Hypothesis Versus Association: The Optimal Hemoglobin Target Debate

IN THIS ISSUE of the American Journal of Kidney Diseases, Levin et al1 present the results of their Canadian randomized trial of early versus delayed anemia management with high (12.0 to 14.0 g/dL [120 to 140 g/L]) versus low (9.0 to 10.5 g/dL [90 to 105 g/L]) hemoglobin (Hb) level targets in patients with chronic kidney disease (CKD). This study is presented within a year of the publication of 2 other trials with similar designs addressing similar questions2,3 and when preliminary results of the Cardiovascular Risk Reduction by Early Anemia Treatment With Epoetin Beta (CREATE) study are coming into the public domain.4

Publication of this trial also is timely given 2 major recent developments in randomized controlled trial methods: prospective registration in the public domain and standardized explicit methods for reporting.5,6 The Canadian trial predates the requirement for prospective registration of key design features for subsequent publication in all major general biomedical journals and nephrology and transplantation journals; therefore, the trial is not reported with a unique identifying number. This will be required in the future. The editorial boards of nephrology journals should be acknowledged for their support of this initiative, which is designed to ensure publication of complete trial data irrespective of whether it is favorable to the investigational intervention.

The second development is an explicit and comprehensive checklist for reporting of 23 domains of trial design (Consolidated Standards for Reporting Trials [CONSORT]).6 It should be emphasized that CONSORT is a mechanism primarily to ensure good reporting, and not necessarily design, and has not been universally and explicitly adopted by nephrology journals. One core feature of CONSORT is nicely illustrated in the Canadian trial. A flow chart shows patient flow from enrollment and randomization to the end of study and any reasons for exclusion; thus, any losses to follow-up or non–intention-to-treat analysis are readily evident. Allocation concealment is a key item required for reporting by the CONSORT; it is the method by which patients are allocated to intervention, to be reported in such a way that it is explicit that it is truly random, preventing the investigators from knowing what the next patient will receive, and thus preventing selection bias. The Canadian trial used sequentially labeled, sealed, opaque envelopes that were randomly ordered by a third party, a method that ensures adequate allocation concealment. Data were analyzed according to the intention-to-treat principle, but of 172 patients randomized, 20 did not contribute follow-up data, with more lost in the control group. Being an open-label study, patients and physicians knew to which group patients were allocated, which may have introduced cointerventions differentially, but the primary end point, left ventricular mass index (LVMI), was measured blind to group assignment. In summary, the methods were reported comprehensively and explicitly, and key domains of allocation concealment, intention-to-treat analysis, blinding of outcomes assessors, and loss to follow-up were all performed well, with the possible exception of the last item.

This trial is based on 2 assumptions using data from observational studies: that there is an association between higher Hb levels and LVMI, and that a lower LVMI improves survival in patients with CKD.7 The trial was performed to test this first assumption, and not the second. Even if a higher Hb level had been found to decrease LVMI, it still would remain uncertain whether early anemia management with higher Hb targets would result in a survival advantage.

The surprising finding is that there was no significant difference in the change in LVMI with early anemia management and a higher Hb target (5.2 ± 30.3 g/m²) compared with delayed management and a lower Hb target (0.4 ± 25.0 g/m²; \( P = 0.28 \)).
The investigators discuss many reasons for this finding, and we suggest that there are 3 main categories. The lack of a significant relationship between the experimental intervention (high Hb level target achieved with early treatment of anemia) and the outcome (reduction in LVMI) may be caused by bias (systematic error), chance (random error), or a true absence of effect.

How may the findings be explained by sources of bias? In this scenario, we postulate there is a real effect, but it was not found because of favoring of the lower-Hb group occurring at the time of randomization, during the follow-up period, or when outcomes were measured. Table 1 in the Canadian trial of Levin et al suggests that selection bias favoring the lower-Hb group is very unlikely, and measurement bias when LVMI was measured also is unlikely given this was blinded to the treatment assignment. What about biased cointerventions? Table 3 shows that both groups received cointerventions equally, although the study was unblended; therefore, this also is an unlikely source of bias. The remaining source is the differential ascertainment of patients to have their outcomes measured: 74 of 87 in the control group compared with 78 of 85 in the intervention group. We are not informed about the characteristics of these patients. It is possible that patients who had worse outcomes were less likely to complete the study, and that this was more likely to occur in the control group, resulting in bias toward the null (ie, less difference between the true groups than is true).

May the findings of Levin et al be explained by chance? Was there lack of power to detect a significant difference in the change in LVMI? The investigators provide explicit assumptions that underpin their power calculations, although the basis for these assumptions is not provided. They achieved their target recruitment goal, and the variability in LVMI and loss to follow-up were planned for. The absolute mean difference in LVMI change from baseline to 24-month echocardiogram measurement between the control and treatment groups was 4.85 g/m² (95% confidence intervals, −4.0 to 13.7 g/m²). A clinically important difference in LVMI may not have been excluded; in short, we do not know. Perhaps more important, the expected difference in LVMI of 15 g/m² did not occur, and this may be because the intervening variable, a substantial difference in Hb levels, did not occur. As the investigators emphasized, the expected decrease in Hb levels in the control group (receiving delayed treatment with no erythropoietin at all or erythropoietin targeting lower Hb levels) was not observed (Hb levels of 11.5 to 11.7 g/dL [115 to 117 g/L] from study month 2 versus 12.6 to 13.0 g/dL [126 to 130 g/L] in the experimental group). The study was not powered for such a small difference in Hb levels, although the investigators do not provide a rationale for their expected difference in LVMI based on expected differences in Hb levels. Why did the groups not separate in Hb indices as much as expected? Perhaps this is another example of the Hawthorn effect: the observation that patients in trials tend to do well.

What if there truly is no significant decrease in LVMI with early anemia management targeting a higher Hb level? In this case, the study of Levin et al demonstrates the very important point that associations found in observational studies are not enough to advocate for the efficacy and broad adoption of interventions, as was done for erythropoietin. As the investigators themselves point out, intervention questions need to undergo trial by proper hypothesis-testing (randomized) studies because observational data are prone to unpredictable bias and confounding that only the randomization process will account for. Changes in Hb levels may not be on the same causal pathway as LVMI in patients with CKD. Even if a decrease in LVMI had been shown, the causal relationship with cardiovascular death would remain unproven. The widely advocated pathway “erythropoietin administration → increase in Hb → decrease in LVMI → reduction in deaths” may not be related causally, such that effects of erythropoietins on Hb level (increase) and LVMI (decrease) really need to be disentangled, and one cannot assume that finding a significant effect of erythropoietins on Hb level (increase) is a good surrogate for a decrease in LVMI, or, more importantly, for improving survival.

The key question is, Are Hb level and LVMI valid “surrogate” end points? In addition, is Hb level an appropriate predictor of improvements in LVMI? And is LVMI itself a valid surrogate for what really matters, which is mortality and
other patient-level (cardiovascular) end points? A surrogate is a measurable outcome (such as a laboratory test) that is responsive to the effect of an intervention (such as an increase in Hb level with erythropoietins) and also associated causally with a clinically important outcome (such as a decrease in all-cause or cardiovascular mortality with erythropoietins). A valid surrogate end point therefore captures the full effect of an intervention, but earlier in the causal chain of events.9-11 These end points generally are preferred to hard end points in randomized trials because of the decreased costs and sample size needed, shorter study duration, and improved sensitivity to differences in treatment. Other examples of surrogates used in patients with CKD include dialysis adequacy and acute rejection.12-14

However, not all surrogates are valid proxies of clinically important patient-level outcomes. For a surrogate to be valid, 2 criteria must be met. First, there must be a strong, independent, and consistent association between the surrogate and the clinically important outcome, which comes from observational studies and has been found in the case of Hb level and mortality and LVMI and cardiovascular mortality. Second, but more important, there also must be evidence that using an intervention that changes a surrogate results in an expected change in the patient-based outcome distal to the surrogate in the same causal pathway for the disease in question. This more stringent criterion requires a randomized trial, which measures both the surrogate and the hard end point.9-11 Do we have this type of data available for Hb levels and LVMI? There are now more than 20 randomized trials of anemia management and Hb level targets in patients with CKD, most of which have been conducted in dialysis patients and have relied on surrogate end points, including Hb level and LVMI. Only 4, the largest of which was the trial of Besarab et al13 (n = 1,236) conducted in hemodialysis patients, provided patient-level mortality data. Invariably, either no effect on patient-level outcome (mortality) was shown when an effect on a surrogate (Hb level or LVMI) was found, or no difference in both (no significant decrease in LVMI with higher Hb levels or erythropoietin doses and no effect on mortality) has been found. The large trial of Besarab et al13 independently and the meta-analysis of results of these trials provided mortality data informing us that the benefits of higher Hb targets (reduced risk for seizures) appear to be outweighed by the harms (increased risk for hypertension and death). These findings in dialysis patients, which are in clear contrast to those of some large observational studies,7 are now compounded further by those of the most recent trials,2,2 including that of Roger et al2 and the CREATE study4 (both conducted in patients with CKD ["predialysis"]). These failed to show a significant cardiovascular benefit with higher Hb levels achieved with early treatment of anemia (similar to the Canadian study) and only found a significant benefit of higher Hb levels on quality-of-life measures. There also are data for patients with anemia of cancer showing increased mortality and disease progression with greater doses of erythropoietins, which warrant careful consideration.16

In conclusion, this important Canadian study should remind us of the pitfalls of relying on association data from observational studies when deciding about interventions and investing little in solid hypothesis testing with randomized trials. Trials are needed and must be performed with proper methods, relying on validated end points. Spending considerable resources in randomized trials that measure only unvalidated surrogates is not in the best interest of our patients or scarce research resources. This study also reminds us that the major challenge in predialysis and dialysis patients is improving overall survival. Both pediatric and adult CKD registry data unfortunately have shown that, despite introduction of a number of novel interventions (including erythropoietins, novel dialysis techniques, better calcium and phosphate metabolism management, and lipid-lowering agents), there has not been the expected significant improvement in survival of patients with CKD during the past 2 decades.17-19 If we really want to improve the survival of these patients, we need to explore novel interventions and better understand the risk factors for CKD and its outcomes, rather than spend considerable time, energy, and resources on interventions for which no substantial evidence of improvements in patient-level end points has been found and is not
likely to be found based on extensive searching that has been done already.\textsuperscript{15}

\textit{Giovanni F.M. Strippoli, MD, MPH (Hons)}

\textit{Jonathan C. Craig, MBChB, PhD}

NHMRC Centre for Clinical Research Excellence in Renal Medicine

Cochrane Renal Group

University of Sydney

Sydney, Australia

REFERENCES


