Statistics and the Relationship of Clinical Research to Clinical Practice

Clinicians will offer better care to their patients if they are able to critically appraise original literature and apply evidence based therapy to their daily practice[1]. Still, determining success or failure of a treatment or preventive agent depends on a number of considerations, including whether an established effective treatment already exists; whether the disease for which the new treatment is sought is severe or life-threatening; the probability and magnitude of harmful effect; and the probability and magnitude of likely benefit.

When clinical trials of a new treatment are conducted, it is assumed there is a true, underlying effect of the treatment that clinical trial can estimate. Investigators use statistical methods to help understand the true effect from the results of the trial. For some time the paradigm for statistical inference has been hypothesis testing. The investigator starts from what is labeled a "null hypothesis": the hypothesis that the statistical procedure is designed to test and, possibly, disprove. Typically, the null hypothesis is that there is no difference between outcomes as a result of the treatments being compared. In a randomized controlled trial to compare an experimental treatment with a placebo, the null hypothesis can be stated: "The true difference in the effects of the experimental and control treatments on the outcome of interest is zero." A challenging issue arises when statistically non-significant results are large enough to be clinically relevant. Methotrexate (MTX) use in treatment of scleroderma (systemic sclerosis, SSc) is a good example of this challenge. SSc is a chronic autoimmune disorder with a highly variable course and an average prevalence of 1-2 cases per 100,000 people worldwide[2,3]; the rarity of the disease, the clinical heterogeneity, and the lack of effective agents to date render therapy a major challenge. Currently, no guidelines for the treatment of SSc are available, and no proven overall disease-modifying therapies have been identified. Three studies of the antimetabolite MTX in SSc patients showed little efficacy in improving clinical variables such as total skin score, carbon monoxide diffusing capacity, forced vital capacity, erythrocyte sedimentation rate, and general health as measured by visual analog scale[4-6]. The largest study[5] is a randomized controlled trial in which 35 patients with diffuse SSc were treated with MTX and 36 patients with placebo. After 12 months, patients in the MTX group had a tendency for better skin scores, better diffusion capacities, and favorable physician global assessments, but these variables and the patient's global assessments were not significantly different from placebo-treated patients. Thus, it is concluded that the findings do not provide evidence that MTX is significantly effective in the treatment of early diffuse SSc.

In this issue of *The Journal*, Johnson, *et al*[7] demonstrate how Bayesian analysis conveys more relevant information to clinicians, using the example of MTX in SSc. When comparing 2 hypotheses using the same information, traditional statistical method would typically result in the rejection or non-rejection of the original hypothesis with a particular degree of confidence, while Bayesian methods would yield statements that one hypothesis was more probable than the other. The Bayesian paradigm states that probability is the only measure of one's uncertainty about an unknown quantity. In a Bayesian clinical trial, uncertainty about an endpoint is quantified according to probabilities. Statistical significance (as defined by $p \leq 0.05$) may bear little relation to clinical significance, and the traditional analysis using p values may be misleading. This includes situations in which an important clinical decision must be based upon a study that has low statistical power. Incorrect interpretation of the p value and 95% confidence interval may lead to misguided clinical interpretation of the study results. In cases of small sample size, labeling a treatment ineffective by relying solely on the p value may not be appropriate, and the Bayesian analysis would be more attractive "primarily" because it uses available data from other studies to potentially reduce study sample size[8]. Bayesian computations have been successfully used in many clinical studies including the assessment of relative cost-effectiveness of treatments in health economics and effect of therapy[9,10].

As stated by the authors[7] "this analysis allows for more

flexible and clinically relevant interpretations of the data." One advantage of the Bayesian approach is that experimental data may be directly interpreted; the 95% CI of the posterior probability distribution corresponds to the set in which 95% of experimental data may be expected based upon the particular experiment that has been performed. Another advantage is that it enables the researcher to interpret more freely the multiple test outcomes (in this case skin score, etc.) without testing for alpha inflation. We agree that Bayesian analysis can convey a more clinically relevant interpretation of the trial data. However, it is not clear if this analysis can be extended to other situations. Thus, no data are available on the applicability of this approach in reverse situations. As an example, consider a situation where an investigator finds significant results using the traditional approach, but when the *a priori* knowledge is examined, the posterior probability of effect may become much lower than the anticipated 95% using the Bayesian approach. In addition, the potential misuse of this approach is possible, as when findings do not achieve the 5% level of significance, tempting researchers to present their data in the Bayesian format. Moreover, substantial *a priori* knowledge may introduce potential ethical concerns in the conduct of trials when transitioning from Phase II to Phase III, whereas in studies using Bayesian approaches that is avoided by the independent replication of the frequentist approach.

The Bayesian approach should not be viewed as a statistical alternative for investigators trying to demonstrate a "treatment effect." Essentially the frequentist and Bayesian approaches to the design and analysis of clinical trials provide complementary information regarding the strength of statistical evidence regarding particular conclusions. Because they address different perspectives, there is much to be gained by considering both analytic approaches. It is generally true that in the absence of prior information, classical and Bayesian approaches should yield equivalent statistics. For certain priors the Bayesian posterior odds test is equivalent in large samples to the classical likelihood ratio test for some significance level and vice versa. The Bayesian paradigm does not create positive results from a negative trial. As the authors stated, "If a treatment has no effect, then on average the posterior probability of treatment being better than placebo in a Bayesian analysis will be 50%." What the authors are proposing here does not therefore entail a radical change in the way clinicians judge the effectiveness of a treatment. Nevertheless, it might help to encourage a wider recognition that the interpretation of confidence intervals is only valid if you work within a Bayesian framework using a uniform prior distribution. If there is important preexisting information that needs to be taken into account, it must be incorporated formally in the analysis or informally in the qualitative process of drawing conclusions. In the case of clinical significance of MTX for SSc, the fairly low probability of harm may justify its use.

SADIK A. KHUDER, PhD,
Professor of Medicine, Public Health
and Homeland Security,
University of Toledo School of Medicine,
Toledo, Ohio;
BASHAR KAHALEH, MD,
Professor of Medicine,
Division of Rheumatology and Immunology,
University of Toledo School of Medicine,
Toledo, Ohio 43617, USA.

Address reprint requests to Dr. Kahaleh.

REFERENCES

1. Evidence-Based Medicine Working Group. Evidence-based medicine: a new approach to teaching the practice of medicine. JAMA 1992;268:2420-5.
2. Denton CP. Systemic sclerosis: clinical features and management. Medicine 2006;34:480-8.
3. Thompson AE, Pope JE. Increased prevalence of scleroderma in southwestern Ontario: a cluster analysis. J Rheumatol 2002;29:1867–73.
4. Das SN, Alam MR, Islam N, et al. Placebo controlled trial of methotrexate in systemic sclerosis. Mymensingh Med J 2005;14:71-4.
5. van den Hoogen FH, Boerbooms AM, Swaak AJ, et al. Comparison of methotrexate with placebo in the treatment of systemic sclerosis: a 24 week randomized double-blind trial, followed by a 24 week observational trial. Br J Rheumatol 1996;35:364-72.
6. Pope JE, Bellamy N, Seibold JR, et al. A randomized, controlled trial of methotrexate versus placebo in early diffuse scleroderma. Arthritis Rheum 2001;44:1351-8.
7. Johnson SR, Feldman BM, Pope JE, Tomlinson GA. Shifting our thinking about uncommon disease trials: the case of methotrexate in scleroderma. J Rheumatol 2009;36:323-9.
8. Diamond GA, Kaul S. Prior convictions: Bayesian approaches to the analysis and interpretation of clinical megatrials. J Am Coll Cardiol 2004;43:1929–39.
9. O'Hagan A, Stevens JW, Montmartin J. Bayesian cost-effectiveness analysis from clinical trial data. Stat Med 2001;20:733–53.
10. Babapulle M, Joseph L, Bélisle P, Brophy J, Eisenberg M. A hierarchical Bayesian meta-analysis of randomised clinical trials of drug-eluting stents. Lancet 2004;364:583-91.
11. Andrews DWK. The large sample correspondence between classical hypothesis tests and Bayesian posterior odds tests. Econometrica 1994;62:1207-32.