

Anaemia of CKD—the CHOIR study revisited[†]

Ajay K. Singh¹, Lynda Szczech², Kezhen L. Tang³, Huiman Barnhart⁴, Shelly Sapp²,
Marsha Wolfson³ and Donal Reddan^{2,5}

¹Renal Division, Brigham and Women's Hospital and Harvard Medical School, Boston, USA ²Renal Division, Duke University Medical Center and Duke Clinical Research Institute, Durham, USA ³Ortho Biotech Clinical Affairs, Bridgewater, NJ, USA ⁴Department of Biostatistics and Bioinformatics, Duke University, Durham, USA ⁵Department of Medicine, University College Galway, Ireland

Keywords: anaemia; CKD; epoetin; haemoglobin; kidney disease

Introduction

Anaemia is a common complication of kidney disease and is treated with erythropoietin. However, the optimal target haemoglobin (Hb) in treating CKD anaemia has become quite controversial. There have been several randomized controlled studies of erythropoietin among kidney patients published over the past several years. The Normal Hematocrit study (NHS), a study of high-risk dialysis patients [1], was published in 1998 and was controversial at the time, as it appeared to contradict conventional wisdom by suggesting that aggressive anaemia correction may not be beneficial to haemodialysis patients with pre-existing cardiovascular disease. More recently, two studies of pre-dialysis CKD patients, the Correction of Hemoglobin Outcomes in Renal Insufficiency (CHOIR) study [2] and the Cardiovascular Risk Reduction by Early Anemia Treatment with Epoetin Beta (CREATE) study [3], focused on patients at early stages of CKD. The publication of these two studies has again attracted much discussion [4]: CHOIR demonstrating increased risk of normalization of hemoglobin [2] and CREATE showing no cardiovascular benefit, but a higher absolute number of cardiovascular events and a higher rate of dialysis events [3]. Specifically relating to the CHOIR study, questions have been raised in a

recent article in NDT by Levin [4], regarding the choice of epoetin dose, the decision to change the Hb target soon after the study was started, the number and reasons for withdrawal, differences in baseline characteristics between the two randomized groups and the decision by the Data Safety Monitoring Board (DSMB) to stop the study. The purpose of this article is to respond to Levin's comments, as they pertain to the CHOIR study.

The CHOIR study

CHOIR was an open-label, randomized trial that studied 1432 patients with CKD: 715 patients were randomized to receive epoetin alpha targeted to achieve an Hb of 13.5 g/dl, and 717 were randomized to receive epoetin alpha targeted to achieve an Hb of 11.3 g/dl [2]. One hundred and thirty centres in the United States participated in the study. The mean and median study duration was 16 months. Key eligibility criteria included age >18 years, estimated GFR of 15–50 ml/min/1.73 m², and the exclusion of patients with refractory anaemia, active malignancy, and active cardiovascular disease. The primary end point was a composite of death, myocardial infarction, congestive heart failure (CHF) hospitalization (excluding hospitalization during which renal replacement therapy occurred) and stroke. Secondary endpoints included the components of the primary endpoint, as well as cardiovascular and total hospitalizations, quality of life (QOL) and the time-to-dialysis. Two hundred and twenty-two composite events occurred: 125 events among the high Hb group and 97 events among the low Hb group $P=0.03$, hazard ratio (HR) of 1.34; with 95% confidence interval of 1.03 and 1.74. The higher rate of composite events was explained largely by a higher rate of death (48% higher risk, $P=0.07$) or CHF hospitalization (41%, $P=0.07$) (Table 1). There was an improvement in QOL in both groups, but no statistically significant incremental benefit in the higher

Correspondence and offprint requests to: Dr A. K. Singh, Renal Division, Brigham and Women's Hospital, 75 Francis Street, Boston, MA 02115, USA. Email: asingh@partners.org

[†]This editorial comment is in response to the editorial by Adeera Levin: Understanding recent haemoglobin trials in CKD: methods and lesson learned from CREATE and CHOIR. *Nephrol Dial Transplant* 2007; 22: 309–312.

Table 1. Summary of CHOIR study

	13.5 Hb Arm	11.3 Hb Arm	Difference	P-value
Deaths	52	36		0.0674
MI	18	20		0.7836
CHF hospitalization	64	47		0.0727
Stroke	12	12		0.9803
Composite events	125	97	28	0.0312
Observed event rate	17.5%	13.5%	4.0%	
K-M estimated event rate by the end of 3rd year	29.5%	24.9%	4.6%	
Hazard ratio (95% CI)	1.337 (1.025, 1.743)			

vs lower Hb group. Furthermore, there was a statistically significantly higher rate of serious adverse events in the higher vs lower Hb group.

CHOIR-related questions

Why were the Hb targets of 11.3 and 13.5 g/dl chosen as targets and why were these targets changed from the initial targets?

In designing the CHOIR trial, there was extensive discussion regarding the choice of the two Hb targets. Several factors determined the choice. From a purely trial design perspective, the widest possible difference between the two Hb levels would have been considered ideal. Initially, an Hb of 10.0–10.5 g/dl for the low Hb group and 13.0–13.5 g/dl for the high Hb group was written into the protocol. However, 11–12 g/dl level emerged as the standard of care in the treatment of pre-dialysis CKD patients—driven in large part by the Kidney Disease Outcome Quality Initiative (KDOQI) guidelines [5,6]. Indeed, we wondered whether the initial very slow enrollment in the trial reflected a reluctance of investigators to randomize patients to a low Hb arm considered to be less than the standard of care. The target Hb levels were, therefore, reviewed. After seeking input from the CHOIR Scientific Advisory Board, as well as individual site PIs, we decided that it was in the best interests of the trial to change the lower Hb level to 11.3 g/dl a level within the KDOQI recommended range [5]. At the time of the protocol amendment, which modified the Hb targets, 347 (24%) of the patients had been enrolled. The number of person years accrued before and after the protocol amendment was 132 and 1807, respectively. While changing the target Hb levels once a trial commences is not ideal, changes such as these are quite common in trials. They frequently reflect changes in standard of care that impact upon recruitment or the emergence of new evidence. Other published studies reflect this as well. We did not feel that changing the Hb target level had a major impact on the results of this study because it was done relatively early in the course of the trial. Hence, even though 24% of patients were treated to the initial targets, only a small fraction of patient years had accrued. A sensitivity analysis of the primary endpoint, by excluding the 347 patients

enrolled before the change of target Hb level, showed that the result was similar with a P -value = 0.0268 and a HR of 1.433 (95% CI, 1.040, 1.973) comparing the high Hb group to the low Hb group.

Why 10 000 units subcutaneously of epoetin-alpha as the initial dose used in the study?

The initial epoetin alpha dose of 10 000 units for CHOIR was the subject of much deliberation by the study leadership. The chosen dose was selected to reflect dosing strategies in use in the United States at the time of study initiation and reflected doses used in other US-based studies [7,8].

The dose of epoetin used in CHOIR [2] was indeed higher than that used in the CREATE study [3]. However, CREATE had fewer diabetics and CREATE patients had lesser degrees of co-morbidity than CHOIR patients. Doses of erythropoietin in other non-US studies [9] are also lower than those in CHOIR. And this may, therefore, also reflect as yet unexplained international differences in epoetin-alpha-prescribing patterns that have been documented extensively in several studies [10,11]. More to the point, perhaps, in CHOIR, the Hb target was achieved successfully for the lower, but not the higher Hb group, suggesting that our algorithm and dose choices were conservative for achieving the higher Hb target in this population.

Imbalance in key baseline covariates

In the CHOIR paper, there were differences in only two baseline characteristics, namely self-reported history of hypertension and a prior history of coronary bypass graft (CABG) among about 100 baseline characteristics (2% of baseline characteristics). This small amount of imbalance is expected in a randomized clinical trial. These differences were unfortunate, because they could potentially have accounted for some of the observed difference between the two groups with respect to the primary composite or one or more of the secondary endpoints. Indeed, the difference in the rate of CABG between the higher and lower Hb groups might be a potential explanation for the higher rate of CHF in the higher vs lower Hb groups. How valid do we think these points are? It is

reasonable to ask whether the data plausibly reflects differences that occurred by chance alone or if there was a breakdown in the randomization process. To support the possibility that this was a chance occurrence is the fact that all other factors were balanced. Indeed, the mean haematocrit values were identical in the two groups. Another supporting factor is that blood pressure at baseline, and during the course of the trial, was very similar—in fact, blood pressure fell for both groups. Furthermore, the use of key concomitant medications such as diuretics, beta blockers, angiotensin blockers and other anti-hypertensive medications were similar between the two groups. Other cardiovascular covariates were also quite similar at baseline between the two groups. In particular, no difference was observed with respect to myocardial infarction or percutaneous coronary intervention.

Adjusting for baseline covariates using a regression model, as has been suggested [4], would be one approach to this issue. Pertinent to this point is the multivariate analysis using Cox proportional hazard regression model, performed and released by Ortho Biotech in the Clinical Study Report [12]. Although point estimates for treatment effect are similar in direction and magnitude to those reported in the primary analysis in the CHOIR paper, the effect of treatment group does not reach conventional statistical levels of significance. This analysis controls for baseline albumin and reticulocyte count, as well as the demographic and clinical factors of age, pre-existing CHF, National Health and Nutrition Examination Survey congestive heart failure (NHANES) CHF score and atrial fibrillation and flutter, and demonstrated a HR of 1.243, $P=0.11$ for the higher as compared with the lower treatment group. Germane to this is the fact that in the Ortho Biotech modelling, the two variables that were different at baseline (i.e. history of CABG and self-reported history of hypertension) were not included in the model since they were not significantly associated with the composite events when other

variables were in the model. Further, albumin and reticulocyte count were also not different between groups at baseline and arguably may reflect factors that are part of the mechanism by which epoetin may or may not effect Hb adequately (i.e. inflammation and iron stores). However, in subsequent and preliminary model-building, when we did adjust these two baseline co-variables that were different at baseline, without other variables in the model, the data yielded affect estimates of the composite events similar to the unadjusted analysis.

Clearly, taking the potential effect of these baseline factors into account, together with the effect of potential confounding and the impact of loss of power, is important in evaluating the significance of post-hoc analyses of a RCT. In any event, further exploration of analyses such as these is planned and will be submitted for publication in the near future. Until more evidence becomes available from our own studies, or emerges from other ongoing studies, we regard a conservative safety message as prudent. A similar conclusion is reached by others in a recent meta-analysis [13] and an accompanying editorial [14].

High rate of withdrawal in CHOIR

Both groups in the CHOIR study experienced a relatively high rate of withdrawal. The withdrawal due to reasons other than primary endpoint and study early termination was 38.3%, representing 549 patients. Specifically, 242 patients (16.9%) withdrew because of the initiation of renal replacement therapy and 307 patients (21.4%) withdrew due to other reasons, such as patient request, investigator request, protocol violation, or loss to follow-up. The reasons for withdrawal are presented in Table 2. The withdrawal rate of 21.4% due to reasons other than RRT initiation is similar to rates observed in other clinical trials involving subjects with CKD (including the CREATE study). Importantly, there were no

Table 2. Information on withdrawals

	Hb 13.5		Hb 11.3		Total	
	(High Hb group)		(Low Hb group)			
	(N = 715)		(N = 717)		N = 1432	
	<i>n</i>	%	<i>n</i>	%	<i>n</i>	%
Number of patients	278	38.9	271	37.8	549	38.3
RRT	131	18.3	111	15.5	242	16.9
Withdrawals (not RRT)	147	20.6	160	22.3	307	21.4
Patient request	53	7.4	76	10.6	129	9.0
Investigator or sponsor request	38	5.3	35	4.9	73	5.1
Pregnancy	1	0.1	1	0.1	2	0.1
Decision by the sponsor	7	1.0	3	0.4	10	0.7
Adverse event	11	1.5	5	0.7	16	1.1
Protocol violation	0	0.0	6	0.8	6	0.4
Other	37	5.2	34	4.7	71	5.0

significant differences between groups among these 307 patients with respect to demographics, medical history, adverse experiences and time-to-withdrawal. It should also be clearly pointed out that all patients randomized were actually included in the analysis. Those who withdrew because they required RRT or for other reasons were included in the time-to-event analysis up until the point they withdrew and they, therefore, contributed considerable numbers of days at risk.

Did the DSMB follow RCT convention with regards to stopping the study?

The DSMB's recommendation to stop the trial was consistent with the guidelines pre-specified in the protocol and the DSMB charter. Indeed, the protocol specified that 'efficacy and futility boundaries will serve as non-binding guidelines for early stopping, but not rules.' At the second interim analysis, neither the O'Brien–Fleming boundary nor the futility boundary was crossed. Consistent with their charter, the DSMB in May 2005 recommended that the study be terminated, after reviewing the results of the planned second interim analysis. This decision was based on data from 1411 patients with 145 observed composite events (80 in the high Hb arm and 65 in the low Hb arm), resulting in an estimated HR of 1.26 (95% CI 0.91, 1.75) ($P=0.16$), indicating a trend for a higher risk in the high Hb arm compared with the low Hb arm. The conditional power for demonstrating a benefit of the high Hb arm by the scheduled end of the study was less than 5% for all plausible values of the true effect for the remaining data after the study was terminated on 26 May 2005.

With regard to adjustments made for interim analyses: it is true that repeatedly testing interim data can inflate Type I error rate (i.e., falsely rejecting the null hypothesis). This has been discussed in detail elsewhere [15]. The alpha spending function is one way to implement group sequential boundaries that control for overall type I error rate. Specifically, the overall alpha (i.e. the overall probability of falsely rejecting the null hypothesis at each interim analysis and at the final analysis) is set by convention at 0.05. The more interim analyses one does, the more alpha you spend during the course of the trial and the smaller final alpha level at the final analysis. We used O'Brien and Fleming alpha-spending function to provide the adjusted alpha levels at each of the interim analyses and final analysis. The alpha level at each interim or final analysis is the largest P -value needed in order to claim statistical significance at this analysis. Even if one were to adjust for all four of the interim analyses as specified in the protocol (i.e. if the study had run its course), the final alpha level would be 0.04. Adjusting for two interim analyses that were performed, the final alpha level was 0.045. Either way (adjusting for two interim analyses or four interim analyses), the results were statistically significant with the reported P -value of 0.03.

Quality of Life measurements in the CHOIR study

QOL has been shown to improve in targeting patients to a higher Hb in several previous studies. However, these benefits in QOL have been modest at best. In CHOIR, three QOL instruments were utilized. These were: Linear Analogue Scale Assessment (LASA), Kidney Disease Questionnaire (KDQ), and the SF-36 Health Survey. These instruments have been validated as measures in many previous studies. The QOL questionnaires were done at baseline, month-6, month-12, month-24 and month-36 of the study. The limitations of QOL assessment have been pointed out elsewhere [16]. In CHOIR, while QOL improved for both the higher and lower Hb groups, no statistically significant incremental improvement was observed for those patients targeted to the higher Hb level, even using longitudinal analytic methods. For one domain of QOL (Role Emotional), patients randomized to the higher Hb group actually fared worse [2].

Patients who dropped out of the study did not add clinical information

In CHOIR, we used survival analyses to analyse the effect of the intervention on the primary and selected secondary endpoints. In these types of analyses, survival times are analysed by measuring follow-up time from randomization to the occurrence of last observation. In CHOIR, the main analysis assessed the effect of intervention on the composite of all-cause mortality, CHF, hospitalization, stroke and myocardial infarction. The analysis compared time to events among treatment groups. Until patients in CHOIR withdrew, reached an endpoint, or needed renal replacement therapy (RRT), these patients contributed time and risk/benefit, and thus contributed data to the analysis. Therefore, 1432 patients are included in the analysis and withdrawals are censored, rather than excluded in the analysis. Hence, it is not valid, as has been suggested elsewhere [4], that this is a study of anything other than the 1432 subjects that were randomized in the study. Thus, following convention and since all 1432 subjects contributed to the analysis, we provided baseline data on the randomized population and not on a sub-set of the randomized population.

Conclusions

Anaemia treatment with epoetin is an important issue because anaemia is a common complication and because both the benefit and risk of treatment may be related to the achieved Hb level. The CHOIR study has provided insights into the optimal target Hb in CKD patients. It has limitations, like all other forms of clinical research, and it needs to be interpreted in the totality of evidence in this area. In this regard, as noted earlier, a recent meta-analysis [13] further confirms that the directionality of other published studies in this

area is concordant with the results generated by CHOIR. Nevertheless, while further studies are awaited, and notwithstanding commercial motivations, a conservative interpretation of the results will maximize patient safety. It is important to draw lessons from CHOIR. First, there was an increased risk for poor outcomes among patients randomized to the higher Hb level. This should motivate adherence to the FDA guideline of maintaining Hb levels below 12 g/dl. While CHOIR was unable to exclude the possibility that a high epoetin dose could have been important, as has been suggested elsewhere [4], the study was not designed to address this question. Secondly, CHOIR showed that there was no incremental improvement in QOL when targeting a higher vs lower Hb level. While other studies have shown improvement in QOL, these improvements have been modest, and as has been pointed out by others [16], open to methodological concerns. Finally, extremely large sample sizes will be needed in future studies to answer important questions regarding the efficacy and safety issues raised by recent studies.

Conflict of interest statement. Dr Singh reports having received consulting fees from Ortho Biotech Clinical Affairs, Amgen, Roche, Merck (Germany), Abbott, Watson and Horizon Blue Cross Blue Shield, and lecture fees from Ortho Biotech Clinical Affairs, Roche, Amgen, Abbott, Watson, Scios, Pfizer and Genzyme; having served on advisory boards for Ortho Biotech Clinical Affairs, Roche, Watson, Advanced Magnetics and Amgen; and having received grant support from Ortho Biotech Clinical Affairs, Dialysis Clinic, Roche, Amgen, Baxter, Watson, Aspreva and the National Institutes of Health (NIH). Dr Szezech reports having received consulting fees from Ortho Biotech Clinical Affairs, Nabi Pharmaceuticals, Gilead, Kuraha, Acologix and Roche; lecture fees from Nabi Biopharmaceuticals, GlaxoSmithKline, Genzyme, Abbott, Amgen and Ortho Biotech; and grant support from Ortho Biotech Clinical Affairs and the NIH. Dr Barnhart reports having received consulting fees and grant support from Ortho Biotech Clinical Affairs and grant support from the NIH. Drs Tang and Wolfson report being employees of Ortho Biotech Clinical Affairs. Dr Reddan reports having received consulting fees from Ortho Biotech Clinical Affairs and Shire Pharmaceuticals; lecture fees from Amgen, Novartis, Pfizer, AstraZeneca and General Electric; and grant support from Ortho Biotech Clinical Affairs, Amgen and Novartis. No other potential conflict of interest relevant to this article was reported.

References

1. Besarab A, Bolton WK, Browne JK *et al.* The effects of normal as compared with low hematocrit values in patients with cardiac

- disease who are receiving hemodialysis and epoetin. *N Engl J Med* 1998; 339: 584–590
2. Singh AK, Szczech L, Tang KL *et al.* CHOIR Investigators. Correction of anemia with epoetin alfa in chronic kidney disease. *N Engl J Med* 2006; 355: 2085–2098
3. Druke TB, Locatelli F, Clyne N *et al.* CREATE Investigators. Normalization of hemoglobin level in patients with chronic kidney disease and anemia. *N Engl J Med* 2006; 355: 2071–2084
4. Levin A. Understanding recent haemoglobin trials in CKD: methods and lesson learned from CREATE and CHOIR. *Nephrol Dial Transplant* 2007; 22: 309–312
5. NKF-DOQI clinical practice guidelines for the treatment of anemia of chronic renal failure. National Kidney Foundation-Dialysis Outcomes Quality Initiative. *Am J Kidney Dis* 1997; 30[4 Suppl 3]: S192–S240
6. NKF-K/DOQI Clinical Practice Guidelines for Anemia of Chronic Kidney Disease: update 2000. POWER Study Group. *Am J Kidney Dis* 2001; 37 [1 Suppl 1]: S182–S238
7. Provenzano R, Garcia-Mayol L, Suchinda P *et al.* POWER study group. Once-weekly epoetin alfa for treating the anemia of chronic kidney disease. *Clin Nephrol* 2004; 61: 392–405
8. Provenzano R, Bhaduri S, Singh AK. PROMPT Study Group. Extended epoetin alfa dosing as maintenance treatment for the anemia of chronic kidney disease: the PROMPT study. *Clin Nephrol* 2005; 64: 113–123
9. Roger SD, McMahon LP, Clarkson A *et al.* Effects of early and late intervention with epoetin alpha on left ventricular mass among patients with chronic kidney disease (stage 3 or 4): results of a randomized clinical trial. *J Am Soc Nephrol* 2004; 15: 148–156
10. Papatheofanis FJ, McKenzie RS, Mody SH, Suruki RY, Piech CT. Dosing patterns, hematologic outcomes and costs of erythropoietic agents in predialysis chronic kidney disease patients with anemia. *Curr Med Res Opin* 2006; 22: 837–842
11. Germain M, Ram CV, Bhaduri S, Tang KL, Klausner M, Curzi M. Extended epoetin alfa dosing in chronic kidney disease patients: a retrospective review. *Nephrol Dial Transplant* 2005; 20: 2146–2152. Epub 28 June 2005
12. Ortho Biotech. Clinical Study Report. http://download.veritasmedicine.com/PDF/CR004588_CSR.pdf
13. Phrommintikul A, Haas SJ, Elsik M, Klum H. Mortality and target haemoglobin concentrations in anemia patients with chronic kidney disease treated with erythropoietin: a meta-analysis. *Lancet* 2007; 369: 381–388
14. Strippoli GF, Tognoni S, Navaneethan SD, Nicolucci A, Craig JC. Haemoglobin targets: we were wrong, time to move on. *Lancet* 2007; 369: 346–350
15. DeMets DL, Lan KK. Interim analysis: the alpha spending function approach. *Stat Med* 1994; 13: 1341–1352; discussion 1353–1356
16. Strippoli GF, Manno C, Schena FP, Craig JC, Paoletti E, Cannella G. Haemoglobin and haematocrit targets for the anaemia of chronic renal disease. *Cochrane Database Syst Rev* 2003; 1; CD003967

Received for publication: 29.1.07

Accepted in revised form: 9.3.07