History-Adjusted Marginal Structural Analysis of the Association between Hemoglobin Variability and Mortality among Chronic Hemodialysis Patients

Steven M. Brunelli,*† Marshall M. Joffe,‡‡ Rubeen K. Israni,*† Wei Yang,† Steven Fishbane,§ Jeffrey S. Berns,* and Harold I. Feldman*†‡

*Renal, Electrolyte and Hypertension Division of the Department of Medicine, ‡Center for Clinical Epidemiology and Biostatistics, and ‡Department of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania, and §Winthrop University Hospital, Mineola, New York

Background and objectives: Hemoglobin variability is common among dialysis patients, and has been associated with increased mortality. The causal nature of this association has been difficult to ascertain because of potential time-dependent confounding, for which traditional statistical methods do not control.

Design, settings, participants, & measurements: A retrospective cohort of 34,963 Fresenius Medical care dialysis patients from 1996 was assembled. Hemoglobin variability, absolute hemoglobin level, and temporal hemoglobin trend were measured over rolling 6-mo exposure windows. Their association with mortality was estimated using history-adjusted marginal structural analysis that adjusts for time-dependent confounding by applying weights to observations inversely related to the predictability of observed levels of hemoglobin.

Results: In the primary analysis, each g/dl increase in hemoglobin variability was associated with an adjusted hazard ratio (HR) [95% confidence interval (CI)] for all-cause mortality of 1.93 (1.20 to 3.10). Neither higher absolute hemoglobin level nor increasing hemoglobin trend were significantly associated with mortality; adjusted HR (95% CI) 0.85 (0.64 to 1.11) and 0.60 (0.25 to 1.45), respectively.

Conclusions: Marginal structural analysis demonstrates that hemoglobin variability is associated with increased mortality among chronic hemodialysis patients, and that this effect is more pronounced than appreciated using standard statistical techniques that do not take time-dependent confounding into account.


Emodialysis patients experience substantial changes in hemoglobin concentration over time (1–4). This variability derives both from patterns of care, such as changes in erythropoietin and intravenous iron dosing, as well as from comorbid conditions that either influence sensitivity to these therapies, or that directly affect red cell mass. Recently, we demonstrated an association between increased hemoglobin variability (Hgb-Var) and all-cause mortality among chronic dialysis patients (5).

As is the case in all nonrandomized research, the observed association between Hgb-Var and death was subject to potential confounding. Confounding occurs when a variable (or variables) affects the likelihood of exposure (e.g. Hgb-Var) and of outcome (e.g. death), but does not serve as an intermediate between exposure and outcome. Standard methods of statistical adjustment are often sufficient to adjust for these confounders when they do not vary over time compared with their baseline values.

Unfortunately, there are instances when application of standard statistical techniques may yield biased associations, such as in the presence of time-dependent confounding. Time-dependent confounders (TDC) are variables that have two different roles that are difficult to distinguish: (1) they influence subsequent exposure and outcome like confounders that are fixed over time, and (2) they also serve as intermediates (or links) between exposure and outcome. Although it is appropriate to adjust for their role as confounders, it is inappropriate to adjust for their role as intermediates on the causal pathway from exposure to outcome. If standard statistical techniques are used to adjust for TDC, they adjust for both roles simultaneously and, therefore, may underestimate true associations owing to the adjustment for the role as a causal pathway factor. For example, if Hgb-Var is both influenced by comorbid conditions and promotes these same conditions as part of the mechanism through which it causes poor clinical outcomes, simple adjustment for these comorbidities would lead to biased estimates of the Hgb-Var outcome relationship. Figure 1 serves to illustrate one potential TDC of the association between Hgb-Var and death, namely intravenous iron administration, which

Received October 8, 2007. Accepted February 1, 2008.

Published online ahead of print. Publication date available at www.cjasn.org.

Correspondence: Dr. Harold I. Feldman, Center for Clinical Epidemiology and Biostatistics, 923 Blockley Hall, 423 Guardian Drive, Philadelphia, Pennsylvania 19104. Phone: 215-898-0901; Fax: 215-898-0643; E-mail: hfeldman@mail.med.upenn.edu

Copyright © 2008 by the American Society of Nephrology

ISSN: 1555-9041/303–0777
may independently be associated with Hgb-Var and mortality (Figure 1A), and may also serve on the causal pathway (Figure 1B).

The best means to avoid the possibility of time-dependent confounding is through randomization to fixed treatment protocols (9). Extending the prior example, this would mean randomizing subjects to high and low levels of Hgb-Var, and treating each group with equivalent amounts of intravenous iron. Unfortunately, current therapies do not enable randomization of subjects to varying degrees of Hgb-Var. Nonetheless, observational research can take advantage of newer statistical tools, such as history-adjusted marginal structural models (HA-MSM), which enable estimation of less biased measures of association from observational data in the presence of time-dependent confounding (8,10–11).

Considering that the prior association may have been biased by the presence of time-dependent confounders, we have reanalyzed our data using HA-MSM to better estimate the causal association between Hgb-Var and mortality.

Materials and Methods
Description of the Cohort

Details of the cohort have been previously published (5). Briefly, we assembled an historical cohort of incident and prevalent adult subjects who remained both alive and enrolled at one of the Fresenius Medical Care outpatient dialysis units over either (or both) of the first or last 6 mo of 1996. Eligible subjects survived for a minimum of 6 mo to allow for adequate time over which to characterize Hgb-Var at “baseline.” All data were abstracted from existing electronic records.

Demographic data on age, gender, race, duration of ESRD and diabetic status (adult, juvenile, none) were collected at baseline. Hospitalization data and data on infections were not available. Laboratory values and medication exposure data, including hemoglobin, urea reduction ratio, Kt/V, albumin, bicarbonate, aspartate aminotransferase, calcium, phosphate, intact parathyroid hormone, ferritin, transferrin saturation, iron level, intravenous iron dose, and erythropoietin dose were updated monthly. Data on activated vitamin D treatment were not available.

The primary exposures of interest were Hgb-Var, absolute level of hemoglobin, and temporal trend in hemoglobin. To derive these values, a best-fit (ordinary least squares) line of hemoglobin over time was plotted for each subject: absolute hemoglobin was the intercept of this line; temporal hemoglobin trend was the slope of this line; Hgb-Var was the residual SD (the SD of the individual distances between observed hemoglobin values and the line) (5). Hgb-Var cannot be defined at a single point in time. To balance between maximizing accuracy of Hgb-Var measurements (which is enhanced by longer-exposure windows) and minimizing delay between Hgb-Var exposure and outcome (which is enhanced by narrower windows), we defined hemoglobin parameters over 6-mo windows.

The outcome of interest was death from any cause. Subjects were followed from the first date of their qualifying enrollment period (January 1, 1996 or July 1, 1996); follow-up time began on the first day of the tenth month (allowing for a 3-mo covariate assessment period followed by a 6-mo hemoglobin exposure window), and continued until death or censoring occurred. There were a total of 527,967 patient-months of potential follow-up. Censoring criteria included transfer of care, renal transplantation, or study end: September 30, 1998.

All analyses were performed in SAS 9.1 (Cary, NC).

Outcome Model

The primary outcome was based on a logistic regression analog of the history-adjusted marginal structural Cox proportional hazards model (8,11). To fit this model, each subject-month was considered independently. (Robust variance estimates were used to account for nonindependence of observations within subject.) For each subject-month ($t$), survival was predicted based on the hemoglobin parameters (Hgb-Var, absolute level of hemoglobin, and hemoglobin trend) calculated over the preceding 6-mo window ($t - 6$ through $t - 1$). Adjustment was made for covariates measured over the 3-mo period preceding this window (months $t - 9$ through $t - 7$) (Figure 2). Covariates were considered before the hemoglobin exposure window because prior values of these covariates could be confounders, but could not serve as causal pathway intermediates. Concurrent values of covariates (those from months $t - 6$ to $t - 1$), which may have either “confounding” or “pathway intermediate” effects were adjusted for by application of weights (see Appendix). Analysis was restricted to subject-months for which all associated hemoglobin and covariate data were available (260,028 and 390,621 follow-up months for weighted and unweighted models, respectively). To facilitate clinical interpretation, HR are expressed per g/dl (Hgb-Var and absolute hemoglobin) or g/dl/mo (hemoglobin trend). (Given that 1 g/dl has different meaning for each parameter, caution is advised in contrasting the relative effects of each parameter on mortality.)

Marginal structural models adjust for time-dependent confounding by applying weights to person-months of follow-up that are inversely related to the level of concordance between the predicted hemoglobin and the actual observed hemoglobin. In this fashion, subjects whose hemoglobin was least explained by their prior characteristics contributed most to the analysis, thereby minimizing the influence of hemoglobin as an intermediate outcome variable in the estimation of the Hgb-Var/mortality relationship (8,10–11). For example, a healthy, compliant patient having, as expected, a high hemoglobin level, or a noncompliant patient with chronic infection having, as expected, a much lower hemoglobin level, would both carry little weight in the outcome model. However, if the healthy compliant patient had low hemoglobin, this would be unexplained by available characteristics, and this observation would be weighted more heavily in the outcome model. (Details on calculation of weights can be found in Appendix).
To examine the effects of marginal structural analysis, we fit unweighted models (otherwise analogous to the HA-MSM) as comparators.

**Results**

The cohort consisted of 34,963 subjects: mean (SD) in age and duration of ESRD were 60.1 (15.1) yr and 3.6 (3.6) yr, respectively. The cohort was 50.6% men, 48.9% Caucasian, 44.7% African-American, and 45% diabetic (11.1% of these were of juvenile onset).

During follow-up, 9160 subjects (26.2%) died, 9965 (28.5%) transferred care or were lost to follow-up, and 15,838 (45.3%) survived to study end.

Hgb-Var was distributed with a mean (SD) of 0.60 (0.33) g/dl and median (interquartile range (IQR)) of 0.53 (0.36 to 0.76) g/dl. Absolute level of hemoglobin was distributed with mean (SD) 10.25 (1.25) g/dl. Temporal trend in hemoglobin was distributed with a mean (SD) of 0.02 (0.29) g/dl/mo (5). [To facilitate clinical interpretation, HR (below) are reported in clinical units (g/dl or g/dl/mo). A difference of 1 g/dl in Hgb-Var represents a greater relative difference than a difference of 1 g/dl in absolute hemoglobin level, owing to the tighter distribution (smaller SD) of the former.]

**Association between Hemoglobin Parameters and Death**

In the marginal structural analysis, each g/dl rise in Hgb-Var was associated with an adjusted HR (95% CI) for death of 1.93 (1.20 to 3.10). Using an unweighted model, which does not address time-dependent confounding, the HR (95% CI) was 1.34 (1.24 to 1.44; Table 1).

In the marginal structural analysis, each g/dl increase in absolute hemoglobin level was associated with an adjusted HR (95% CI) for death of 0.85 (0.64 to 1.11). Using an unweighted model, the HR (95% CI) was 0.71 (0.68 to 0.74).

In the marginal structural analysis, each g/dl/mo increase in temporal hemoglobin trend was associated with a HR (95% CI) for death of 0.60 (0.25 to 1.45). Using an unweighted model, the HR (95% CI) was 0.28 (0.24 to 0.32).

**Discussion**

Hemoglobin variability is common among patients treated with chronic maintenance dialysis (1–4). Previously, we dem-

---

**Table 1.** Hazard ratios for hemoglobin variability (Hb-Var), absolute level of hemoglobin (Absolute-Hb), and temporal trend in hemoglobin (Trend-Hb) in history-adjusted marginal structural and unweighted analyses

<table>
<thead>
<tr>
<th></th>
<th>History-Adjusted Marginal Structural Model</th>
<th>Unweighted Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hb-Var (per 1 g/dl)</td>
<td>1.93 (1.20 to 3.10), <em>P</em> = 0.005</td>
<td>1.34 (1.24 to 1.44), <em>P</em> &lt; 0.001</td>
</tr>
<tr>
<td>Absolute-Hb (per 1 g/dl)</td>
<td>0.85 (0.64 to 1.11), <em>P</em> = 0.24</td>
<td>0.71 (0.68 to 0.74), <em>P</em> &lt; 0.001</td>
</tr>
<tr>
<td>Trend-Hb (per 1 g/dl/mo)</td>
<td>0.60 (0.25 to 1.45), <em>P</em> = 0.26</td>
<td>0.28 (0.24 to 0.32), <em>P</em> &lt; 0.001</td>
</tr>
</tbody>
</table>

*Hazard ratios are expressed as point estimate (95% confidence interval); associated *P* values are also displayed. Estimates are adjusted for other variables in the table, as well as for age, race, gender, duration of ESRD, diabetes, hemoglobin, urea reduction ratio, Kt/V, albumin, bicarbonate, aspartate aminotransferase, calcium, phosphate, intact parathyroid hormone, ferritin, transferrin saturation and iron levels, as well as for treatment with intravenous iron and erythropoietin.
onstrated an association between higher levels of Hgb-Var and mortality among chronic hemodialysis patients, (5) however these estimates were potentially biased due to time-dependent confounding. In the study presented here, we used history-adjusted marginal structural analysis to adjust for potential time-dependent confounding, and demonstrated that the association between Hgb-Var and mortality was more potent than previously appreciated. Each g/dl increase in Hgb-Var was associated with a 93% increase in all-cause mortality. Conversely, the protective effects of higher hemoglobin levels and temporal hemoglobin trend appeared to be less pronounced and failed to reach statistical significance.

Several reports have demonstrated substantial Hgb-Var among dialysis patients (1–4). This observed variability may derive from patterns of care, and from comorbidities that influence sensitivity to these therapies. Only one previous report has examined the association between Hgb-Var and mortality, demonstrating a 33% increase in mortality per g/dl increase in Hgb-Var (5). (These findings were very similar to those yielded by the unweighted model in this study, differing only because of use of rolling versus static exposure windows.) However, findings from that study were likely biased by the effects of time-dependent confounding, explaining the discrepancy between its findings and those in study presented here.

Other than in chronic diseases such as kidney failure, hemoglobin levels are typically maintained within a narrow range to ensure consistent oxygen delivery to peripheral tissues. Repeated oscillations in hemoglobin may result in interruption of tissue oxygen delivery, with resulting ischemic tissue damage and inappropriate activation of cardiovascular compensatory mechanisms such as cardiac myocyte hypertrophy and left ventricular hypertrophy and dilation (24–26). In addition, hemoglobin variation has been shown to promote autonomic dysfunction, which in turn predisposes patients to sudden death (27,28).

History-adjusted marginal structural models use inverse probability of exposure weighting (or a stabilized analog) to adjust for the confounding effects of TDC, but not their pathway intermediate effects (8,10,11). In the absence of unmeasured confounders, weighting in this manner eliminates bias due to time-dependent confounding. Unfortunately, their application simultaneously creates imprecision (i.e. widens CI), as is evident comparing the CI width between marginal structural and unweighted analyses.

Comparing findings from HA-MSM and unweighted models, the association between Hgb-Var and death appears potentiated: HR (95% CI) 1.93 (1.20 to 3.10) versus 1.34 (1.24 to 1.44), respectively. Intuitively, this makes sense considering that patients with greater comorbid disease burdens are expected to have more variable response to exogenous erythropoietin (i.e. greater Hgb-Var), and also to be at increased risk of death. Conversely, the protective effects of absolute hemoglobin level and temporal hemoglobin trend become attenuated as weighting increases. Again, this makes intuitive sense considering that healthier patients are expected to have greater hemoglobin levels and more positive hemoglobin trends, and be at decreased risk of death.

Reassurance as to the effectiveness of HA-MSM to appropriately control for TDC can be found by analogy. Observational studies have demonstrated a significant and potent association between absolute hemoglobin level and mortality, (1,14–17) a finding not borne out by randomized trials (18,19). In examining our data, the association between higher hemoglobin levels and mortality becomes attenuated upon use of marginal structural analysis, and no longer reaches statistical significance, consistent with the results of randomized trials.

Other examples in the literature where the related marginal structural models have been applied include analysis of the Multicenter AIDS Cohort study, in which marginal structural analysis confirmed that zidovudine treatment was associated with improved survival among HIV-positive men, whereas standard statistical methods suggested it was harmful (11). Similarly, in a post hoc analysis of the Physicians’ Health Study, marginal structural analysis demonstrated that aspirin use was more beneficial in preventing cardiovascular mortality than appreciated by traditional statistical techniques (20). In the renal literature, marginal structural models have been used to confirm a survival benefit associated with activated vitamin D treatment of chronic dialysis patients (21), to demonstrate the safety of intravenous iron administration, (22) and to examine the effects of carnitine supplementation on hospitalization in dialysis patients (23).

Several limitations of the study presented here should be noted. First, the success of HA-MSM to estimate causal associations is contingent upon the untestable assumption that no residual confounding is present; this is especially true of “endogenous” variables such as hemoglobin that are not under direct human control (8,11). Through inverse probability of exposure (IPE) weighting, we attempted to account for many candidate TDC, but the possibility exists that there were others for which we could not fully account. This limitation is particularly relevant here given the limited data available on comorbid disease, infection, or hospitalization.

Second, as with all observational studies, there was the opportunity for information and selection biases. The latter is particularly concerning given the survival requirement imposed by our metric of Hgb-Var.

Third, the cohort we studied is nearly a decade old, and there have been temporal changes in the management of anemia in dialysis patients since that time. Although we know of no reason why the fundamental relationship between Hgb-Var and absolute level of hemoglobin and death would be different now as compared with a decade ago, we cannot exclude this possibility.

Fourth, although the population studied was geographically and demographically diverse, all subjects were treated in one network of dialysis providers. Generalization of results to other populations or settings should be undertaken cautiously. In particular, although the results presented here suggest that reducing Hgb-Var might be beneficial, it is premature to conclude that specific approaches to doing this would lead to reductions in mortality.

Finally, it should be noted that no statistical technique is a valid substitute for well done randomized clinical trials when...
such trials are feasible. However, history-adjusted marginal structural modeling (as well as the related marginal structural modeling) is one potential analytical technique that can be applied to observational data when randomization is infeasible.

In conclusion, this study demonstrates that hemoglobin variability is associated with increased mortality among chronic hemodialysis patients using marginal structural analysis, which controls for time-dependent confounding in a fashion superior to traditional multivariable statistical analyses. Furthermore, this association was more pronounced than those demonstrated using methods that do not take time-dependent confounding into account. If confirmed in other observational studies that also permit generalization beyond the population studied here, hemoglobin variability may ultimately become an important therapeutic target. Before this can occur, clinical trials will need to demonstrate the superiority of anemia management strategies that reduce hemoglobin variability. Studies such as this provide further impetus to develop and implement such trials.

Appendix
Calculation of Inverse Probability of Exposure Weights

Inverse probability of exposure weights were estimated using an unordered polychotomous logistic regression model. The dependent variable was the observed hemoglobin exposure for each subject-month divided into quartiles: ≤9.7 g/dl, 9.7 to 10.6 g/dl, 10.6 to 11.3 g/dl, and >11.3 g/dl. We selected polychotomous logistic regression rather than dichotomous logistic regression or ordinal logistic regression because the former does not provide for sufficient resolution of hemoglobin level, whereas the latter makes stronger assumptions about the form of the association between predictors and hemoglobin level that may not apply. The predictor variables in this model were laboratory measures and medication exposure from the previous 3 mo, and hemoglobin concentration over each of the prior 6 mo.

We used stabilized weights in place of raw weights, because they generally produce more stable estimates (or “estimates with higher precision”) (8,11). Stabilized weights are a ratio: the denominator is precisely the probability of exposure outlined above; the numerator is the probability of exposure as predicted by the prior 6-mo hemoglobin concentration and laboratory and medication data from the 3 mo before the hemoglobin exposure window.

Censoring Weights
Discontinued or censored follow-up occurred in our study cohort at the time of death, at the end of the study, or when a subject was lost to follow-up. The latter censoring event may have been associated with the exposure or outcome (mortality) (i.e. informative censoring), thereby, creating opportunities for bias. This was minimized by estimating a stabilized censoring weight for each subject using models analogous to those for stabilized IPE weights, except where the outcome for each month was censoring (1 = yes, 0 = no) (8,11). The weights that we applied in the outcome analysis were the product of the stabilized IPE times censoring weight (8,11).

Acknowledgments
This study was supported by an unrestricted grant from Hoffman-La Roche Company. In addition, R.K.I. was supported by NIH grant F32 DK072877-01, and S.M.B. in part, by an NIH grant T32DK07785, and American Heart Association Fellow-to-Faculty transition award 0775021N. WY had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. This work was previously presented in abstract form at the American Society of Nephrology meeting in San Diego, CA, November 2006. The authors express great appreciation to Dr. Michael Lazarus and Fresenius Medical Care for providing unrestricted access to the data for these analyses as well as those used for our prior study in this area (5).

Disclosures
None.

References