Original Article

Estimating Treatment Effect of High Hemoglobin Using the Principal Stratification Approach

Junji Moriya^{*1} and Yutaka Matsuyama^{*2}

^{*1}Biometrics, Kyowa Hakko Kirin Pharma, Inc.

^{*2}Department of Biostatistics, School of Public Health,
the University of Tokyo

e-mail: jmoriya@kyowa-kirin-pharma.com

Anemia is a common complication of chronic kidney disease (CKD) and end-stage renal disease. A high hemoglobin level targeted in the treatment of anemia has been controversial because recent overseas studies have reported that it did not affect renal survival or increased the risk of cardiovascular events. In the motivation study, patients with CKD were randomly assigned to high or low hemoglobin target group (11.0–13.0 or 9.0–11.0 g/dL). The comparison of groups for the composite of renal events as the primary endpoint revealed no significant differences ($p = 0.111$). In these studies, *ad hoc* analyses suggested that a high hemoglobin level may potentially reduce cardiovascular events. However, those results could not precisely estimate the effect of treatment with high hemoglobin because of post-treatment selection bias. To address this problem, we used the method based on principal stratification approach to estimate the causal effect of Partial Responders by which the treatment effect of high hemoglobin can be evaluated. The results suggested that not only Partial but also Always Responders may benefit more from high hemoglobin treatment than Never Responders. These data suggest that patients with CKD can receive benefit from high hemoglobin treatment, who can respond to that treatment.

Key words: Principal Stratification, Anemia, CKD, Erythropoiesis-Stimulating Agent.

1. Introduction

Chronic kidney disease (CKD) is a disease in which a patient's kidney function is chronically deteriorating. If left untreated, CKD can progress to end-stage renal disease (ESRD) and the patient will require dialysis or renal transplantation. CKD is strongly correlated with hypertension, diabetes mellitus, and metabolic syndrome. There are reportedly more than 10 million patients with CKD and more than 30,000 patients on dialysis in Japan [Japanese Society of Nephrology.(2012)]. The number of those patients will continue to increase in the future, so CKD has been recognized as an important disease. Anemia is a common complication of CKD

and ESRD. Patients with anemia have worse symptoms such as fatigue, palpitations, or vertigo due to a reduced red blood cell count. Also, since anemia causes a lack of oxygen throughout the body, the heart works harder to carry oxygen throughout the body, which creates a chronic burden. Therefore, improving the symptoms of anemia in patients with CKD and ESRD is very important. An erythropoiesis-stimulating agent that has been used for more than 20 years in Japan has played an important role in the treatment of anemia. Additionally, the Japanese Society for Dialysis Therapy recommended a target hemoglobin range of 10.0–11.0 g/dL in patients on hemodialysis and 11.0–13.0 g/dL in patients with CKD [Tsubakihara, Y., et al. (2010)].

Some large-scale overseas studies recently reported negative results that the high target hemoglobin level may be related to the increased risk of mortality or cardiovascular events; therefore, the target hemoglobin level in patients with CKD has been set to approximately 12.0 g/dL [Besarab, A., et al. (1998); Drüeke T.B., et al. (2006); Pfeffer, M.A., et al. (2009); Singh A.K., et al. (2006); Tsubakihara, Y., et al. (2010)]. In the above studies, exploratory analyses were performed that consider hemoglobin levels after the start of treatment because some patients did not achieve or maintain the target hemoglobin range. Those results suggested that the high target hemoglobin level may decrease the risk of renal-related events such as mortality or the need for dialysis [Inrig, J.K., et al. (2012); Solomon, S.D., et al. (2010); Szczech, L.A., et al. (2008); Szczech, L.A., et al. (2010)] but could not accurately estimate the effect of high hemoglobin levels because they used hemoglobin levels after the start of treatment.

As described in Section 2 below, this motivation study was performed in Japan to test the treatment effect of a high hemoglobin target on renal survival [Tsubakihara, Y., et al. (2012)]. In this study, different drugs were administered to patients with renal anemia between randomized groups. Patients assigned to the high hemoglobin group received darbepoetin- α , while those assigned to the low hemoglobin group received epoetin- α . Many studies of erythropoiesis-stimulating agents have been performed to date, but there has been no report on whether erythropoiesis-stimulating agents obviously benefit patients with renal anemia, in a manner not related to improved hemoglobin. Therefore, we believe that the treatment effect between the randomized groups in our motivation study is derived by the difference in target hemoglobin level ranges rather than the erythropoiesis-stimulating agent administration. When the treatment effects between different hemoglobin level ranges are estimated, hemoglobin values during the study should be analyzed as a covariate or time-dependent covariate, but that is not appropriate in the motivation study because hemoglobin levels after the start of treatment are affected by erythropoiesis-stimulating agent administration. If hemoglobin values observed after the start of treatment are used for the usual analysis methods such as subgroup analysis or a Cox regression model, the estimated effect may be biased. Therefore, we believe that this can be recognized as a post-treatment variable problem. Some methods based on causal inference have been proposed by many researchers to address post-treatment variable problem. However,

in the case that inverse-probability-of-treatment weighting (IPTW) [Robins, J.M., et al. (2000); Hernán, M.A., et al. (2000)] or g-estimation methods [Mark, S.D., Robins, J.M. (1993)] are applied for the motivation study in which different target hemoglobin level ranges are set between randomized groups as described in Section 2, we must suppose the counterfactual variable corresponding to unrealistic situation such as hemoglobin levels of all patients could be controlled within a particular range. From that, we believe that the principal stratification [Frangakis, C.E., Rubin, D.B. (2002)] can be applied to the motivation study than IPTW or g-estimation methods more appropriately. Therefore, we think about the method based on the principal stratification approach in this paper. The brief explanation of the principal stratification is shown next.

Estimates made from stratified analyses and adjustments for post-treatment variables can be biased [Rosenbaum, P.R. (1984)]. To address this problem, Frangakis and Rubin proposed the principal stratification approach to post-treatment variables [Frangakis, C.E., Rubin, D.B. (2002)]. The principal stratification approach to estimate the treatment effect is a cross-classification of patients defined by the joint potential values of the post-treatment variable under each of the treatments under comparison. In the clinical trials in which the patients are randomly assigned to a low or high hemoglobin treatment group, patients can be classified into the following four potential subgroups according to treatment group and hemoglobin response:

- Always Responder (AR): Those who would respond in either treatment group.
- Partial Responder (PR): Those who would not respond in the low hemoglobin treatment group but would respond in the high hemoglobin treatment group.
- Defier Responder (DR): Those who would respond in the low hemoglobin treatment group but would not respond in the high hemoglobin treatment group.
- Never Responder (NR): Those who would not respond in either treatment group.

Frangakis and Rubin defined the principal causal effect as a comparison of potential outcomes of primary interest within a principal stratum. In particular, Frangakis and Rubin defined associative effect (AE) as the principal causal effect of a population consisting of patients who have different potential values depending on treatment groups, that is, patients in the PR and DR groups, and dissociative effect (DE) as the principal causal effect for a population consisting of patients who have the same values among treatment groups, that is, patients in the AR and NR groups. However, as described in Section 3, we do not believe the existence of the DR group in this study because the monotonicity assumption can be made from the clinical point of view.

There is the realistic problem that many patients have difficulty maintaining or cannot achieve a high hemoglobin range if they undergo treatment with erythropoiesis-stimulating agents. As the IPTW or g-estimation methods, which require unrealistic suppositions such as all patients could be controlled within the high hemoglobin range, lead to unrealistic results, we decided to apply the principal stratification approach, which can give a more realistic result.

Furthermore, we thought that the treatment effect of high hemoglobin could be evaluated by estimating the principal causal effect for the PR group who could achieve the high hemoglobin range only within the high hemoglobin treatment group. We believe that we can show the benefit of high hemoglobin treatment if AE estimated by the principal stratification approach exceeds the result from the ITT analysis in the motivation study. If the estimated principal causal effect for the PR group exceeds the results from the ITT analysis, we can think of the following two possibilities. First, it is possible that the effect was caused by maintaining a high hemoglobin range. Second, it is possible that the effect was caused by erythropoiesis-stimulating agents, independently of hemoglobin. However, as described above, from the clinical background that there has been no report on whether erythropoiesis-stimulating agents themselves benefit patients, we believe that we can deny the second possibility. We also believe that the principal causal effect for the PR group is used in the evaluation of the treatment effect caused by high hemoglobin. Therefore, the merit of applying the principal stratification approach in this paper is that we can consider a more realistic situation than we can with the IPTW and g-estimation methods, and the purpose of applying this approach is to evaluate the treatment effect of high hemoglobin by using the estimated principal causal effect.

However, even if the treatment effect of high hemoglobin can be evaluated by applying the principal stratification approach, we cannot definitively specify in which subgroup any patient is included from the data observed. The result from ITT analysis in an overseas clinical trial showed that high hemoglobin levels made renal function worse [Drüeke T.B., et al. (2006)], whereas other results of exploratory analyses using hemoglobin levels observed after starting the treatment showed that high hemoglobin levels led to benefits for patients [Inrig, J.K., et al. (2012); Solomon, S.D., et al. (2010); Szczech, L.A., et al. (2008); Szczech, L.A., et al. (2010); Tsubakihara, Y., et al. (2012)]. However, as we described above, those results of exploratory analyses may be biased owing to the post-treatment variable problem. To discuss the treatment effect of high hemoglobin on renal function, the estimation derived from the appropriate analysis method is necessary, but that kind of estimation has not been reported to date. Therefore, although we cannot definitively specify in which subgroup any patient is included from the data observed, we believe that showing the benefit of high hemoglobin treatment itself leads to a meaningful suggestion for future discussion about hemoglobin target levels.

To make our discussion of this paper more clear, we present our concept of DE before the next Section. We do not set the assumption in this paper that DE is zero. In the case that DE is not zero, we can think about the two possibilities described above. Since, as already described, the second possibility can be denied by the clinical background, our opinion is that in the situation where DE is not zero, it is possible that the effect was caused by maintaining a high hemoglobin range. We can explain about the reason as follows. As described in Section 4, we define the post-treatment variable to frame potential subgroups as whether hemoglobin

values observed within 2 weeks after the start of treatment could achieve 11.0 g/dL or not. This means that hemoglobin, which is a time-dependent and continuous variable, is summarized by specifying the time and changing to binary data. Post-treatment variable, as defined in this paper, cannot sufficiently and entirely explain the hemoglobin values. Because of these reasons, we believe that it is worth considering the situation where DE is not zero. Therefore, we do not assume that DE is zero in this paper.

In this paper, we use the principal stratification approach to examine the post-treatment variables of the data of our motivation study and estimate the treatment effect of high hemoglobin level as the causal effect. We also interpret the adequacy of high hemoglobin treatment for patients with anemia from the estimated treatment effect.

The structure of this paper is as follows. In Section 2, we provide a brief summary of the motivation study. In Section 3, we explain the method based on the principal stratification approach for the post-treatment variables and give the required assumptions. In Section 4, we explain our results from the application of the method to the data of our motivation study. Details of the steps of the EM algorithm used through the method are provided in the Appendix.

2. Study Population and Definitions

A total of 321 patients with CKD were randomly assigned to a high hemoglobin target (11.0–13.0 g/dL, 161 patients; High group) or low hemoglobin target (9.0–11.0 g/dL, 160 patients; Low group). Patients assigned to the High group received darbepoetin- α , while those assigned to the Low group received epoetin- α . Two interim safety and efficacy analyses were planned during a 3-year follow-up period using an O'Brien-Fleming α -spending function. The significance levels at the first and second interims were allocated to 0.0060 and 0.0151, respectively. The final significance level was allocated to 0.0472. The primary endpoint was a composite of the time from baseline to the first occurrence of serum creatine level doubling, initiation of dialysis (maintenance dialysis), renal transplantation, or death. The patients' baseline characteristics are shown in Table 1. All factors were well-balanced between groups for the randomization. In the last analysis at week 144, the cumulative renal survival rate in the High group was 39.9% (95% confidence interval [CI], 30.7–49.1%) by the Kaplan-Meier method. In the same way, the cumulative renal survival rate in the Low group was 32.4% (95% CI, 24.0–40.8%). Renal survival rates were compared between the High and Low groups by log-rank test ($p = 0.111$). The Cox regression model was used to adjust groups and baseline covariates, while the estimated hazard ratio (HR) for the High versus Low group was 0.71 (95% CI, 0.52–0.98) [Tsubakihara, Y., et al. (2012)].

However, many patients assigned to the High group could not achieve the hemoglobin target range (11.0–13.0 g/dL) or could not keep their hemoglobin levels within the target range. These patients had difficulty increasing their hemoglobin levels from baseline to the target range or

Table 1. Patients' baseline characteristics

Characteristic	High Hb group ($n = 161$)	Low Hb group ($n = 160$)
Age (years)	65.2 ± 11.8	64.1 ± 11.7
Male (%)	80 (49.7)	71 (44.4)
Diabetes as comorbidity (%)	50 (31.1)	50 (31.3)
Weight (kg)	56.7 ± 11.2	56.6 ± 11.1
Height (cm)	157.8 ± 9.5	157.5 ± 8.3
Baseline Hb (g/dL)	9.2 ± 0.8	9.2 ± 0.9
Baseline creatinine (mg/dL)	3.54 ± 1.06	3.57 ± 1.08
Estimated GFR (mL/min/1.73m ²)	14.1 ± 5.0	13.9 ± 4.9
Plus-minus value are mean $\pm$ standard deviation; Hb, hemoglobin; GFR, glomerular filtration rate		

keeping the hemoglobin levels within the target range. Substantial numbers of patients in the High (44/161) and Low (17/160) groups failed to reach the target hemoglobin levels at more than half of the measurement time points during the study. Therefore, in the comparison of the subgroups of patients whose median hemoglobin levels observed during the study were within the target range, the renal survival rate was significantly better in the High group than in the Low group ($p = 0.020$) [Tsubakihara, Y., et al. (2012)]. As described in Section 1, however, the result from that subgroup analysis may be biased owing to the post-treatment variable problem.

3. Method

3.1 Estimation Considering Post-Treatment Variables by Principal Stratification

Here we introduce the method suggested by Tanaka to estimate the treatment effect based on the principal stratification approach [Tanaka, S., et al. (2007); Tanaka, S. (2008)] and consider a similar situation as in the motivation study described in Section 2.

In this situation, n patients are randomly assigned 1:1 to the Low or High hemoglobin treatment group and the primary endpoint is the renal survival time to death or renal failure. We let R_i be the vector of the treatment assignment for the i -th patient ($i = 1, \dots, n$),

$$R_i = \begin{cases} 0, & \text{Low hemoglobin treatment group} \\ 1, & \text{High hemoglobin treatment group} \end{cases}$$

where r_i is the actual value of R_i . The index i indicates the i -th patient. Let $T_i (\geq 0)$ be the survival time to death or renal failure and $C_i (\geq 0)$ be the censoring time. Let $T_i^{obs} = \min(T_i, C_i)$ be the observed renal survival time and $\delta_i = I(T_i^{obs} = T_i)$ be the non-censoring indicator. In the motivation study, we assume that the post-treatment variable is the achievement of the particular hemoglobin level such as the lower limit of the target hemoglobin range and is shown as

$$S_i^{obs} = \begin{cases} 0, & \text{Non response} \\ 1, & \text{Response} \end{cases}.$$

Let $\mathbf{X}_i = (X_{i1}, \dots, X_{iq})$ be the q -elements vector of the covariate. The observed data of the i -th patient defined above is shown as $\mathbf{O}_i = (T_i^{obs}, \delta_i, S_i^{obs}, R_i, \mathbf{X}_i)$. Let $t = 0$ be the start time of the study and $t = K$ be the end time of the study. For simplicity, censoring is assumed to occur completely at random with no tie in survival times.

We consider the assumptions required for estimating the principal causal effect. We first defined the potential outcome variables as follows: let $T_i^{(0)}$ and $T_i^{(1)}$ be the potential survival time of i -th patient under $R_i = 0$ and $R_i = 1$, respectively. Similarly, let $S_i^{(0)}$ and $S_i^{(1)}$ be the potential post-treatment variable of the i -th patient under $R_i = 0$ and $R_i = 1$, respectively.

[Assumption 1] Stable unit treatment value assumption (SUTVA) [Rubin, D.B. (1978)]

The potential outcome for the i -th patient is unrelated to R_j of the j -th patient ($i \neq j$).

Under SUTVA, let $T_i^{(r)}$ and $S_i^{(r)}$ be the variable of the survival time and the post-treatment variable when the i -th patient is assigned to $R_i = r$, respectively. We think that this assumption can be made for our motivation study because this study was designed as a randomization study and it can be plausible that the observed data $(T_i^{(r)}, S_i^{(r)}, R_i)$ are independent and identical distributed across patient. We then assume the consistency assumption, which means that actual observed data are linked to the potential outcome variables.

[Assumption 2] Consistency assumption [Rubin, D.B. (1978)]

$$T_i^{obs} = \begin{cases} \min(T_i^{(0)}, C_i) & \text{if } r = 0 \\ \min(T_i^{(1)}, C_i) & \text{if } r = 1 \end{cases}, \quad S_i^{obs} = \begin{cases} S_i^{(0)} & \text{if } r = 0 \\ S_i^{(1)} & \text{if } r = 1 \end{cases} \quad (1)$$

We can observe the results of $T_i^{(r)}$ and $S_i^{(r)}$ from the actual group to which the patient was assigned. Since part of the $\mathbf{S}_i = (S_i^{(0)}, S_i^{(1)})$ is not observed, we can describe $\mathbf{S}_i$ as

$$S_i^{mis} = \begin{cases} S_i^{(1)}, & r = 0 \\ S_i^{(0)}, & r = 1 \end{cases} \quad (2)$$

In a randomized study, we can set the following assumption.

[Assumption 3] Randomization assumption

$$T_i^{(0)}, T_i^{(1)}, S_i^{(0)}, S_i^{(1)} \prod R_i \mid \mathbf{X}_i \quad (3)$$

Also, we make the following assumption for $S_i^{(0)}$ and $S_i^{(1)}$:

[Assumption 4] Monotonicity assumption [Angrist, J.D., et al. (1996)]

$$S_i^{(0)} \leq S_i^{(1)} \quad (4)$$

The monotonicity assumption can classify patients into three principal strata of AR, PR, and NR (does not include DR). From the clinical point of view, $S_i^{(0)} > S_i^{(1)}$ is not appropriate; as such, we believe that a monotonicity assumption can be made for the motivation study.

Next we set the Cox proportional model to estimate AE and DE in the case of survival time as the outcome. Here we assume that AE and DE can be defined as the hazard functions for

$T_i^{(0)}$ and $T_i^{(1)}$ at time t .

[Assumption 5] Proportional hazard model

$$\lambda_i^{(0)}(t) = \lambda_{NR}(t) \exp(I_i^{AR} \beta_{AR} + I_i^{PR} \beta_{PR} + \mathbf{X}_i \boldsymbol{\beta}_x) \quad (5)$$

and

$$\lambda_i^{(1)}(t) = \lambda_i^{(0)}(t) \exp\{I_i^{PR} \beta_{AE} + (1 - I_i^{PR}) \beta_{DE}\} \quad (6)$$

where I_i^{AR} and I_i^{PR} are the indicator functions of each principal stratum and $\lambda_{NR}(t)$ is the baseline hazard for NR. Also, β_{AR} , β_{PR} , β_{AE} , and β_{DE} are unknown parameters corresponding to indexes and $\boldsymbol{\beta}_x$ is the unknown parameter vector of the covariate. β_{AR} and β_{PR} are the effects of principal strata for Always Responder and Partial Responder, respectively. Similarly, β_{AE} and β_{DE} are the Associative Effect and the Dissociative Effect, respectively. β_{AE} is the effect within the principal strata of NR and AR, and β_{DE} is the effect of PR. These are the parameters that we want to estimate in this paper. Therefore, we must estimate

$$\boldsymbol{\beta} = (\beta_{AR}, \beta_{PR}, \beta_{AE}, \beta_{DE}, \boldsymbol{\beta}_x^T)^T. \quad (7)$$

Actually, we cannot know the principal strata to which the patients with $R_i = S_i^{obs} = 0$ or $R_i = S_i^{obs} = 1$ belong; however, R_i and $\mathbf{S}_i = (S_i^{(0)}, S_i^{(1)})$ are any of the following six patterns under the monotonicity assumption: $(R_i, S_i^{(0)}, S_i^{(1)}) = \{(0, 0, 0), (1, 0, 0), (0, 0, 1), (1, 0, 1), (0, 1, 1), (1, 1, 1)\}$.

Since one element of $\mathbf{S}_i = (S_i^{(0)}, S_i^{(1)})$ is the unobserved result, the joint distribution of R_i and $\mathbf{S}_i = (S_i^{(0)}, S_i^{(1)})$ is the incomplete multinomial distribution. Since R_i is independent of $\mathbf{S}_i$ from the randomization assumption, so we can say that the parameters of multinomial probability are $(1-p)P_{i00}$, $(1-p)P_{i01}$, $(1-p)P_{i11}$, pP_{i00} , pP_{i01} , and pP_{i11} , in which we let $P_{i00} = \Pr\{\mathbf{S}_i = (0, 0) \mid \mathbf{X}_i\}$, $P_{i01} = \Pr\{\mathbf{S}_i = (0, 1) \mid \mathbf{X}_i\}$, $P_{i11} = \Pr\{\mathbf{S}_i = (1, 1) \mid \mathbf{X}_i\}$, and p be the common probability of assignment to $R_i = 1$ for all of patients. Additionally, we assume the following logistic models.

[Assumption 6] Logistic regression model

$$\Pr(S_i^{obs} = 1 \mid R_i = 0, \mathbf{X}_i) = \frac{\exp(\alpha_{00} + \mathbf{X}_i \boldsymbol{\alpha}_{0x})}{1 + \exp(\alpha_{00} + \mathbf{X}_i \boldsymbol{\alpha}_{0x})} \quad (8)$$

and

$$\Pr(S_i^{obs} = 1 \mid R_i = 1, \mathbf{X}_i) = \frac{\exp(\alpha_{10} + \mathbf{X}_i \boldsymbol{\alpha}_{1x})}{1 + \exp(\alpha_{10} + \mathbf{X}_i \boldsymbol{\alpha}_{1x})} \quad (9)$$

The regression parameters of (8) and (9), $\boldsymbol{\alpha}_{0x}$ and $\boldsymbol{\alpha}_{1x}$, determine the distribution of the principal strata. Using the consistency assumption, monotonicity assumption, and logistic regression model, we can describe P_{i00} , P_{i01} , and P_{i11} as

$$P_{i00} = \Pr\{\mathbf{S}_i = (0, 0) \mid \mathbf{X}_i\} = \Pr\{S_i^{(1)} = 0 \mid \mathbf{X}_i\} = 1 - \frac{\exp(\alpha_{10} + \mathbf{X}_i \boldsymbol{\alpha}_{1x})}{1 + \exp(\alpha_{10} + \mathbf{X}_i \boldsymbol{\alpha}_{1x})}, \quad (10)$$

$$P_{i11} = \Pr\{\mathbf{S}_i = (1, 1) \mid \mathbf{X}_i\} = \Pr\{S_i^{(0)} = 1 \mid \mathbf{X}_i\} = \frac{\exp(\alpha_{00} + \mathbf{X}_i \boldsymbol{\alpha}_{0x})}{1 + \exp(\alpha_{00} + \mathbf{X}_i \boldsymbol{\alpha}_{0x})}, \quad (11)$$

and

$$P_{i01} = \Pr\{S_i = (0, 1) \mid \mathbf{X}_i\} = 1 - (P_{i00} + P_{i11})$$

$$= \frac{\exp(\alpha_{10} + \mathbf{X}_i \boldsymbol{\alpha}_{1x})}{1 + \exp(\alpha_{10} + \mathbf{X}_i \boldsymbol{\alpha}_{1x})} - \frac{\exp(\alpha_{00} + \mathbf{X}_i \boldsymbol{\alpha}_{0x})}{1 + \exp(\alpha_{00} + \mathbf{X}_i \boldsymbol{\alpha}_{0x})}, \quad (12)$$

respectively. Therefore, we can get the estimators $\hat{\boldsymbol{\alpha}}_0$ and $\hat{\boldsymbol{\alpha}}_1$ from the logistic models of the treatment groups. The hazard function for survival time T_i at time t is

$$\lambda_i(t \mid \mathbf{S}_i, R_i, \mathbf{X}_i) = \lambda_0(t) \exp(\mathbf{Z}_i \boldsymbol{\beta}), \quad (13)$$

where $\mathbf{Z}_i = \{I_i^{AR}, I_i^{PR}, R_i I_i^{PR}, R_i(1 - I_i^{PR}), \mathbf{X}_i\}$. However, $\mathbf{Z}_i$ is unidentifiable from the data because some parts of $\mathbf{Z}_i$ are functions of S_i^{mis} . Since we cannot observe S_i^{mis} , $\hat{\boldsymbol{\beta}}$ cannot be estimated by the usual methods. We need to use the EM algorithm to estimate $\hat{\boldsymbol{\beta}}$. Tanaka recommended using the deterministic annealing EM algorithm by Ueda, N. and Nakano, R. (1998) when parameters are estimated because the reliability for the estimator in an early step of the usual EM algorithm is not high. The way to estimate $\hat{\boldsymbol{\beta}}$ using a deterministic annealing EM algorithm is shown in the Appendix. Also, since the asymptotic distribution of the estimator $\hat{\boldsymbol{\beta}}$ follows a normal distribution, the confidential interval of $\hat{\boldsymbol{\beta}}$ can be estimated [Herring, A.H., Ibrahim, J.G. (2001); Reilly, M., Pepe, M.S. (1995)].

4. Application to the Motivation Study

4.1 Settings

Our application is the motivation study described in Section 2. All analyses were adjusted by sex, age, baseline creatinine concentration, baseline hemoglobin concentration, and diabetes. Age, baseline creatinine concentration, and baseline hemoglobin concentration are divided into two categories each: <65 or ≥ 65 years old, <4.0 or ≥ 4.0 mg/dL, and <8.0 or ≥ 8.0 g/dL, respectively. The post-treatment variable S_i^{obs} is defined as whether the hemoglobin levels of the i -th patient observed within Week 2 achieved the criterion value. Here we set the criterion hemoglobin level at 11.0 g/dL, which is the boundary level between treatment groups in the motivation study. As described in Section 1, we want to think about the principal strata consisted by the responders, which reflect the differences among the patients' conditions. Although many definitions of response can be made, from the situation of the motivation study that the first visit after the start of treatment in this study is Week 2 and 11.0 g/dL is an intermediate value of hemoglobin between groups, we believe that the above definition of response is simple and conservative for composing the principal strata for our purposes.

We believe in the two types of situations: the situation in which DE for NR is the same as AR (Situation 1) and that in which DE for NR differs from AR (Situation 2). In Situation 2, we convert (6) into

$$\lambda_i^{(1)}(t) = \lambda_i^{(0)}(t) \exp\{I_i^{PR} \beta_{AE} + (1 - I_i^{PR})(1 - I_i^{AR}) \beta_{DE(NR)} + (1 - I_i^{PR})(I_i^{AR}) \beta_{DE(AR)}\} \quad (6)'$$

to break the unknown parameter of DE into separate parameters for AR and NR, which lets $\beta_{DE(NR)}$ and $\beta_{DE(AR)}$ be the unknown parameters of DE for NR and that for AR, respectively.

4.2 Results

The medians of the existence probabilities of principal strata estimated by equation (10) to (12) are 91%, 7%, and 2% for the NR, PR and AR groups.

Table 2 shows the results of the application in both situations. In Situation 1, the HR of DE (95% CI) was 0.72 (0.52–1.00), while that of AE was 0.62 (0.26–1.49). The results of Situation 2 in Table 2 show that the HR of DE for NR was 0.73 (0.53–1.02), while that for AR was 0.22 (0.08–0.55). Also, the HR of AE was 0.67 (0.30–1.51). The HR of DE for NR was similar to that of DE in Situation 1, while that of DE for AR was smaller than that of DE in Situation 1. The HR of AR in Situation 2 (3.01) was larger than that of AE in Situation 1 (1.38). The other variables were basically the same in both situations.

Table 2. Results of application to the motivation study

	Variable	HR	95% CI
Situation 1	Dissociative effect	0.72	0.52–1.00
	Associative effect	0.62	0.26–1.49
	Always Responder	1.38	0.72–2.62
	Partial Responder	0.91	0.84–0.98
	Sex (male*, female)	0.87	0.63–1.20
	Baseline hemoglobin (<8.0 g/dL*, ≥8.0 g/dL)	0.42	0.20–0.90
	Baseline creatinine (≥4.0 mg/dL*, <4.0 mg/dL)	0.51	0.36–0.71
	Age (≥65 years*, <65 years)	0.82	0.59–1.14
	Diabetes as comorbidity (non-diabetes*, diabetes)	1.17	0.83–1.65
Situation 2	Dissociative effect for Never Responder	0.73	0.53–1.02
	Dissociative effect for Always Responder	0.22	0.08–0.55
	Associative effect	0.67	0.30–1.51
	Always Responder	3.01	1.61–5.62
	Partial Responder	0.92	0.85–1.00
	Sex (male*, female)	0.87	0.63–1.20
	Baseline hemoglobin (<8.0 g/dL*, ≥8.0 g/dL)	0.42	0.20–0.90
	Baseline creatinine (≥4.0 mg/dL*, <4.0 mg/dL)	0.50	0.36–0.70
	Age (≥65 years*, <65 years)	0.82	0.59–1.13
	Diabetes as comorbidity (non-diabetes*, diabetes)	1.17	0.83–1.65

HR, hazard ratio; CI, confidence interval

*Reference

5. Discussion

Some large-scale overseas studies and the motivation study described here were unable to show the efficacy of high hemoglobin treatment. However, many patients allocated to the high hemoglobin treatment could not achieve or maintain a value within the target hemoglobin range. Some exploratory analyses were performed for those studies including the motivation study,

and the results of those analyses suggested that high hemoglobin levels led to the potential for decreased renal events such as death or the start of dialysis. However, since those results were derived from *ad hoc* analyses, they could not precisely show the effect of high hemoglobin levels due to the fact that the adjustment analyses using hemoglobin levels observed after starting treatment include bias because of the post-treatment variable problem that the estimates made from stratified analyses and adjustments for post-treatment variables can be biased. To address the post-treatment variable problem, Frangakis and Rubin suggested the principal stratification approach. Also, Tanaka suggested the method based on the principal stratification approach to estimate the treatment effect by considering post-treatment variables.

Taking into consideration the property of our motivation study, we believed that the principal stratification approach is more appropriate than other methods of causal inference. Additionally, we believed that we could show the treatment effect of high hemoglobin by using the principal causal effect for the PR group. Therefore, we decided to apply the method based on the principal stratification approach.

In this paper, we applied the method suggested by Tanaka to data from the motivation study of patients with CKD in 2005 to estimate the treatment effect of high hemoglobin considering the post-treatment variables. In our application, we set the post-treatment variable as whether hemoglobin levels observed after starting treatment achieved the particular criterion value (11.0 g/dL). We also created a situation (Situation 1) in which the same DE was used for NR and AR and a situation (Situation 2) in which DE for NR and DE for AR differed. In Situation 1, the HR of DE and AE was 0.72 (95% CI, 0.52–1.00) and 0.62 (0.26–1.49), respectively. That of DE was similar to the original result of 0.71 (0.52–0.98) by the intention to treat (ITT) analysis, while that of AE was smaller than the original result by ITT. These results mean that although the estimated 95% CI of HR is not narrow, high hemoglobin treatment may be more efficient for PR under our setting that the treatment effect between groups is derived by the difference from target hemoglobin level ranges, not from erythropoiesis-stimulating agents administered. In Situation 2, the HR of DE for NR, that of DE for AR, and that of AE were 0.73 (0.53–1.02), 0.22 (0.08–0.55), and 0.67 (0.30–1.51), respectively. The results of DE for NR and AE were similar to those of Situation 1, but the result of DE for AR was smaller than that in Situation 1. Also, from the results that the estimated HR and 95% CI of DE for NR group almost is equal to the original estimation by ITT and the median of estimated existence probability of NR group is 91%, it can be said that the original result may mostly be affected by NR group. The estimated HR of DE for AR is small and the estimated upper limit of 95% CI for that HR is even 0.55. It means that high hemoglobin treatment for AR drastically decreases the risk of renal event. These may suggest that patients in the AR group can benefit from high hemoglobin treatment under our setting about the treatment effect. The HR of AR was 1.38 in Situation 1, 3.01 in Situation 2, and >1.00 in both situations. These results suggest that patients in the AR

group were originally at higher risk than those in the NR group, but that result is inconsistent with the intuition in clinical situations in which patients in the AR group are in better condition than patients in the NR group. Since the 95% CI of the AR group was very wide, the reason the result is inconsistent with our intuition may be that the HR for the AR group could not be stably estimated because not many patients were classified into that group as described in Section 4.2 (the median of estimated existence probability for AR group was 2%).

Although some estimates are unstable, we could estimate the effect of high hemoglobin treatment for each principal stratum by using the method based on the principal stratification approach. From those results, although we cannot say that the estimated HR of AE substantially exceeded the result from ITT analysis, we believe that patients classified into PR group can receive benefit from high hemoglobin treatment. Moreover, although we need to consider that the estimated HR of DE for AR was very low by coincidence owing to the low existence probability of the AR group, the patients classified into the AR group may be able to considerably decrease the risk of a renal event.

Tanaka recommended using the deterministic annealing EM algorithm to consider that the reliability for the estimator in an early step of usual EM algorithm is not high when the parameters are estimated. Other types of the EM algorithm were recently suggested by many researchers. As such, we applied this method using other types of the EM algorithm [Wei, G.C.G., Tanner, M.A. (1990); Meng, X.L., Robins, J.M. (1993)], but the results did not differ from that of the deterministic annealing EM algorithm (data not shown).

There are some limitations to this application. We must set some assumptions for estimating the principal causal effect, but they cannot be validated from the data. We made the monotonicity assumption from the clinical point of view, but for the future we will need to think about the sensitivity analysis in consideration of the case that the assumption is not plausible. As we will see, not only from our concept of DE described in Section 1 but also the fact that the estimated HR of DE for NR was 0.73, we cannot deny that the definition of the response might not be plausible enough to make DE zero or quite small. In particular, we set the criterion value of 11.0 g/dL because it is the boundary value between treatment groups in the motivation study, but it is possible that a different criterion value is more appropriate. Similarly, we used the hemoglobin levels within Week 2, but it is necessary to consider the more appropriate duration in which the hemoglobin levels should be included. More exploratory analyses must be performed of alternate criterion values or alternate durations. We believe that the treatment effect between groups is derived by the difference from target hemoglobin level ranges rather than from erythropoiesis-stimulating agent administration because no report to date has demonstrated that erythropoiesis-stimulating agent administration benefits patients with renal anemia in a manner other than improved hemoglobin. Therefore, we cannot apply the results of this paper for a medicine that our setting is inappropriate.

The high hemoglobin level has been controversial, and there is still debate in the medical literature about its efficacy. Although, we need to consider that the definition of response in this paper may not be plausible enough, the results provided by the principal stratification approach suggested that among patients with CKD, there may be a subpopulation for which high hemoglobin treatment is recommended. Therefore, factors that can define such a subpopulation are expected to be determined in the future, so that more patients could benefit from high hemoglobin treatment.

In conclusion, our hypothesis was that the treatment effect of high hemoglobin could be evaluated by estimating the principal causal effect of the PR group, and here we applied the method based on the principal stratification approach to data from the motivation study and estimated the principal causal effects for the NR, AR and PR groups. Although the results should be cautiously interpreted, because there were some concerns, our findings suggested that not only patients in the PR group but also those in the AR group may benefit more from high hemoglobin treatment compared to those in the NR group. We conclude that patients with CKD may be able to receive benefit from high hemoglobin treatment, who can respond to that treatment.

REFERENCES

- Angrist, J.D., Imbens, G.W., Rubin, D.B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, **91**, 444–472.
- Besarab, A., Bolton, W.K., Browne, J.K., et al. (1998). The effects of normal as compared with low hematocrit values in patients with cardiac disease who are receiving hemodialysis and epoetin. *The New England Journal of Medicine*, **339**, 584–590.
- Dempster, A.P., Laird, N.M., Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, **39**, 1–37.
- Drüeke T.B., Locatelli, F., Clyne, N., et al. (2006). Normalization of hemoglobin level in patients with chronic kidney disease and anemia. *The New England Journal of Medicine*, **355**, 2071–2084.
- Frangakis, C.E., Rubin, D.B. (2002). Principal stratification in causal inference. *Biometrics*, **58**, 21–29.
- Hernán, M.A., Brumback, B., 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.
- Herring, A.H., Ibrahim, J.G. (2001). Likelihood-based methods for missing covariates in the Cox proportional hazards model. *Journal of the American Statistical Association*, **11**, 292–302.

- Inrig, J.K., Barnhart H.X., Reddan D., et al. (2012). Effect of hemoglobin target on progression of kidney disease: a secondary analysis of the CHOIR (Correction of Hemoglobin and Outcomes in Renal Insufficiency) trial. *American Journal of Kidney Diseases*, **60**, 390–401.
- Japanese Society of Nephrology. (2012). CKD Practice Guideline 2012. *The Japanese Journal of Nephrology*, **54**.
- Mark, S.D., Robins, J.M. (1993). A method for the analysis of randomized trials with compliance information: an application to the Multiple Risk Factor Intervention Trial. *Controlled Clinical Trials*, **14**, 79–97.
- Meng, X.L., Rubin, D.B. (1993). Maximum likelihood estimation via the ECM algorithm: a general framework. *Biometrika*, **80**, 267–278.
- Pfeffer, M.A., Burdmann, E.A., Chen, C.Y., et al. (2009). A trial of darbepoetin alfa in type 2 diabetes and chronic kidney disease. *The New England Journal of Medicine*, **361**, 2019–2032.
- Reilly, M., Pepe, M.S. (1995). A mean score method for missing and auxiliary covariate data in regression models. *Biometrika*, **82**, 299–314.
- Rosenbaum, P.R. (1984). The consequences of adjustment for a concomitant variable that has been affected by the treatment. *Journal of the Royal Statistical Society. Series A (General)*, **147**, 656–666.
- Robins, J.M., Hernán, M.A., Brumback, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology*, **11**, 550–560.
- Rubin, D.B. (1978). Bayesian inference for causal effects: the role of randomization. *The Annals of Statistics*, **6**, 34–58.
- Singh A.K., Szczech, L., Tang, K.L., et al. (2006). Correction of anemia with epoetin alfa in chronic kidney disease. *The New England Journal of Medicine*, **355**, 2085–2098.
- Solomon, S.D., Uno, H., Lewis, E.F., et al. (2010). Erythropoietic response and outcomes in kidney disease and type 2 diabetes. *The New England Journal of Medicine*, **363**, 1146–1155.
- Szczech, L.A., Barnhart, H.X., Inrig, J.K., et al. (2008). Secondary analysis of the CHOIR trial epoetin- α dose and achieved hemoglobin outcomes. *Kidney International*, **74**, 791–798.
- Szczech, L.A., Barnhart, H.X., Sapp, S., et al. (2010). A secondary analysis of the CHOIR trial shows that comorbid conditions differentially affect outcomes during anemia treatment. *Kidney International*, **77**, 239–246.
- Tanaka, S., Matsuyama, Y., Ohashi, Y. (2007). Estimating surrogate endpoints defined by principal causal effects. *2007 Joint Statistical Meetings, Salt Lake City, Utha, July 29 - August, 2007*, Abstract **123**.

- Tanaka, S. (2008). Evaluation of surrogate endpoints via principal stratification with an application to advanced prostate cancer trail data. *Tokyo University, Division of Health Sciences and Nursing, Graduate School of Medicine, Dissertation*.
- Tsubakihara, Y., Nishi, S., Akiba, T., et al. (2010). 2008 Japanese Society for Dialysis Therapy: guidelines for renal anemia in chronic kidney disease. *Therapeutic Apheresis and Dialysis*, **14**, 240–275.
- Tsubakihara, Y., Gejyo, F., Nishi, S., et al. (2012). High target hemoglobin with erythropoiesis-stimulating agents has advantages in the renal function of non-dialysis chronic kidney disease patients. *Therapeutic Apheresis and Dialysis*, **16**, 529–540.
- Ueda, N., Nakano, R. (1998). Deterministic annealing EM algorithm. *Neural Networks*, **11**, 271–282.
- Wei, G.C.G., Tanner, M.A. (1990). A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms. *Journal of the American Statistical Association*, **85**, 699–704.

Appendix

Let the parameter of the conditional distribution for S_i^{mis} be $\pi_i = \Pr(S_i^{mis} = 1 \mid \mathbf{O}_i)$. Here we consider the score function of β , $\mathbf{U}_\beta(\beta \mid \pi)$, and the score function of λ_0 , $\mathbf{U}_\lambda(\lambda_0 \mid \pi, \beta)$, where let π and λ_0 are the vector of π_i and λ_0 (T_i^{obs}). Since we assume that the missing mechanism of S_i^{mis} is complete at random, the consistent estimators of β and λ_0 can be estimated by the mean score method for the estimating equations, $\mathbf{U}_\beta(\beta \mid \pi) = \mathbf{0}$ and $\mathbf{U}_\lambda(\lambda_0 \mid \pi, \beta) = \mathbf{0}$ [Herring, A.H., Ibrahim, J.G. (2001); Reilly, M., Pepe, M.S. (1995)].

The detailed steps are as follows.

[Step 1] Estimate the maximum likelihood estimator $\hat{\alpha}_0$ and $\hat{\alpha}_1$ from (8) and (9), then compute $\hat{P}_{i00}$.

[Step 2] Set $\hat{\pi}^{(0)}$, $\hat{\beta}^{(0)}$, and $\gamma_0 = 0.1$ as the initial values of $\hat{\pi}$, $\hat{\beta}$, and γ , respectively.

[Step 3] Solve $\mathbf{U}_\lambda(\lambda_0 \mid \hat{\pi}_\gamma^{(m_\gamma-1)}, \hat{\beta}^{(m_\gamma-1)}) = \mathbf{0}$ in (m_γ) st step for γ , then compute $\hat{\lambda}_{ijk}^{(m_\gamma-1)}$ and $\hat{S}_{ijk}^{(m_\gamma-1)}$ by using

$$\lambda_{ijk} = \lambda\{T_i^{obs} \mid S_i = (j, k), R_i, \mathbf{X}_i\}$$

and

$$S_{ijk} = S\{T_i^{obs} \mid S_i = (j, k), R_i, \mathbf{X}_i\} = \Pr\{T_i > T_i^{obs} \mid S_i = (j, k), R_i, \mathbf{X}_i\}.$$

[Step 4] Compute $\hat{\pi}_\gamma^{(m_\gamma)}$ by assigning $\hat{P}_{ijk}$, $\hat{\lambda}_{ijk}^{(m_\gamma-1)}$, and $\hat{S}_{ijk}^{(m_\gamma-1)}$ to the following equations.

$$\begin{aligned} \pi_{i,\gamma} &= 1 & \text{if } R_i = 0, S_i^{obs} &= 1 \\ \pi_{i,\gamma} &= 0 & \text{if } R_i = 1, S_i^{obs} &= 0 \\ \pi_{i,\gamma} &= \frac{(P_{i01}\lambda_{i01}S_{i01})^\gamma}{(P_{i00}\lambda_{i00}S_{i00})^\gamma + (P_{i01}\lambda_{i01}S_{i01})^\gamma} & \text{if } R_i = 0, S_i^{obs} &= 0, \delta_i = 1 \end{aligned}$$

$$\begin{aligned}
\pi_{i,\gamma} &= \frac{(P_{i01}S_{i01})^\gamma}{(P_{i00}S_{i00})^\gamma + (P_{i01}S_{i01})^\gamma} & \text{if } R_i = 0, S_i^{obs} = 0, \delta_i = 0 \\
\pi_{i,\gamma} &= \frac{(P_{i11}\lambda_{i11}S_{i11})^\gamma}{(P_{i01}\lambda_{i01}S_{i01})^\gamma + (P_{i11}\lambda_{i11}S_{i11})^\gamma} & \text{if } R_i = 1, S_i^{obs} = 1, \delta_i = 1 \\
\pi_{i,\gamma} &= \frac{(P_{i11}S_{i11})^\gamma}{(P_{i01}S_{i01})^\gamma + (P_{i11}S_{i11})^\gamma} & \text{if } R_i = 1, S_i^{obs} = 1, \delta_i = 0
\end{aligned}$$

[Step 5] Solve $\mathbf{U}_\beta(\boldsymbol{\beta} \mid \hat{\boldsymbol{\pi}}_\gamma^{(m_\gamma)}) = \mathbf{0}$, then get $\hat{\boldsymbol{\beta}}^{(m_\gamma)}$.

[Step 6] Repeat Step 3–5 until the convergence condition is satisfied. At that point, multiply γ by $c = 1.1$, then return to Step 3.

[Step 7] When $\gamma > 1$, set the converge value $\hat{\boldsymbol{\beta}}^{(\infty)}$ as the final estimation result.