*University of Pennsylvania*
UPenn Biostatistics Working Papers

*Year* 2007 *Paper* 19

# Methodological Issues in the Study of the Effects of Hemoglobin Variability

Marshall Joffe* Wei Yang†
Steve Brunelli‡ Harold Feldman**

*University of Pennsylvania, mjoffe@mail.med.upenn.edu
†
‡steven.brunelli@uphs.upenn.edu
**University of Pennsylvania, hfeldman@mail.med.upenn.edu

# Methodological Issues in the Study of the Effects of Hemoglobin Variability

Marshall Joffe, Wei Yang, Steve Brunelli, and Harold Feldman

**Abstract**

We consider estimating the effect of hemoglobin variability on mortality in hemodialysis patients. Causal effects can be defined as comparisons of outcomes under different hypothetical interventions. Defining measures of the effect of hemoglobin variability and clinical outcomes is complicated by the fact that hypothetical interventions on variability used to define its effect inevitably involve manipulation of related variables. We propose a model-based definition of the effect of the hemoglobin variability as a parameter for variability in a causal model for the effect of an overall intervention on hemoglobin levels over time. We consider this problem using history-adjusted marginal structural models, and apply this approach to data from a large observational database. We consider issues arising when the variable of interest is endogenous, and consider in principle alternate estimands.

# Methodological Issues in the Study of the Effect of Hemoglobin Variability

Marshall M. Joffe

Wei Yang

Steven Brunelli

Harold I. Feldman

Marshall M. Joffe is Associate Professor of Biostatistics, Department of Biostatistics and Epidemiology, University of Pennsylvania, Philadelphia, PA 19104-6021 (E-mail: mjoffe@mail.med.upenn.edu), Wei Yang is Staff Biostatistician, Center for Clinical Epidemiology and Biostatistics, University of Pennsylvania, Philadelphia, PA 19104-6021 (E-mail: weiyang@mail.med.upenn.edu), Steven Brunelli is Fellow, Division of Epidemiology, Center for Clinical Epidemiology and Renal Electrolyte and Hypertension Division of the Department of Medicine, University of Pennsylvania School of Medicine, Philadelphia, Pennsylvania 19104-6021 (E-mail: Steven.Brunelli@uphs.upenn.edu), and Harold I. Feldman is Professor of Medicine and Epidemiology, Departments of Medicine and of Biostatistics and Epidemiology, University of Pennsylvania School of Medicine, Philadelphia, PA 19104-6021 (E-mail: hfeldman@mail.med.upenn.edu).

**Abstract**

We consider estimating the effect of hemoglobin variability on mortality in hemodialysis patients. Causal effects can be defined as comparisons of outcomes under different hypothetical interventions. Defining measures of the effect of hemoglobin variability and clinical outcomes is complicated by the fact that hypothetical interventions on variability used to define its effect inevitably involve manipulation of related variables. We propose a model-based definition of the effect of the hemoglobin variability as a parameter for variability in a causal model for the effect of an overall intervention on hemoglobin levels over time. We consider this problem using history-adjusted marginal structural models, and apply this approach to data from a large observational database. We consider issues arising when the variable of interest is endogenous, and consider in principle alternate estimands.

# 1 Introduction

Individuals with end-stage renal disease dependent on chronic hemodialysis commonly manifest anemia. That anemia is typically treated with a combination of two therapies: erythropoetin (EPO) and intravenous iron. Despite these therapies, or, in part, because of the way that they are administered, hemoglobin (Hgb) levels often vary substantially over time. Further, the amount of this individual variability in Hgb varies substantially among individuals (Lacson, Ofshun and Lazarus 2007;Berns, Elzein, Hafez, Fishbane, Meisels and Deoreo 2003;Fishbane and Berns 2005). It has been hypothesized that increased Hgb variability is associated with and leads to increased mortality.

Understanding of the effect of Hgb variability requires an appropriate definition of the causal effects of that variability, as well as consideration of the assumptions under which those effects are estimable. Definition and estimation of the effects of Hgb variability is complicated by two factors: Hgb variability is not definable instantaneously, but only over time, and Hgb itself and its variability are subject to human control or manipulation only indirectly and incompletely. After considering how to define the effects of Hgb variability, we outline appropriate methods for estimating the effect of Hgb variability on mortality. We then apply these methods to data obtained from the Fresenius Medical Corporation, a large national chain provider of hemodialysis in the U.S.

# 2 Defining the Effect of Hemoglobin Variability

## 2.1 The Potential Outcomes Model

Under the potential outcomes model(Neyman 1990;Rubin 1974), causal effects are

defined as comparisons of what would happen to the same individual or group under two or more potential manipulations or conditions. Hgb variability is not meaningful instantaneously, but is meaningful as a description of how Hgb varies over time. Thus, in considering possible manipulations, it is most straightforward to consider directly manipulating a person's Hgb level over a period of time; Hgb variability would then be a function of the fluctuations in Hgb over time under this planned manipulation.

We formalize these ideas using the potential outcomes model. Let $A_k$ denote a subject's Hgb level at time *k*, and let $Y_k$ indicate whether the subject is alive at time *k* (1=dead). We use overlines to denote the history of a variable; thus, $\overline{A}_k \equiv \{A_1,...,A_k\}$ denotes the course of Hgb through time *k*. Let $Y_k^{\overline{a}_{k-1}}$ denote the outcome that would have been observed at time *k* had a subject's Hgb level through time *k*-1 been manipulated achieve Hgb history $\overline{a}_{k-1}$. Causal effects are comparisons of outcomes $Y_k^{\overline{a}_{k-1}}$ and $Y_k^{\overline{a}'_{k-1}}$ for different Hgb histories $\overline{a}_{k-1}$, $\overline{a}'_{k-1}$ for the same individual or group (Robins 1986;Robins 1987).

## 2.2 History-adjusted Marginal Structural Models

One way to parametrize these comparisons is through history-adjusted marginal structural models (HA-MSMs)(van der Laan, Petersen and Joffe 2005b). HA-MSMs can be used to model the effect of Hgb levels over a particular interval on the outcome following that interval. We use double subscripting to refer to the history of a process over a specified interval. Thus, $\overline{a}_{k,\Delta} \equiv \{a_k,...,a_{k+\Delta-1}\}$ refers to a potential exposure history starting at *k* and lasting $\Delta$

intervals. HA-MSMs model the expectation or distribution of $Y_{k+\Delta}^{\bar{A}_{k-1},\bar{a}_{k,\Delta}}$ given some function of the observed covariates measured through time $k$ $V_k \equiv r(\bar{L}_k, \bar{A}_{k-1})$ for some known function $r(\cdot)$; $V_k$ can be the null set, the entire history $\bar{L}_k, \bar{A}_k$, or some function thereof (e.g., the current value of covariates $L_k$). Thus, these are models for $E(Y_{k+\Delta}^{\bar{A}_{k-1},\bar{a}_{k,\Delta}} | V_k)$ or for $f(Y_{k+\Delta}^{\bar{A}_{k-1},\bar{a}_{k,\Delta}} | V_k)$.

For discrete-time failure outcomes, we consider models of the form

$$g\{E(Y_{k+\Delta}^{\bar{A}_{k-1},\bar{a}_{k,\Delta}} | V_k, Y_{k+\Delta-1}^{\bar{A}_{k-1},\bar{a}_{k,\Delta}} = 0)\} = q(V_k, \bar{a}_{k,\Delta}). \quad \textbf{(1)}$$

$g(\cdot)$ can be a standard link function (e.g., logit) or complementary log-log. $q(V_k, \bar{a}_{k,\Delta})$ is used to parametrize 1) the association of "baseline" covariates $V_k$ with the potential outcomes, 2) the effect of the exposure (Hgb) experienced from $k$ to $k+\Delta-1$, and 3) any modification of the exposure effect by "baseline" covariates. For simplicity, we suppose for now that there is no modification (on the chosen scale) of the effect of Hgb variability by covariates $V_k$; we may thus write $q(V_k, \bar{a}_{k,\Delta}) = q_1(V_k) + q_2(\bar{a}_{k,\Delta})$; $q_1$ represents the association of covariates $V_k$ with the outcome, and $q_2$ represents the effect of Hgb exposure over time.

### 2.3 Defining the Effect of Hemoglobin Variability

We may choose to represent the effect $q_2$ of Hgb exposure as being a function in part of Hgb variability under that exposure pattern. For any $\Delta \geq 3$, one might summarize an exposure

pattern $\bar{a}_{k,\Delta}$ through a linear regression: $a_j=\beta_{0k}+\beta_{1k}(j-k)+\epsilon_k$; here $\beta_0$ and $\beta_1$ are the least squares slopes and intercepts, and $\epsilon$ is the residual from this regression. The residual variability of the $a_j$ after adjusting for the regression line may be described as $\sigma_k \equiv (\frac{1}{\Delta-2}\sum_{j=k}^{k+\Delta-1}\epsilon^2)^{0.5}$; we call $\sigma$ the residual standard deviation of Hgb.

One way to model the effect of hemoglobin over the period $[k,k+\Delta)$ is as a function of the intercept, slope, and residual standard deviation; e.g., $q_2(\bar{a}_{k,\Delta})=\alpha_0+\beta_{0k}\alpha_1+\beta_{1k}\alpha_2+\sigma_k\alpha_3$. In this model, the effect of Hgb is expressed through separate regression parameters for the Hgb intercept, slope, and residual standard deviation. We call the regression parameter for the residual standard deviation the effect of Hgb variability, as it represents the effect of variability in Hgb apart from overall Hgb levels and trends.

### 2.4 Difficulties with Definition

Hgb is not modified directly. For many other interventions whose effects are of interest (e.g., taking a particular dose of a drug), the immediate goal of the intervention (e.g., the amount of drug ingested) may be achieved in a standardized fashion (e.g., ingestion of the desired amount); the exposure $A_k$ is simply set by some person (e.g., physician or patient). The standard approach to modify Hgb in hemodialysis patients is through a combination of synthetic erythropoetin (EPO) and intravenous (IV) iron. It may be impossible to specify in advance the dose of these agents required over time to achieve the Hgb goals, and so $A_k$, while strongly affected by human intervention, is not as completely under human control.

To speak formally about the effect of Hgb variability, or of Hgb in general, it is necessary to consider the hypothetical intervention that might be applied to get Hgb to reach a certain level (Robins 1986;Rubin 1986). Although current technology does not allow one to completely control Hgb levels, one can imagine interventions and technology that could lead to closer control. In addition, one will need to assume that the effect of Hgb does not depend on the method used to make Hgb reach those levels (e.g., different combinations of EPO and iron that might lead to identical Hgb levels), at least among the several methods used among subjects in the study and in the future to achieve given Hgb levels. This assumption does not imply that the methods used to change Hgb exert an effect on mortality only through their effect on Hgb levels; in other words, EPO or IV iron may affect mortality partly by modifying Hgb and partly through other mechanisms. These other pathways through which EPO or iron affect mortality do not prevent meaningful definition of the effect of Hgb on mortality.

## 3 Identification and Estimation

We use methods proposed for HA-MSMs for estimating the effect of Hgb, including the effect of variability. This approach uses inverse probability of treatment weighting to control for confounding, and inverse probability of censoring weighting to adjust for censoring. To estimate parameters in our HA-MSM $g\{E(Y_{k+\Delta}^{\bar{A}_{k-1},\bar{a}_{k,\Delta}}|V_k,Y_{k+\Delta-1}^{\bar{A}_{k-1},\bar{a}_{k,\Delta}}=0)\}$ **(1)**, we fit the corresponding observed data model $g\{E(Y_{k+\Delta}|V_k,\bar{A}_{k-1},\bar{A}_{k,\Delta},Y_{k+\Delta-1}=0)\}$ using weights. Let $\kappa_{i.k}\equiv pr(A_{i.k}=a'_{i.k}|\bar{L}_{i.k},\bar{A}_{i.k-1},Y_{i.k}=0)$ and $\tau_{i.k.j}\equiv pr(A_{i.k}=a'_{i.k}|V_{i.j},\bar{A}_{i.k-1},Y_{i.k}=0)$, where the statement $A_{i.k}=a'_{i.k}$ means that the (random

variable) Hgb at time $k$ equals its observed value $a'_{i.k}$. Let $D_k=1$ if a subject is not lost to follow-up between $k$ and $k+1$, 0 if the subject is lost to follow-up. Let $\theta_{i.k} \equiv pr(D_{i.k}=1|\bar{L}_{i.k},\bar{A}_{i.k},Y_{i.k}=0,D_{i.k-1}=0)$ and let $\rho_{i.k.j} \equiv pr(D_{i.k}=1|V_{i.j},\bar{A}_{i.k},Y_{i.k}=0,D_{i.k-1}=0)$. The weights used for each person-interval are $\prod_{k'=k}^{k+\Delta-1} (\tau_{i.k./k}/\kappa_{i.k'})(D_{i.k'}/\rho_{i.k./k}/\theta_{i.k'})$; the first term $(\prod (\tau_{i.k./k}/\kappa_{i.k'}))$ is the stabilized inverse probability of treatment weight, and the second term $(\prod (D_{i.k'}/\rho_{i.k./k}/\theta_{i.k'}))$ is the stabilized inverse probability of censoring weight. Fuller discussion and justification of this weighting scheme for use in MSMs and HA-MSMs appear elsewhere (Hernan, Brumback and Robins 2000;van der Laan et al. 2005b).

The treatment weights derive from models of Hgb levels, not for models of Hgb variability. This is formally justified because we are modeling the effect of a pattern of Hgb (of which variability is a part), not directly modeling the effect of variability. Additionally, because our measures of variability derive from monthly Hgb measures, variability is defined only over a longer time scale. Directly modeling variability would require modeling a composite variable defined from a number of individual measurements over several months. The longer time frame could interfere with the ability of inverse probability weighting to control confounding of the effect of Hgb by time-varying covariates, especially if Hgb level has immediate effects on other variables which quickly affect subsequent Hgb.

In modeling Hgb level, we divided Hgb levels into four categories: ≤9.7 g/dl, 9.7-10.6 g/dl, 10.6-11.3 g/dl, and >10.3 g/dl. We then used logistic regression models for unordered

categorical outcomes to model Hgb levels. For comparison, we considered models for ordered categorical outcomes; the fit of the proportional odds model, one such model, was substantially poorer. Further, since the models for Hgb are not the main target of our inference, the improved interpretability of the simpler ordinal logistic regression model is of relatively little concern. We considered several approaches to modeling current Hgb as a function of previous Hgb: in some, we modeled current Hgb directly as a function of Hgb levels in previous months; in others, we modeled current Hgb level as a function of previous Hgb levels summarized by the intercept, slope, and standard deviation of an individual's observed Hgb levels over the previous six months. Since the model for Hgb level is a nuisance model; thus, we are less concerned with interpretability of these models than we are for the HA-MSM for mortality. Similarly, we used logistic regression to model censoring as a discrete-time process. We derived weights as above from these models.

Table 1 presents selected parameters from a polychotomous regression model for Hgb levels; in particular, it presents parameters modeling the comparison of the lowest Hgb category to the highest Hgb category. Earlier Hgb levels are the variable which most strongly predicts current Hgb. Interestingly, having moderately high Hgb is associated with a lower conditional probability of a having a low Hgb the next month than having a high Hgb. Two of the strongest other predictors of Hgb were the treatments aiming at increasing Hgb: EPO dose and iron dose. In both cases, higher doses in the previous month were associated with lower probabilities of having low Hgb later (controlling for the other variables in the model). Interestingly, higher earlier EPO dose was associated with higher probabilities of having low current Hgb; this could reflect the fact that these are subjects with higher levels of EPO required to maintain a given Hgb

level. Higher albumin, a marker for good nutritional status, is associated with lower probability of having low Hgb, and higher ferritin is associated with higher probability of low Hgb.

Figure 1 presents modified box plots of the inverse probability of treatment weights, the censoring weights, and the overall weights, both stabilized and unstabilized. Not surprisingly, the treatment weights are much more variable than the censoring weights, and so these play a larger role in generating the overall weights and the instability of our estimates. As expected, the stabilized weights are less variable than the unstabilized one. Interestingly, the proportion of the total weight contributed by any single observation is larger for the stabilized weights (0.11% versus 0.01%).

We fit logistic regression HA-MSMs to the data, using the stabilized weights derived above. Weights which vary too much can interfere with inference (Elliott and Little 2000); these weights can reflect the fact that some levels of exposure are extremely rare for subjects with certain prior exposure and covariate levels. Highly variable weights can lead (properly) to large standard errors and imprecise inference. In the presence of large and highly variable weights, weight trimming is sometimes applied (Elliott and Little 2000). In this approach weights larger than (or smaller than) a specified level are trimmed to that level; for example, if the level used for trimming were 10, a subject whose weight was estimated to be 50 based on the inverse probability weighting methods would be reassigned a weight of 10.

It is not always clear what level should be used for trimming. In the survey sampling literature, one approach is to choose weights to minimize the model-based mean squared error (MSE). Let $\hat{\Psi}^t$ denote the estimate where weights greater than $t$ are trimmed. If one assumes that the untrimmed estimate $\hat{\Psi}^\infty$ is asymptotically unbiased, we can naively estimate the MSE of

a trimmed estimator by $(\hat{\Psi}^t - \hat{\Psi}^\infty)^2 + \hat{Var}(\Psi^t)$; we can then choose the trimming level $t$ to minimize the MSE.

Table 2 provides estimates of the effect of Hgb on mortality using different schemes for trimming. Higher initial Hgb (intercept) and greater increases in Hgb over time (slope) are both associated with less mortality. Weighting decreases the magnitude of these associations; increasing trimming increases the magnitude, as it produces analyses closer to the unweighted ones. Thus, at least some of the associations between lower Hgb and mortality appears to be due to time-varying confounding. Increasing residual variability (standard deviation) is associated with more mortality; unlike the other Hgb parameters, its magnitude appears to increase with weighting.

Table 3 presents the model-based MSE for each Hgb parameter for each trimming scheme. For residual standard deviation, the MSE is minimized when trimming weights at 100. For the intercept and slope parameters, the untrimmed weights produce the smallest MSE.

## 4 Alternative Estimands

The possible feedback mechanisms between Hgb and EPO or IV iron may pose some problems for defining the effect of treatment. Figure 2 shows possible relations among Hgb levels, EPO dose, and mortality. EPO dose affects Hgb, and Hgb levels also influence subsequent dosing of EPO. Finally, we draw arrows from EPO to mortality to indicate that EPO may have a direct effect on mortality separate from any effect of EPO on mortality mediated by Hgb.

In characterizing the effect of Hgb in this setting, we can consider both the overall effect of Hgb on mortality and the direct effect, controlling for EPO use. The direct effect of Hgb on mortality is the effect of Hgb on mortality were we to ensure that EPO dosing is not affected by Hgb (or alternatively, were the direct effect of EPO somehow to be blocked); we imagine intervening to change EPO dosing patterns from their current dependence on Hgb. The overall effect of Hgb is the effect of Hgb not controlling for EPO; that is, the effect of Hgb were EPO dosing to follow the patterns it takes in our data. No intervention on EPO is contemplated in defining overall effects. Figures 3 and 4 show two manipulated graphs; the first represents the direct effect of Hgb (controlling for EPO as a possible intermediate and also as a confounder); the second represents the overall effect of Hgb, controlling for EPO only as a potential time-varying confounder; the second represents the overall effect of Hgb, controlling for EPO only as a potential time-varying confounder.

If we are interested in studying the "pure" or "biologic" effect of Hgb, we should be more interested in the direct effect. The direct effect would represent roughly the effect of an intervention which alters Hgb but does not affect mortality through other pathways. The HA-MSM approach that we have adopted instead takes the estimand to be the overall effect of Hgb. Methods to estimate the direct effect are more involved; it is not clear how estimates of Hgb effect would differ from the overall effect estimates reported here.

## 5 Discussion

There are widely divergent views on the propriety of discussing the causal effect of variables for which direct intervention is not possible. Some have adopted the slogan or motto,

"no causation without manipulation," (Rubin 1986); this slogan can be taken to mean that, if the variable whose effect is being considered is not subject to manipulation in the current setting, it is meaningless to consider its causal effect. In this view, not only is it meaningless to consider the effects of such nonmanipulable variables as sex or race, but it is also meaningless to consider the effect of medication usage (as opposed to treatment assignment) in a randomized trial if there is noncompliance to the assigned treatment (Imbens and Rubin 1997). A complementary view is that, if there are multiple versions of the intervention (i.e., different ways to achieve different Hgb patterns), and those versions lead to different outcomes (e.g., different EPO and iron dosage patterns that lead to identical Hgb patterns have different consequences), it is not meaningful to speak of the causal effect of the intervention (Rubin 1986); in this view, methods developed primarily to answer causal questions may not be of interest in these settings.

In our opinion, these attitudes are too strong, as they forbid the consideration of important causal questions. In common usage, people speak of the causal effect of medication use or of Hgb levels; it is a useful exercise to attempt to make this common understanding more precise. Consideration of the effects of variables that are not subject to direct manipulation can serve to advance understanding of the mechanisms that lead to clinical outcomes of interest (e.g., mortality) and can lead to prediction of the effects of interventions to change this variable.

A different view is that methods developed for answering causal questions, such as the HA-MSM estimand (estimated using the inverse probability weighted estimator), are of interest even when it is difficult to imagine direct intervention on the variable of interest (e.g., Hgb) given current technology (van der Laan, Haight and Tager 2005a). While we sympathize with this viewpoint in general, it is unsatisfying in the current setting, because it does not fully

address the questions motivating our analysis (see Hernan (2005) for discussion in an analogous setting); part of the motivation for considering the effect of Hgb variability is to consider what would happen were we to use a hypothetical intervention (e.g., a new pharmacologic agent or new reimbursement policies for EPO) to modify Hgb levels and variability. The HA-MSM estimand may estimate the effect of Hgb patterns under the current mixture of Hgb management protocols leading to those patterns (van der Laan et al. 2005a).

In this setting, we would hope that Hgb levels could serve as a surrogate for the intervention that sets Hbg levels over time at particular values and so fixes Hgb variability. These ideas might be formalized in terms of the joint effects of an actual intervention applied for Hgb management and of the Hgb levels themselves (van der Laan et al. 2005a;Pearl 2001;Robins and Greenland 1992;Taylor, Wang and Thiebaut 2005;Freedman, Graubard and Schatzkin 1992); in this view, Hgb level (including its variability) would be a good surrogate for the actual intervention if the direct effect of method of managing those levels (i.e., that part of its effect that is not mediated by its effect on Hgb level) were small. Methods for evaluating the usefulness of Hgb as a surrogate are beyond the scope of this paper.

In conclusion, we believe that it is worthwhile to try to define and estimate the effect of Hgb variability, and we have attempted to do so. Nonetheless, we are cognizant of the difficulties involved, and so caution in interpretation is warranted. Nonetheless, our analysis and interpretation suggest that increased Hgb variability may lead to adverse outcomes. Stronger evidence might be obtained by a trial with arms with tighter and less tight control of Hgb levels.

Table 1. Logistic regression model for monthly hemoglobin.

| Variable | Coefficient | Standard error | chi-squared | d.f. | p-value |
|---|---|---|---|---|---|
| MONTH | | | 2507 | 75 | <.0001 |
| Hb previous month | 3.07 | 0.02 | 81664 | 9 | <.0001 |
| | 0.74 | 0.01 | | | |
| | -1.11 | 0.01 | | | |
| HB two months ago | 0.36 | 0.01 | 3803 | 9 | <.0001 |
| | 0.22 | 0.01 | | | |
| | -0.06 | 0.01 | | | |
| HB three months ago | -0.08 | 0.01 | 551 | 9 | <.0001 |
| | -0.02 | 0.01 | | | |
| | 0.04 | 0.01 | | | |
| HB four months ago | 0.19 | 0.01 | 959 | 9 | <.0001 |
| | 0.10 | 0.01 | | | |
| | -0.05 | 0.01 | | | |
| HB five months ago | 0.38 | 0.01 | 2459 | 9 | <.0001 |
| | 0.16 | 0.01 | | | |
| | -0.09 | 0.01 | | | |
| HB six months ago | 0.29 | 0.01 | 1589 | 9 | <.0001 |
| | 0.12 | 0.01 | | | |
| | -0.06 | 0.01 | | | |
| CO2 previous month | 0.0167 | 0.0022 | 65 | 3 | <.0001 |
| CO2 2 months ago | -0.0120 | 0.0023 | 38 | 3 | <.0001 |
| CO2 3 months ago | -0.0202 | 0.0022 | 82 | 3 | <.0001 |
| SQRT_FERRITIN previous month | 0.0655 | 0.0018 | 1381 | 3 | <.0001 |
| SQRT_FERRITIN 2 months ago | -0.0261 | 0.0021 | 177 | 3 | <.0001 |
| SQRT_FERRITIN 3 months ago | -0.0267 | 0.0017 | 237 | 3 | <.0001 |
| IRON previous month | -0.0033 | 0.0004 | 62 | 3 | <.0001 |
| IRON 2 months ago | -0.0034 | 0.0004 | 64 | 3 | <.0001 |
| IRON 3 months ago | 0.0004 | 0.0004 | 1 | 3 | 0.8314 |
| TSAT previous month | 0.0080 | 0.0010 | 71 | 3 | <.0001 |
| TSAT 2 months ago | 0.0046 | 0.0010 | 24 | 3 | <.0001 |
| TSAT 3 months ago | -0.0018 | 0.0010 | 5 | 3 | 0.1499 |

| | | | | | |
|---|---|---|---|---|---|
| URR previous month | -0.0056 | 0.0014 | 27 | 3 | <.0001 |
| URR 2 months ago | -0.0003 | 0.0015 | 6 | 3 | 0.1032 |
| URR 3 months ago | -0.0008 | 0.0015 | 2 | 3 | 0.6458 |
| KTV previous month | -0.2011 | 0.0480 | 18 | 3 | 0.0005 |
| KTV 2 months ago | 0.0132 | 0.0498 | 7 | 3 | 0.0654 |
| KTV 3 months ago | -0.0091 | 0.0485 | 4 | 3 | 0.2501 |
| CALCIUM previous month | 0.0714 | 0.0105 | 74 | 3 | <.0001 |
| CALCIUM 2 months ago | 0.0121 | 0.0115 | 4 | 3 | 0.2847 |
| CALCIUM 3 months ago | -0.0285 | 0.0105 | 13 | 3 | 0.0057 |
| PTH previous month | 0.0003 | 0.0000 | 73 | 3 | <.0001 |
| PTH 2 months ago | -0.0001 | 0.0001 | 4 | 3 | 0.2256 |
| PTH 3 months ago | -0.0001 | 0.0000 | 13 | 3 | 0.0043 |
| PHOSPH previous month | -0.0442 | 0.0047 | 93 | 3 | <.0001 |
| PHOSPH 2 months ago | 0.0347 | 0.0050 | 54 | 3 | <.0001 |
| PHOSPH 3 months ago | 0.0443 | 0.0047 | 96 | 3 | <.0001 |
| AST_GOT previous month | -0.0031 | 0.0003 | 117 | 3 | <.0001 |
| AST_GOT 2 months ago | -0.0007 | 0.0003 | 26 | 3 | <.0001 |
| AST_GOT 3 months ago | 0.0002 | 0.0003 | 9 | 3 | 0.0234 |
| ALBUMIN previous month | -0.2795 | 0.0247 | 232 | 3 | <.0001 |
| ALBUMIN 2 months ago | 0.0082 | 0.0262 | 6 | 3 | 0.1348 |
| ALBUMIN 3 months ago | 0.0633 | 0.0248 | 18 | 3 | 0.0005 |
| SQRT_EPO previous month | -0.0141 | 0.0001 | 17543 | 3 | <.0001 |
| SQRT_EPO 2 months ago | 0.0069 | 0.0001 | 3400 | 3 | <.0001 |
| SQRT_EPO 3 months ago | 0.0072 | 0.0001 | 4626 | 3 | <.0001 |
| SQRT_IRONDOSE previous month | -0.0365 | 0.0007 | 3341 | 3 | <.0001 |
| SQRT_IRONDOSE 2 months ago | -0.0155 | 0.0008 | 453 | 3 | <.0001 |
| SQRT_IRONDOSE 3 months ago | -0.0015 | 0.0007 | 8 | 3 | 0.0369 |

Description of selected variables in the model C02: bicarbonate concentration (in blood); SQRT_FERRITIN: the square root of the ferritin concentration; IRON: iron concentration; TSAT: transferring saturation; URR: urea reduction ratio; CALCIUM: calcium concentration; PTH: parathyroid hormone concentration; PHOSPH: phosphate concentration; ALBUMIN: albumin concentration; SQRT_EPO; square root of the erythropoetin dose; SQRT_IRONDOSE: square root of the intravenous iron dose.

Table 2: MSM/6 months exposure window/weight calculated using previous 6 months' hgb value

| | MSM | | | | | MSM(W≤100) | | | | | MSM(W≤10) | | | | | Unweighted Model | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Parameter | beta | se | 95% CI | | p | beta | se | 95% CI | | p | beta | se | 95% CI | | p | beta | se | 95% CI | | p |
| Intercept | -2.287 | 3.292 | -8.740 | 4.165 | 0.487 | -0.095 | 2.188 | -4.383 | 4.193 | 0.965 | 1.635 | 1.489 | -1.283 | 4.552 | 0.272 | 1.813 | 0.799 | 0.246 | 3.379 | 0.023 |
| Hgb Intercept | -0.168 | 0.142 | -0.447 | 0.111 | 0.237 | -0.283 | 0.091 | -0.462 | -0.104 | 0.002 | -0.367 | 0.036 | -0.437 | -0.297 | <.0001 | -0.343 | 0.020 | -0.382 | -0.304 | <.0001 |
| Hgb Slope | -0.515 | 0.452 | -1.402 | 0.372 | 0.255 | -0.857 | 0.308 | -1.461 | -0.252 | 0.006 | -1.162 | 0.123 | -1.403 | -0.922 | <.0001 | -1.278 | 0.069 | -1.414 | -1.143 | <.0001 |
| Hgb Residual SD | 0.656 | 0.242 | 0.181 | 1.131 | 0.007 | 0.514 | 0.182 | 0.157 | 0.872 | 0.005 | 0.410 | 0.084 | 0.246 | 0.574 | <.0001 | 0.292 | 0.039 | 0.217 | 0.368 | <.0001 |
| Age (per year) | 0.043 | 0.006 | 0.031 | 0.055 | <.0001 | 0.039 | 0.004 | 0.030 | 0.047 | <.0001 | 0.036 | 0.002 | 0.032 | 0.041 | <.0001 | 0.033 | 0.001 | 0.031 | 0.036 | <.0001 |
| Sqrt Duration ESRD | 0.123 | 0.062 | 0.002 | 0.244 | 0.046 | 0.161 | 0.049 | 0.065 | 0.257 | 0.001 | 0.210 | 0.033 | 0.145 | 0.274 | <.0001 | 0.181 | 0.018 | 0.146 | 0.216 | <.0001 |
| Sex (Female) | -0.212 | 0.102 | -0.412 | -0.011 | 0.038 | -0.142 | 0.084 | -0.308 | 0.023 | 0.092 | -0.145 | 0.056 | -0.254 | -0.036 | 0.009 | -0.177 | 0.029 | -0.235 | -0.120 | <.0001 |
| Race (Asian/Pacific islander) | -0.464 | 0.241 | -0.936 | 0.008 | 0.054 | -0.424 | 0.229 | -0.873 | 0.025 | 0.064 | -0.301 | 0.199 | -0.692 | 0.089 | 0.130 | -0.114 | 0.122 | -0.352 | 0.124 | 0.347 |
| Race (Black) | -0.320 | 0.092 | -0.502 | -0.139 | 0.000 | -0.271 | 0.069 | -0.406 | -0.136 | <.0001 | -0.286 | 0.054 | -0.392 | -0.179 | <.0001 | -0.318 | 0.030 | -0.377 | -0.259 | <.0001 |
| Race (Native American) | -0.000 | 0.244 | -0.478 | 0.478 | 1.000 | 0.009 | 0.244 | -0.469 | 0.487 | 0.970 | 0.089 | 0.212 | -0.326 | 0.503 | 0.676 | 0.054 | 0.134 | -0.207 | 0.316 | 0.683 |
| Race (Other) | 0.064 | 0.239 | -0.404 | 0.532 | 0.788 | 0.101 | 0.230 | -0.350 | 0.551 | 0.661 | -0.051 | 0.135 | -0.316 | 0.215 | 0.709 | -0.125 | 0.072 | -0.266 | 0.016 | 0.082 |
| Adult Onset Diabetes | 0.126 | 0.124 | -0.117 | 0.368 | 0.310 | 0.232 | 0.086 | 0.064 | 0.400 | 0.007 | 0.295 | 0.056 | 0.186 | 0.404 | <.0001 | 0.262 | 0.030 | 0.204 | 0.320 | <.0001 |
| Juvenille Diabetes | 0.380 | 0.128 | 0.130 | 0.631 | 0.003 | 0.413 | 0.124 | 0.170 | 0.655 | 0.001 | 0.479 | 0.110 | 0.262 | 0.696 | <.0001 | 0.628 | 0.066 | 0.500 | 0.756 | <.0001 |
| Epoetin dose (per 10,000) | -0.006 | 0.009 | -0.024 | 0.011 | 0.494 | -0.008 | 0.008 | -0.024 | 0.008 | 0.358 | -0.001 | 0.007 | -0.014 | 0.012 | 0.859 | 0.000 | 0.004 | -0.008 | 0.008 | 0.994 |
| URR (per 10) | 0.501 | 0.759 | -0.986 | 1.988 | 0.509 | 0.038 | 0.554 | -1.048 | 1.123 | 0.946 | -0.450 | 0.423 | -1.279 | 0.380 | 0.288 | -0.428 | 0.225 | -0.869 | 0.013 | 0.057 |
| URR Square (per 100) | -0.064 | 0.063 | -0.188 | 0.060 | 0.315 | -0.023 | 0.045 | -0.110 | 0.065 | 0.610 | 0.021 | 0.034 | -0.045 | 0.088 | 0.528 | 0.027 | 0.018 | -0.008 | 0.062 | 0.133 |
| Serum albumin | -1.045 | 0.144 | -1.328 | -0.762 | <.0001 | -1.123 | 0.100 | -1.318 | -0.928 | <.0001 | -1.137 | 0.079 | -1.292 | -0.982 | <.0001 | -1.151 | 0.044 | -1.238 | -1.065 | <.0001 |
| AST/SGOT (per 25) | -0.023 | 0.028 | -0.077 | 0.032 | 0.413 | -0.012 | 0.025 | -0.060 | 0.037 | 0.640 | 0.026 | 0.023 | -0.020 | 0.071 | 0.272 | 0.103 | 0.029 | 0.047 | 0.159 | 0.000 |
| Bicarbonate | -0.051 | 0.025 | -0.100 | -0.003 | 0.038 | -0.030 | 0.015 | -0.060 | -0.001 | 0.042 | -0.019 | 0.009 | -0.037 | -0.000 | 0.045 | -0.017 | 0.005 | -0.027 | -0.006 | 0.001 |
| Calcium | 0.090 | 0.061 | -0.030 | 0.210 | 0.144 | 0.059 | 0.048 | -0.036 | 0.154 | 0.224 | 0.040 | 0.036 | -0.031 | 0.110 | 0.269 | 0.090 | 0.019 | 0.052 | 0.127 | <.0001 |

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Ferritin (per 100)** | 0.011 | 0.006 | -0.001 | 0.024 | 0.068 | 0.011 | 0.005 | 0.001 | 0.022 | 0.030 | 0.013 | 0.004 | 0.004 | 0.021 | 0.003 | 0.017 | 0.003 | 0.012 | 0.022 | <.0001 |
| **Phosphate** | 0.026 | 0.042 | -0.056 | 0.108 | 0.530 | 0.057 | 0.027 | 0.004 | 0.110 | 0.034 | 0.069 | 0.019 | 0.032 | 0.105 | 0.000 | 0.066 | 0.010 | 0.046 | 0.086 | <.0001 |
| **TSAT** | -0.012 | 0.006 | -0.024 | -0.001 | 0.030 | -0.014 | 0.005 | -0.024 | -0.003 | 0.012 | -0.011 | 0.004 | -0.020 | -0.002 | 0.015 | -0.010 | 0.003 | -0.015 | -0.005 | <.0001 |
| **Hgb** | -0.026 | 0.100 | -0.223 | 0.171 | 0.795 | 0.047 | 0.070 | -0.091 | 0.185 | 0.506 | 0.100 | 0.033 | 0.035 | 0.164 | 0.002 | 0.074 | 0.018 | 0.038 | 0.109 | <.0001 |
| **Iron (per 50)** | -0.013 | 0.187 | -0.379 | 0.354 | 0.946 | 0.097 | 0.138 | -0.173 | 0.367 | 0.480 | 0.104 | 0.109 | -0.109 | 0.318 | 0.338 | 0.042 | 0.062 | -0.079 | 0.163 | 0.500 |
| **KTV** | 0.738 | 0.461 | -0.166 | 1.641 | 0.110 | 0.522 | 0.388 | -0.238 | 1.283 | 0.178 | 0.259 | 0.334 | -0.395 | 0.914 | 0.437 | -0.153 | 0.170 | -0.486 | 0.179 | 0.367 |
| **PTH (per 20)** | 0.001 | 0.001 | -0.002 | 0.004 | 0.560 | 0.000 | 0.001 | -0.002 | 0.003 | 0.812 | 0.001 | 0.001 | -0.002 | 0.003 | 0.615 | 0.002 | 0.001 | 0.000 | 0.003 | 0.020 |
| **Irondose** | 0.304 | 0.124 | 0.061 | 0.547 | 0.014 | 0.201 | 0.081 | 0.043 | 0.360 | 0.013 | 0.143 | 0.041 | 0.063 | 0.223 | 0.000 | 0.092 | 0.021 | 0.051 | 0.134 | <.0001 |

Table 3. Estimated mean squared error for different weighted estimators

| Parameter | MSM | MSM ($W \leq 100$) | MSM ($W \leq 10$) | Unweighted model |
|---|---|---|---|---|
| Intercept | 0.14 | 0.15 | 0.20 | 0.18 |
| Slope | 0.45 | 0.46 | 0.66 | 0.77 |
| Residual standard deviation | 0.24 | 0.23 | 0.26 | 0.37 |

Figure 1

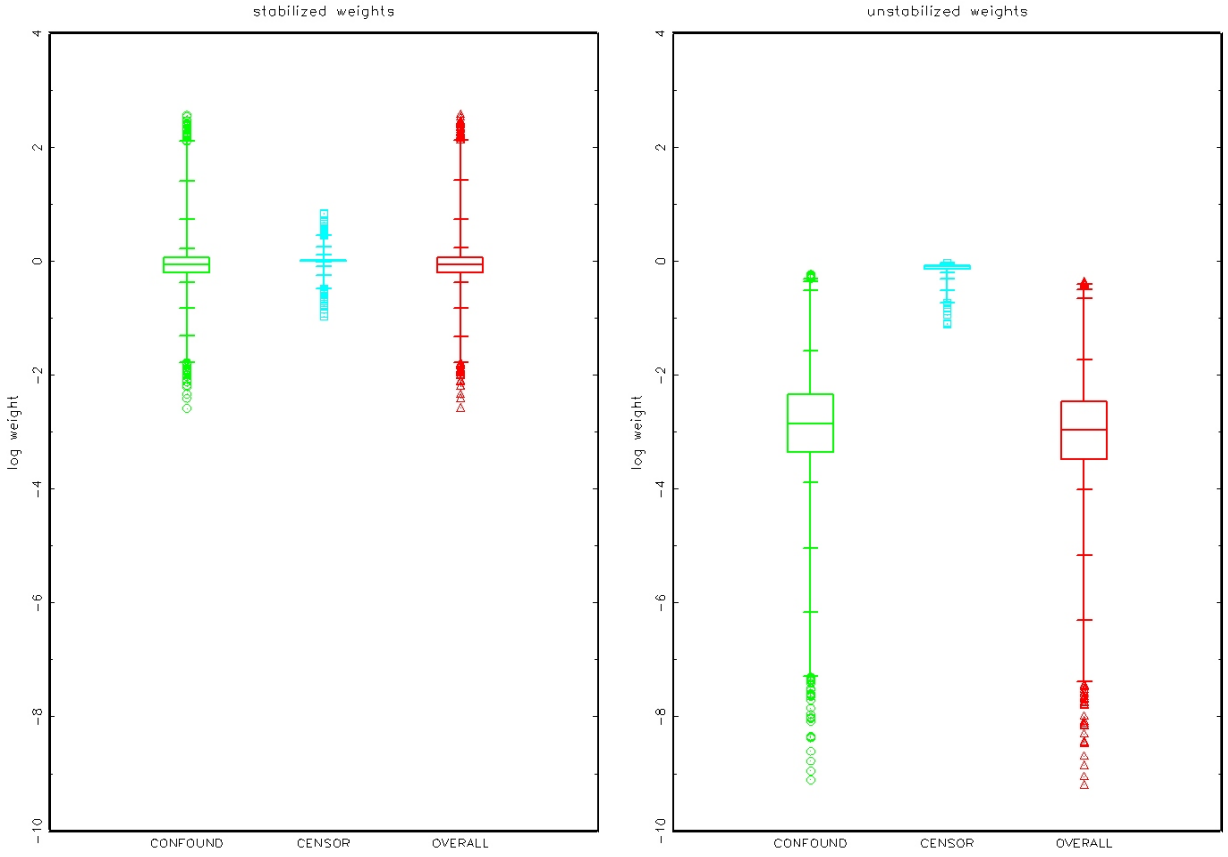


Figure 2. Directed Acyclic Graph showing relationship between variables in study

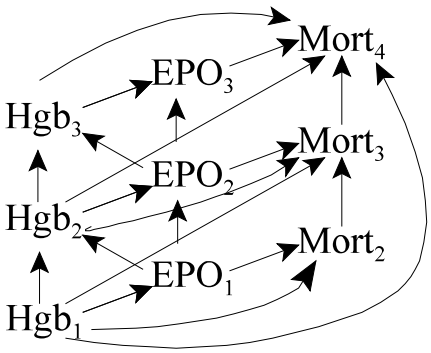



Figure 3. Manipulated graph diagramming overall effect of hemoglobin

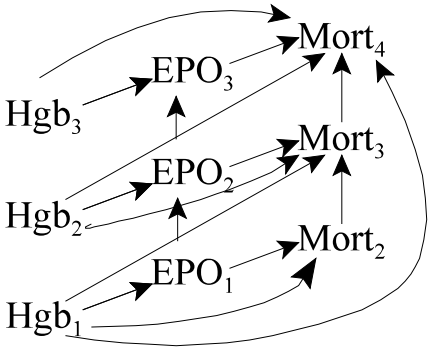
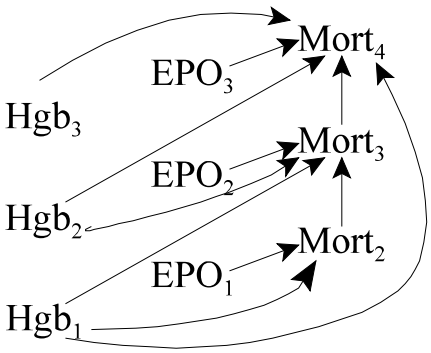


Figure 4. Manipulated graph diagramming direct effect of hemoglobin



Reference List

Berns, J. S., Elzein, H., Hafez, L., Fishbane, S., Meisels, I. S., and Deoreo, P. B. (2003), "Hemoglobin variability in epoetin-treated hemodialysis patients," *Kidney International*, 64, 1514-1521.

Elliott, M. R. and Little, R. J. A. (2000), "Model-based alternatives to trimming survey weights," *Journal of Official Statistics*, 16, 191-209.

Fishbane, S. and Berns, J. S. (2005), "Hemoglobin cycling in hemodialysis patients treated with recombinant human erythropoetin," *Kidney International*, 68, 1337-1343.

Freedman, L., Graubard, B., and Schatzkin, A. (1992), "Statistical validation of intermediate endpoints for chronic disease," *Statistics in Medicine*, 11, 167-178.

Hernan, M. A. (2005), "Invited commentary: hypothetical interventions to define causal effects-afterthought or prerequisite?," *American Journal of Epidemiology*, 162, 618-620.

Hernan, M. A., Brumback, B., and Robins, J. M. (2000), "Marginal structural models to estimate the causal effect of zidovudine on the survival of HIV-positive men," *Epidemiology*, 11, 561-570.

Imbens, G. W. and Rubin, D. B. (1997), "Bayesian inference for causal effects in randomized

experiments with noncompliance," *Annals of Statistics*, 25, 305-327.

Lacson, E. Jr., Ofshun, N., and Lazarus, J. M. (2007), "Effect of variability in anemia management on hemoglobin outcomes in ESRD," *American Journal of Kidney Diseases*, 41, 111-124.

Neyman, J. (1990), "On the application of probability theory to agricultural experiments. Essay on principles. Translated by D.M. Dabrowska and edited by T. P. Speed," *Statistical Science*, 5, 465-472.

Pearl, J. (2001), "Direct and indirect effects," in *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence*, San Francisco: Morgan Kaufmann.

Robins, J. (1986), "A new approach to causal inference in mortality studies with a sustained exposure period- application to control of the healthy worker survivor effect," *Mathematical Modelling*, 7, 1393-1512.

-----(1987), "A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods," *Journal of Chronic Diseases*, 40, 139S-161S.

Robins, J. and Greenland, S. (1992), "Identifiability and exchangeability for direct and indirect effects," *Epidemiology*, 3, 143-155.

Rubin, D. B. (1974), "Estimating causal effects of treatments in randomized and nonrandomized studies," *Journal of Educational Psychology*, 66, 688-701.

-----(1986), "Comment: which ifs have causal answers," *Journal of the American Statistical Association*, 81, 961-962.

Taylor, J. M. G., Wang, Y., and Thiebaut, R. (2005), "Counterfactual links to the proportion of treatment effect explained by a surrogate marker," *Biometrics*, 61, 1102-1111.

van der Laan, M. J., Haight, T. J., and Tager, I. B. (2005a), "van der Laan et al. respond to "Hypothetical interventions to define causal effects"," *American Journal of Epidemiology*, 162, 621-622.

van der Laan, M. J., Petersen, M. L., and Joffe, M. M. (2005b), "History-adjusted marginal structural models and statically-optimal dynamic treatment regimens," *International Journal of Biostatistics*, 1.