Identifying Saturation of Heterogeneity in Preclinical Datasets to Inform Decision to proceed with Translation to humans

Ezgi Tanriver- Ayder[1], Nicky Welton[2], Hugo Pedder[2], Malcolm Macleod[1]

[1] University of Edinburgh, [2] University of Bristol

Background: Conventional drug development involves sequence of in vitro and in vivo studies prior to testing safety and efficacy in human populations. Translation is initiated once data from animal studies are considered sufficiently convincing to justify the costs and risks of human testing. However, thresholds of evidence at which translation might occur are not well defined, and the judgement often appears subjective. Drug discovery in stroke, Alzheimer's disease and multiple sclerosis has been characterised by substantial efficacy in animal studies which has not been observed in clinical trials. Part of this translation failure might be due to overstatement of efficacy in animal studies at high risk of bias. Alternatively, it may be that the circumstances of animal testing have been too specific to reliably predict efficacy in humans. That is, within an animal study, treatment effects are highly homogenous (due to selected circumstances population, timing and outcome). This feature of animal studies leads high level of heterogeneity between studies due to differences in design. In clinical trials by contrast, heterogeneity predominantly arises within the study between patients, due to for instance a range of ages, sex and disease severity. Understanding heterogeneity in results from animal studies is important to evaluate whether current animal study evidence is sufficient to justify translation, or whether further preclinical research is necessary.

Objectives: The aim of this study was empirically to investigate how estimates of heterogeneity from meta-analyses of animal studies change as evidence accumulates, and the precision with which any impact of features of the included studies that are associated with heterogeneity are identified.

Methods: Using seven datasets of different sizes from systematic reviews of drugs tested in animal models of stroke, we perform cumulative random effects meta-analyses and inspect the changes in the proportion of Cochran's Q over its degrees of freedom, I^2 statistic and the between-study variability. Additionally, we perform cumulative meta-regression to explore the change in the heterogeneity measures, the proportion of heterogeneity explained by a covariate and the precision of the coefficient estimates.

Results: As the number of included studies increased, Q value increased. In contrast, I^2 and Q/df first increased with small number of studies then decreased slowly and stabilised when around 180 studies were included. Similarly, for between-study variability estimated using REML, we first observed an early peak, a trough, a second peak and a gradual declined. Cumulative meta-regression showed similar trends to random effects analysis except that the precision of the estimate of beta coefficients increased with inclusion of more studies. Cumulative change in proportion of heterogeneity explained by a variable showed an initial peak with a later decrease where it gradually stabilised after around 100 studies.

Conclusion: These preliminary findings suggest that it may be possible to identify systematic characteristics of heterogeneity within preclinical datasets which, when considered alongside the size of effects observed, risks of bias in individual studies and publication bias, can be used to guide decisions to proceed with human testing.