Are Concerns About Reliability in the Trial to Assess Chelation Therapy Fair Grounds for a Hasty Dismissal?
An Alternative Perspective
Sanjay Kaul, MD

Preconceived notions are the locks on the door to wisdom.
—Merry Browne

Trial to Assess Chelation Therapy (TACT) was a prospective, randomized, double-blind, placebo-controlled trial designed to evaluate the effect of ethylenediaminetetraacetic acid-based intravenous chelation therapy on cardiovascular outcomes. A total of 1708 stable patients >50 years of age with a previous myocardial infarction were enrolled at 134 North American sites and followed for a median of 55 months. The overall results showed that chelation therapy reduced the risk of primary major adverse cardiac events plus end point—a composite of death, myocardial infarction, stroke, coronary revascularization, or hospitalization for angina. Several findings in this report are worthy of the spotlight. First, the results in the expanded diabetes mellitus cohort were indistinguishable from those in the original cohort, providing reassurance that the definition of diabetes mellitus was modified before data analysis.

Second, the treatment benefit was greater in patients enrolled at non-CAM sites (hazard ratio, 0.38; 95% confidence interval [CI], 0.24–0.62) compared with complementary and alternative medicine sites (hazard ratio, 0.76; 95% CI, 0.53–1.08; nominal P for interaction=0.022), thereby alleviating the concern for potential trial misconduct at CAM sites.

Third, adjustment for multiple comparisons using the conservative Bonferroni correction that accounted for 9 prespecified subgroups yielded a statistically significant difference in favor of chelation therapy for the primary end point, thereby controlling the false-positive error rate and mitigating the risk of a misleading conclusion. Adjustment for multiplicity did not, however, yield significant differences for additional end points. This is understandable because trials are seldom designed adequately to assess treatment effects across multiple end points within subgroups. Had the investigators adjusted for 3 additional subgroups that were not prespecified but presented in the original trial report (high-dose vitamins, CAM sites, prior revascularization), the primary end point would still be significant at the adjusted P-value threshold of 0.003 (0.036/12=0.003) as would treatment by diabetes mellitus interaction: adjusted P for interaction=0.043 (adjusted P=1−[1−unadjusted P]^k, where k is the number of subgroups).

Nonetheless, the number of subgroups explored is arguably large for a trial this size, compounding the challenge of non-replication in future trials.

Fourth, patients with diabetes mellitus experienced a significant reduction in stringent MACE that was similar in magnitude to risk reduction in MACE plus end point. This suggests that treatment effect was not driven by soft and less reliable end points such as coronary revascularization and hospitalization for angina. Inclusion of soft but prevalent end points is primarily motivated by trial feasibility considerations, but they nevertheless increase the potential for ascertainment error and misclassification that typically biases the results toward the null, which is conservative for a superiority trial such as TACT and anticonservative for a noninferiority trial. However, because the trial was not adequately powered for secondary outcomes, these analyses should be deemed exploratory.

Fifth, 115 of 633 patients (18%) withdrew consent or were lost to follow-up and did not experience a primary end point event before censoring. This amount of missing data, while...
vexing, is not unexpected given the onerous treatment protocol involving multiple weekly infusions during a prolonged period of time. Even so, 61% completed all 40 infusions, and 73% completed 30 infusions. The missing data can challenge the interpretation of trial results, especially if there is evidence of differential dropout and informative censoring. Although it has been claimed (albeit without supportive evidence) that potential unmasking of treatment allocation could lead to differential dropout, the higher dropout in the placebo arm would be expected to lead to bias against chelation treatment. There does not seem to be major imbalance in either the key prognostic covariates or the reasons for missingness among the dropouts in the 2 treatment groups, indicating low potential for informative censoring. Furthermore, the results of sensitivity analyses conducted by the investigators do not support the assertion that missing data biases the results in favor of chelation treatment. Nonetheless, the sensitivity analysis approach adopted by TACT investigators does not follow the recommended statistical procedure for addressing missing data such as multiple imputation. The TACT investigators should be encouraged to conduct these analyses expeditiously and report them posthaste.

Sixth, it has been argued that the trial was unethical because there was no compelling clinical or preclinical evidence that chelation therapy has significant efficacy against atherosclerotic cardiovascular disease, and given that chelation therapy can cause harm, the risk was not minimal. A Bayesian analysis would not look kindly on the results because of the low prior probability of treatment effect (the so-called implausibility argument). This is an uncharitable (and unwarranted) interpretation of the data because previous systematic reviews concluded, “insufficient evidence to decide on the effectiveness or ineffectiveness of chelation therapy in improving clinical outcomes among people with atherosclerotic cardiovascular disease.” It is axiomatic that absence of evidence of efficacy is not the same as evidence of the absence of efficacy.

From a Bayesian perspective, the strength of evidence is often summarized using a Bayes factor, which is a measure of how well 2 competing hypotheses (the null and the alternate) predict the data. The Bayes factor and the corresponding strength of evidence for the primary end point result in TACT overall, and diabetic cohorts are shown in Table 1. The P value of 0.035 for TACT overall cohort translates into a Bayes factor of 0.108, which means the evidence supports the null hypothesis =1/9th as strongly as it does the alternative. This reduces the null probability from 50% pretrial (justified by suspension of one’s belief in treatment effect) to 10% post-trial. Although this does not represent strong evidence against the null, it does reduce the level of skepticism surrounding chelation therapy. In the diabetic cohort, the nominal P value of 0.0002 translates into a Bayes factor of 0.002 (1/500), which reduces the extremely skeptical prior null probability of 95% to 4% post-trial, indicating very strong evidence against the null.

A formal Bayesian analysis of TACT shown in Table 2 provides additional insights. The choice of noninformative or vague prior can be reasonably justified, reflecting the uncertainty associated with the possible benefit of chelation therapy. In this case, the posterior is driven entirely by the evidence (as in the frequentist approach). Two types of skeptical priors are also used, each centered on a null difference with probability distribution reflecting 5% and 1% probability of extreme treatment effects (odds ratio [OR] <0.75 or >1.33), respectively. In the overall cohort, the posterior OR shifts from 0.82 using the noninformative prior to 0.87 (95% CI excluding OR of 1) using skeptical prior 2. Similarly, for the diabetic cohort, the posterior OR shifts from 0.59 to 0.77 (95% CI excludes OR of 1) using adjusted results and to 0.89 (95% CI does not exclude OR of 1) using adjusted results.
results. Thus, if we start from a position of skepticism, integrating the results of the TACT trial reduces the degree of skepticism. This is exactly how Bayesian analysis helps modify prior beliefs by incorporating new evidence and upgrading knowledge. One would require a prior of 2-fold increase in risk with chelation (not supported by published evidence) to completely nullify the results of the TACT trial, an arguably implausible conjecture!

The presumption that the TACT trial provides actionable evidence for clinical practice is not endorsed even by the TACT investigators who are appropriately circumspect in their conclusions and interpretation of the results. The best case interpretation of the data is that the evidence is inconclusive. Under such circumstances, the only worthwhile course of action is to conduct additional studies to replicate the findings and to adjudicate the uncertainties about the mechanism of benefit, explore incremental benefit associated with concomitant high-dose vitamin therapy, and help understand why the clinical outcome benefit was not accompanied by improvement in quality of life. That is how ideally science and knowledge should progress.

Finally, TACT highlights the double standard when it comes to accepting inconvenient results not aligned with our preconceived notions on so-called dubious quack cures such as chelation versus eagerly anticipated results of de rigueur cures such as gene transfer or stem cell therapy. Although the debate surrounding TACT is clearly warranted, the arguments that the TACT results are not valid or reliable are overstated. Consequently, the calls for a hasty dismissal are unfair and undoubtedly unjustified.

Disclosures

None.

References


Key Words: Editorials ■ Bayes theorem
Are Concerns About Reliability in the Trial to Assess Chelation Therapy Fair Grounds for a Hasty Dismissal?: An Alternative Perspective
Sanjay Kaul