Commentary: Biostatistics, biological mechanisms and Bayes: lessons from the magnesium trials

KL Woods

The controversy which has surrounded the interpretation of the randomized controlled trials (RCT) of intravenous Mg²⁺ in acute myocardial infarction has focused attention on the place of both meta-analysis and the mega-trial in therapeutic research. Stated simply, a fixed-effects meta-analysis of seven small controlled trials indicated a highly significant 55% reduction in odds of death among Mg²⁺ treated patients. These early findings prompted a conventionally powered RCT in 2300 patients (LIMIT-2). Patients allocated to receive intravenous Mg²⁺ for 24 hours experienced relative reductions of 24% (95% CI: 1–43%) in 28-day mortality, 25% (95% CI: 7–39%) in early left ventricular failure and 21% (95% CI: 5–35%) in coronary heart disease mortality during 1–5 years’ follow-up.1 Separately, a very similar Mg²⁺ regimen was examined in ISIS-4, a factorial mega-trial which enrolled 58,000 patients worldwide (the other treatments being oral captopril and oral nitrate).2 ISIS-4 showed no significant mortality reduction in the Mg²⁺ group. How can these findings be reconciled? What does this experience tell us about the interpretation of trials in general and the therapeutic role of Mg²⁺ in particular? Higgins and Spiegelhalter have now examined the potential contribution of classical and Bayesian meta-analysis to the resolution of these inconsistencies.3

They conclude that neither a fixed-effects nor a random-effects model of classical meta-analysis adequately describes the results of the 15 magnesium trials now available. Both are, however, informative. The fixed-effects model reveals the extreme heterogeneity which arises when the ISIS-4 result is added to the other 14 (P < 0.0002). The inference is that ISIS-4 differed in some rather fundamental way, which should prompt a close examination of the delivered protocol. A meta-analysis with a random-effects model yields a highly significant pooled odds ratio (0.53, 95% CI: 0.36–0.77) for all 15 trials including ISIS-4. If the underlying assumptions of that model are accepted, it can be inferred from the totality of the evidence that Mg²⁺ treatment is beneficial despite the null result of ISIS-4.4

Choice of model entails a judgement of how much weight should be accorded to a mega-trial by virtue of its size alone. The rationale for the very large, simple trial is the precision with which small treatment effects can be distinguished from random error. However, size per se does not make the estimate intrinsically more valid or more generalizable than a conventional trial. The simplification of recruitment and data collection increases the risks of protocol deviation, poor data quality, misclassification and non-trial use of trial treatments—all of which create a bias towards the null.4 In the nitrate comparison of ISIS-4, for example, 60% of patients received non-trial nitrates (presumably for perceived clinical need). The null result of the nitrate comparison is uninterpretable, and influential opinion has ignored it.5 In the captopril comparison, no survival benefit was seen in over 8000 patients with heart failure at entry, yet conventional trials have convincingly shown that absolute mortality reduction is greatest in this type of patient (for whom ACE inhibitors are now standard therapy).5

What does the Bayesian perspective add to meta-analysis? Higgins and Spiegelhalter test the assertion that a sceptical prior applied to the meta-analysis of the early magnesium trials would have allowed the non-significant result of ISIS-4 to be anticipated. They show that extreme scepticism would have been required—certainly more than is justified by the evidence from a dozen laboratories that Mg²⁺ reduces the size of experimental myocardial infarction.6 Choosing an appropriate prior reintroduces the idea of biological plausibility into the interpretation of empirical evidence from trials.

Designing a clinical trial or a meta-analysis requires a mechanistic hypothesis for the treatment effect being examined. Too often this is unstated and unexamined in empirical research. The RCT adopts the paradigm of the laboratory experiment, but often without the same concern to ensure that the conditions are optimized to test the hypothesis in question. Statisticians judge the quality of a trial design on a range of criteria, but seldom explicitly ask whether the details of the protocol as delivered (drug, dose, timing, patients sampled) are capable of revealing a treatment effect if one exists. A mechanistic model must inform the judgement on whether it is appropriate to pool trials.

Higgins and Spiegelhalter conclude that it is vital to explore possible reasons for heterogeneity between trials. They review the evidence that trial size or underlying risk might explain (in the statistical sense) the heterogeneity of the magnesium trials, but do not comment on possible biological effect modification by the timing of treatment. How plausible is it that delayed (8 hours), predominantly post-thrombolytic, administration of Mg²⁺ as in ISIS-4 has a different effect from early (3 hours) pre-thrombolytic administration as in LIMIT-2?7 Since the publication of LIMIT-2 and ISIS-4, several groups of laboratory investigators have tested this hypothesis by developing models of myocardial ischaemia and reperfusion to replicate the trial conditions as closely as possible. These confirm that myocardial protection by Mg²⁺ occurs at the point of reperfusion, when
(paradoxically) most ischaemic myocardial damage occurs. The laboratory studies show that myocardial protection by Mg\(^+\) is critically dependent on timing of treatment and is abolished if treatment is delayed until after reperfusion has occurred.\(^8-10\) These data make the null result of ISIS-4 and the heterogeneity of the trials unsurprising. Acute myocardial infarction is a rapidly evolving process and, for treatments intended to minimize myocardial loss, time is a powerful effect modifier—as is well recognized for thrombolytic drugs.

Trialists have perhaps taken too far the dictum of Bayes’ contemporary, John Hunter: ‘Don’t think, try the experiment’. We ignore consideration of biological plausibility at our peril when we design clinical trials or judge the appropriateness of meta-analysis. The strength of Bayesian analysis is that it allows the full range of available evidence to influence subjective probability, and thereby helps to close the damaging gap which has opened up between mechanism and empiricism in therapeutic research.

References


