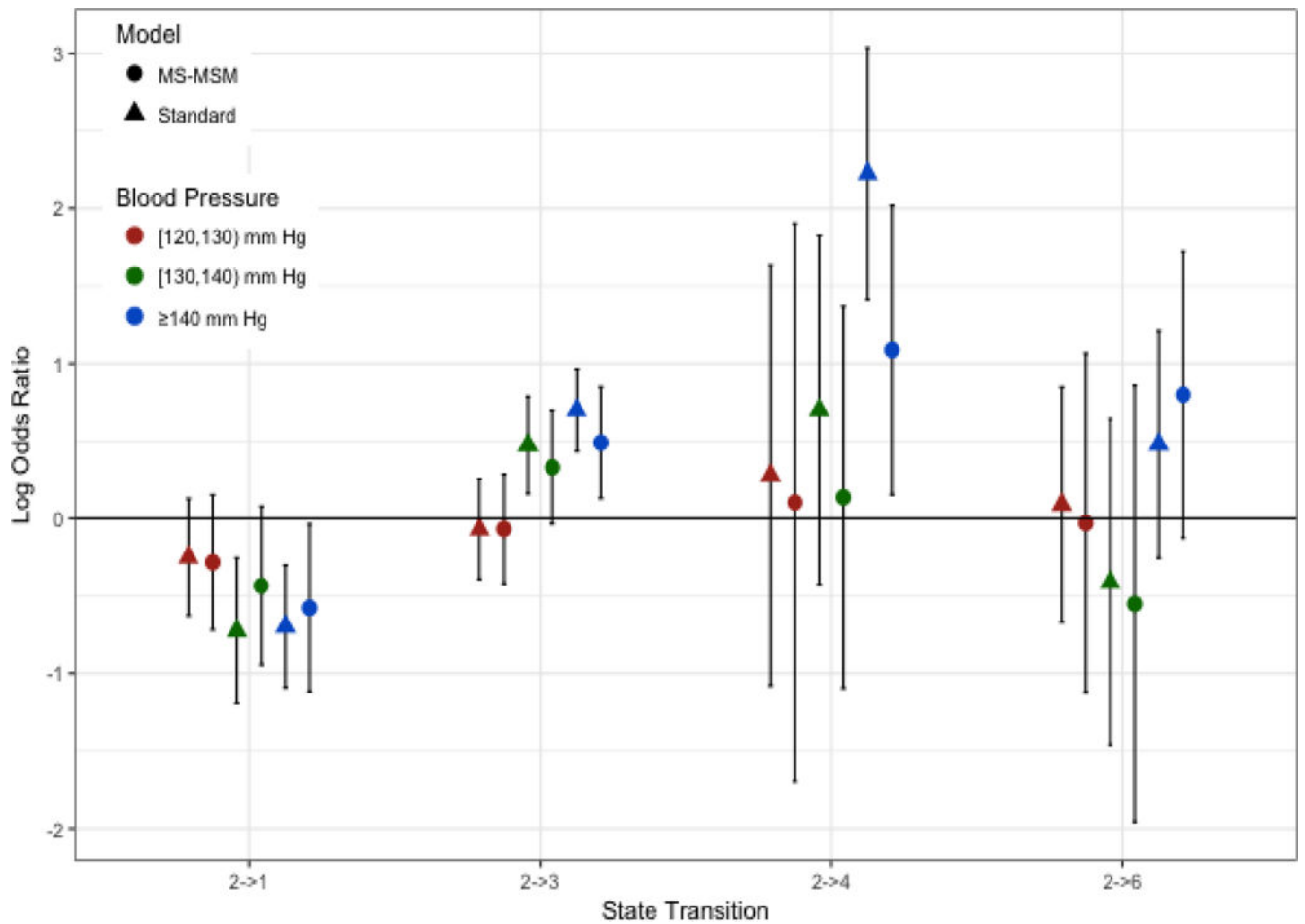state of < 120 mm Hg

Figure 5.
Log Odds Ratios of Blood Pressure Effects for Transitions from State 4. Log odds ratios shown are for the odds of transitioning into the indicated state versus staying in state 4 for always being in the indicated blood pressure category versus always being in the reference state of < 120 mm Hg
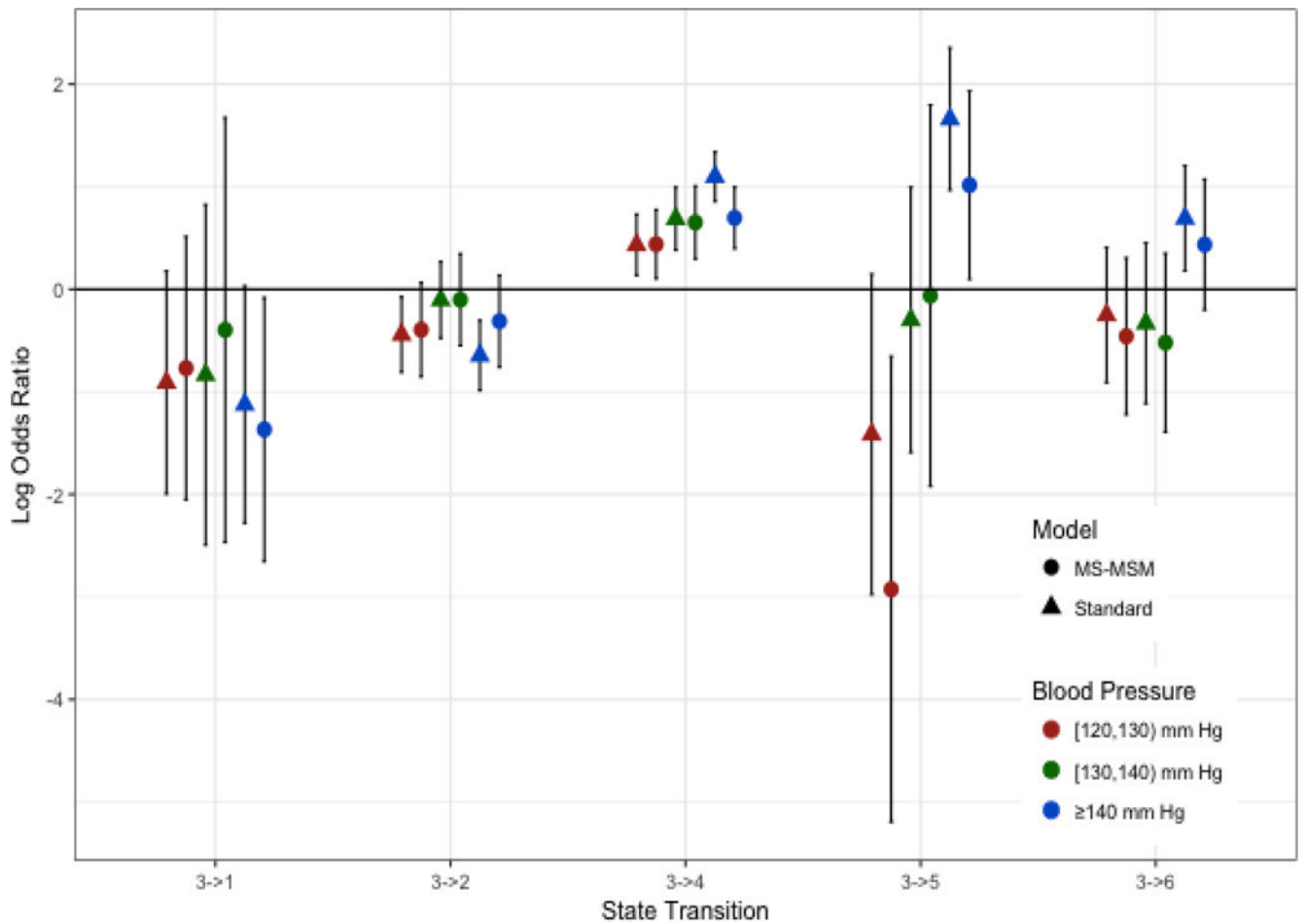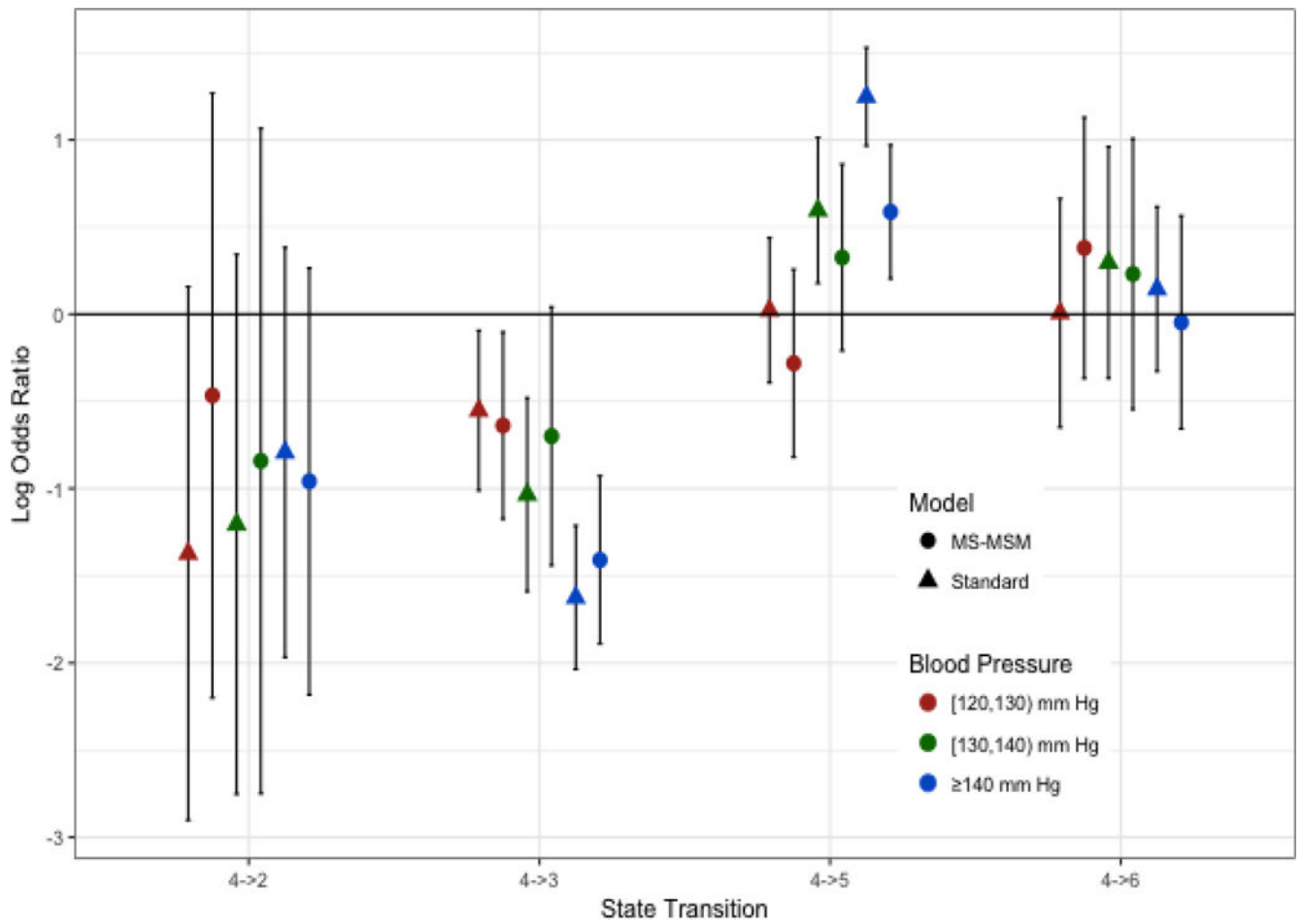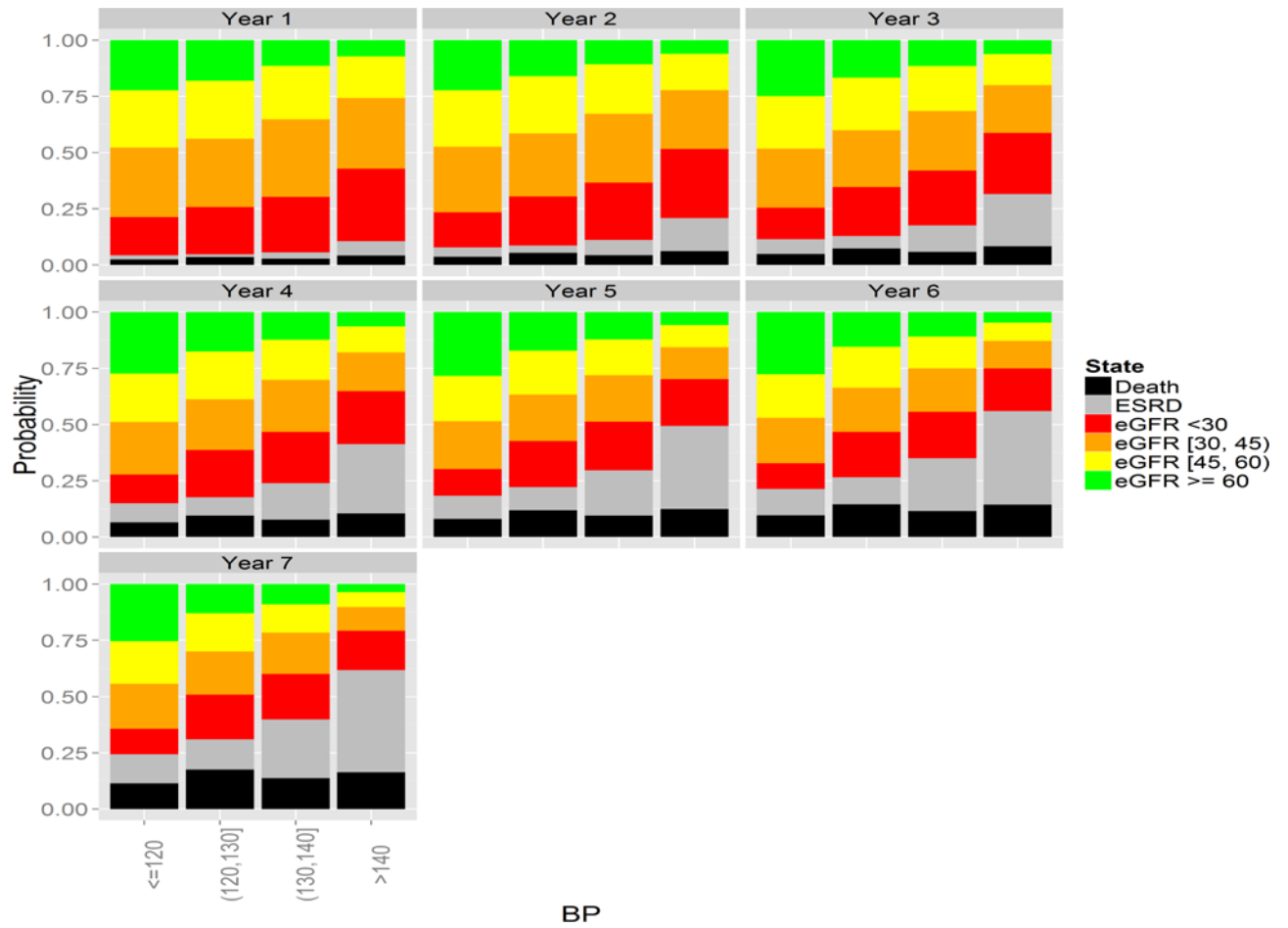
Figure 6.
Yearly Marginal Distribution of States by Systolic Blood Pressure Trajectory.

Table 1

Simulation Study Results: MS-MSM Coefficients

Estimation of Multistate Marginal Structural Model (MS-MSM) parameters using 1) Standard, unweighted regression, 2) Inverse Probability Weighting without weight truncation, and 3) Inverse Probability Weighting with weight truncation at the 98th percentile. Parameters $\hat{\beta}_{k'}^{'}$ indicate the effect of treatment in the logit in the causal multinomial transition state model for transition from state k' to state k. Estimation was stratified by k', and for each model, k' was the reference for the multinomial transition model. Mean Estimate – sample mean of estimated causal parameters where the true value is ψ $= -0.5$. Bias – the difference of the true parameter and the mean estimate. —%Bias— – Absolute value of the bias divided by true parameter × 100%.

Estimator	Parameter	Mean Estimate	Bias	%Bias	MCSE	Mean ASE
Unweighted	$\beta^{1,2}$	-0.47	0.02	4.57	0.037	0.038
	$\beta^{1,4}$	-0.37	0.14	27.0	0.055	0.054
	$\beta^{2,3}$	-0.46	0.04	7.27	0.044	0.043
	$\beta^{2,4}$	-0.32	0.18	35.3	0.049	0.048
	$\beta^{3,4}$	-0.37	0.13	25.2	0.07	0.07
Weighted	$\beta^{1,2}$	-0.51	-0.01	1.32	0.095	0.075
	$\beta^{1,4}$	-0.52	-0.02	4.02	0.122	0.100
	$\beta^{2,3}$	-0.51	-0.01	2.71	0.122	0.090
	$\beta^{2,4}$	-0.52	-0.02	3.65	0.130	0.107
	$\beta^{3,4}$	-0.54	-0.04	7.22	0.193	0.134
Weights Truncated at 98th Percentile	$\beta^{1,2}$	-0.50	0.00	0.20	0.055	0.055
	$\beta^{1,4}$	-0.49	0.01	1.92	0.088	0.086
	$\beta^{2,3}$	-0.49	0.01	1.25	0.066	0.064
	$\beta^{2,4}$	-0.47	0.03	6.53	0.079	0.800
	$\beta^{3,4}$	-0.48	0.02	3.64	0.111	0.102

MCSE – Monte Carlo Standard Error calculated as the standard deviation of $\hat{\beta}$ estimates across 1000 simulated datasets. Mean ASE – Mean Asymptotic Standard Error, calculated as the sample average of asymptotic standard error estimates across 1000 replicates.

Table 2

Simulation Study Results: Marginal Probability

$P\left(Y_{J+1}^{\bar{a}}=k\right)$. Estimation of the marginal probability of state at the end of follow up 1) Standard, unweighted regression, 2) Inverse Probability Weighting without weight truncation, and 3) Inverse Probability Weighting with weight truncation at the 98th percentile. The true values of $P(Y=k)$ were (0.0006, 0.030, 0.152, 0.812), and (0.40, 0.307, 0.082, 0.21) for states 1–4 for always control and always treated joint exposores, respectively. Mean Estimate – sample mean of estimated marginal probability over 1000 replicates. % Bias – the difference of the true parameter and the mean estimate divided by true parameter × 100%.

Estimator		=(0,0,…0) (Control)				=(1,1,….1) (Treated)		
	Parameter	Mean Estimate	% Bias	MCSE	Parameter	Mean Estimate	% Bias	MCSE
Unweighted	$P\left(Y_{J+1}^{\bar{a}}=1\right)$	0.007	21.0	0.0014	$P\left(Y_{J+1}^{\bar{a}}=1\right)$	0.365	−9.06	0.026
	$P\left(Y_{J+1}^{\bar{a}}=2\right)$	0.044	44.3	0.0058	$P\left(Y_{J+1}^{\bar{a}}=2\right)$	0.286	−6.79	0.0222
	$P\left(Y_{J+1}^{\bar{a}}=3\right)$	0.235	54.4	0.0234	$P\left(Y_{J+1}^{\bar{a}}=3\right)$	0.087	5.19	0.0113
	$P\left(Y_{J+1}^{\bar{a}}=4\right)$	0.714	−12.0	0.0244	$P\left(Y_{J+1}^{\bar{a}}=4\right)$	0.262	25.3	0.0212
Weighted	$P\left(Y_{J+1}^{\bar{a}}=1\right)$	0.006	0.80	0.0017	$P\left(Y_{J+1}^{\bar{a}}=1\right)$	0.398	−0.67	0.0667
	$P\left(Y_{J+1}^{\bar{a}}=2\right)$	0.030	0.10	0.0086	$P\left(Y_{J+1}^{\bar{a}}=2\right)$	0.305	−0.57	0.0674
	$P\left(Y_{J+1}^{\bar{a}}=3\right)$	0.150	−1.49	0.0435	$P\left(Y_{J+1}^{\bar{a}}=3\right)$	0.083	0.78	0.0386
	$P\left(Y_{J+1}^{\bar{a}}=4\right)$	0.814	0.27	0.0456	$P\left(Y_{J+1}^{\bar{a}}=4\right)$	0.213	1.82	0.0451
Weights Truncated at 98th Percentile	$P\left(Y_{J+1}^{\bar{a}}=1\right)$	0.006	3.04	0.0014	$P\left(Y_{J+1}^{\bar{a}}=1\right)$	0.395	−1.43	0.0369

Estimator		$\bar{a}=(0,0,\ldots 0)$ (Control)				$\bar{a}=(1,1,\ldots 1)$ (Treated)		
	Parameter	Mean Estimate	% Bias	MCSE	Parameter	Mean Estimate	% Bias	MCSE
	$P\left(Y^{\bar{a}}_{J+1}=2\right)$	0.032	6.97	0.0057	$P\left(Y^{\bar{a}}_{J+1}=2\right)$	0.302	−1.77	0.0326
	$P\left(Y^{\bar{a}}_{J+1}=3\right)$	0.168	10.5	0.0281	$P\left(Y^{\bar{a}}_{J+1}=3\right)$	0.084	1.87	0.0162
	$P\left(Y^{\bar{a}}_{J+1}=4\right)$	0.793	−2.25	0.0293	$P\left(Y^{\bar{a}}_{J+1}=4\right)$	0.219	4.59	0.0259

MCSE – Monte Carlo Standard Error calculated as the standard deviation of estimates across 1000 simulated datasets.

Table 3

Marginal Probability of State at the End of Observation Period

Weighted Estimates. Marginal probabilities are shown as the percentage of subjects in the population who would be in a given state at the end of the observation period under the joint exposure of always being in the indicated SBP category. 95% Confidence intervals were obtained by the bootstrap percentile method with 1000 replicates.

| SBP | eGFR Level | | | | Absorbing | |
	60	[45,60)	[30,45)	< 30	ESRD	Death
< 120	22.9 (19.4,26.4)	18.9 (16.0, 21.8)	20.5 (17.6, 23.4)	11.5 (9.3, 13.7)	13.4 (11.0, 15.8)	12.7 (9.7, 15.7)
[120,130)	14.3 (9.9, 18.7)	16.0 (11.9, 20.1)	18.5 (14.0, 23.0)	19.9 (15.1, 24.7)	14.0 (9.5, 18.5)	17.3 (11.7, 22.9)
[130,140)	10.3 (6.5, 14.1)	12.2 (8.4, 16.0)	17.2 (12.6, 21.8)	19.4 (13.7, 25.1)	27.8 (20.0, 35.6)	13.1 (6.8, 19.4)
140	4.2 (3.1, 5.3)	7.3 (5.8, 8.8)	11.1 (9.3, 12.9)	17.5 (14.6, 20.4)	43.4 (38.4, 48.4)	16.6 (12.2, 21.1)

Table 4

Marginal Probability of State at the End of Observation Period

Unweighted Estimates. Marginal probabilities are shown as the percentage of subjects in the population who would be in a given state at the end of the observation period under the joint exposure of always being in the indicated SBP category. 95% Confidence intervals were obtained by the bootstrap percentile method with 1000 replicates.

SBP	eGFR Level					Absorbing	
	60	[45,60)	[30,45)	< 30	ESRD	Death	
<120	27.4 (24.8, 30.0)	19.9 (17.7, 22.1)	20.6 (18.4, 22.8)	10.9 (9.2, 12.6)	9.3 (7.6, 11.0)	11.9 (9.7, 14.1)	
[120,130)	14.2 (10.9, 17.5)	18.1 (14.9, 21.3)	20.2 (16.9, 23.5)	19.7 (16.3, 23.1)	13.3 (10.1, 16.5)	14.5 (10.6, 18.4)	
[130,140)	8.0 (5.2, 10.8)	12.4 (9.5, 15.3)	18.3 (15.0, 21.6)	20.7 (16.5, 24.9)	26.4 (20.7, 32.1)	14.2 (9.6, 18.8)	
140	2.1 (1.2, 3.0)	4.7 (3.6, 5.8)	8.1 (6.8, 9.4)	15.1 (13.0, 17.2)	53.9 (50.2, 57.6)	16.1 (13.0, 19.2)	