editorial

Is MOOSE a loose goose?

John P Brown

Professor and Chair, Department of Community Dentistry, University Texas Health Science Center San Antonio, Texas, USA

A proposal for reporting meta-analysis of observational studies in epidemiology ('MOOSE') has been published. Meta-analysis of observational studies is outwardly similar in methodology to that applied to randomised controlled trials (RCT), but it does differ inherently. It is less certain and less definitive. Single summary estimates of effect derived from it are suspect. RCT and observational studies are complementary in nature, however, and they provide assessments of efficacy and effectiveness respectively. Meta-analysis of both types of study is needed, and the MOOSE proposal is a useful one.

The York review of fluoridation of drinking water (see <http://york.ac.uk/inst/crd/fluorid.htm>) seemed to satisfy no-one. The reviewers were critical of the quality of available evidence. Fluoridation proponents were disappointed at the degree of benefit reported. Opponents were upset that no associations of harm were evident. Little discussion was provoked, however, about the applicability and methods of meta-analysis applied to observational studies versus experimental studies (RCT).

Water fluoridation is not the result of an experimental study, but of a natural experiment. It is observational. It does not aim to deliver a certain dose of fluoride to all members of a community. Rather it aims to reproduce the observation that at a certain (optimal) fluoride level, caries reduction is maximised with minimal occurrence of mild fluorosis of enamel, and without harmful effects on health generally. The misunderstanding of this basis for fluoridation has resulted in studies being designed which assumed that fluoride tablets were a gold standard. In current times there has also been the expectation by some that the goal was to achieve zero prevalence of fluorosis while maintaining caries prevention

across a population, using various fluoride sources, such as water and toothpaste.

Randomisation is not possible in studies of fluoridation. A neutral site for dental examinations and assessment of radiographs without knowledge of site have helped. But the relative absence of caries and presence of mild fluorosis does bias examiners.

Dietary, risk-factor exposure, health educational and other community-based oral studies are also observational in nature, not experimental. Confounding and study subject selection biases are not balanced but inherent in observational studies, whose internal validity can be threatened by unrecognised confounding factors. Does this mean that observational studies are second rate, poor science, fail to meet the 'gold' standard, and ought to be rejected by both research funders and editors? It must be understood that although meta-analysis of observational studies and RCT appear to be similar processes, they are based on different scientific methods. Black¹ stated the case for the complementary nature of these two approaches. He described the limitations of RCT as being sometimes unnecessary, inap-

propriate, impossible or inadequate. When an intervention is dramatic, the likelihood of confounders is so small that an experimental study may be unnecessary. RCT are not appropriate for measuring infrequent adverse outcomes, prevention of rare events, outcomes far in the future or when the randomisation itself reduces the effectiveness of the therapy. This occurs where the patient or the investigator needs to fully participate in the intervention.

An RCT can be impossible due to cultural, political and social circumstances. Clinicians and patients can be reluctant to participate, ethical objections may be insurmountable, government policy may preclude certain studies and legal obstacles may not be overcome. Contamination across test and control therapies is a barrier in some RCT, and the immensity of the task of evaluating all healthcare and health interventions means that a variety of methods must be considered.

RCT may be inadequate because of factors that weaken their external validity or the potential for making generalisations from their findings. This possibility is often ignored, even in systematic reviews. Investigators and patients may be atypical, recruitment rates low and all participants may receive better care than that generally available. RCT indicate efficacy, not effectiveness in everyday use. We need to know about both.

In fact, the two approaches are complementary: whereas RCT provide an indication of minimum effect, observational studies offer estimates of maximum effect.

Now comes a welcome proposal for reporting meta-analysis of observational studies in epidemiology ('MOOSE')². The MOOSE group define an observational study as an etiological or effectiveness study using data from an existing database, a cross-sectional study, a case series, a case-control design, a design with historical controls or a cohort design. They do not address meta-analysis of pooled data, but only the pooling of published study results.

The MOOSE checklist describes how the background of such a report should be formed. A search strategy is given and common weaknesses described. The strategy should be described in enough detail to allow replication. The methods checklist, based on the group's review of 32 observational studies, indicated that problems in classification, coding and quality were common. It is recommended that results include graphical summaries and any combined estimates, tables of descriptive information for each study, sensitivity testing, subgroup analysis, and indications of statistical uncertainty. The discussion section of a report is to include issues of bias, confounding and quality. The conclusions section should include alternative explanations, appropriate generalisations and describe the further research that is warranted. The MOOSE group made recommendations for determining usefulness and implementation of their recommendations and noted that, 'meta-analyses are themselves observational studies, even when applied to RCT', an interesting puzzle.

The weakness of the MOOSE proposal may lie in its insufficient warnings of the controversial nature and limits of meta-analyses applied to observational studies. To be fair, these are not ignored but merit wider understanding. Egger and colleagues³ and, especially, Shapiro with commentators Petitti and Greenland⁴ have criticised this application. Shapiro states, 'meta-analysis offers the

Holy Grail of attaining statistically stable estimates for effects of low magnitude', and in so doing ignores the limits of epidemiological inference from observational studies. These studies can rarely eliminate or balance all sources of bias and never exclude the possibility of unidentified and uncontrolled confounding factors. The assignment of quality-weighting to studies does not eliminate bias and confounding in meta-analysis of observational studies, in the way that randomisation does in RCT. Simply estimating size effects does not reach truth. These types of meta-analyses are inherently less certain and less definitive and the quality of any meta-analysis cannot transcend that of its individual studies. Further, the validity and reliability of quality rating system needs to be proved, not assumed.

Yet there is a problem in these criticisms. They seem to condemn observational studies as well, and Black argued well for their complementary nature to RCT.

The inappropriate use of a single conclusive summary-estimate leads to premature conclusions. In other words random-effect models should not be used to combine obviously disparate results into meaningless summaries of precise-looking single estimates of weak effects. Broader study inclusion criteria may be considered if the problematic single summary measure is forsaken. Providing rigorous comparisons of studies and identification of patterns of results is a useful outcome. Appropriate comparative meta-analysis of observational studies directs the search for sources of heterogeneity, allows more rigorous comparison of studies and helps define and promote better observational studies in the future. Inappropriate synthetic meta-analysis of observational studies will mislead us, discredit the method and further decrease support for essential and complementary observational studies.

Have we come to expect more of meta-

analysis of observational studies than the method can presently deliver? This is not to infer that it should be abandoned, because the systematic consideration of sources of heterogeneity between observational studies can provide more insight than a biased and misleading overall measure of effect. The York review of fluoridation wisely reported ranges of effects. But it was subtitled as an efficacy study. It inherently dealt with effectiveness. This was misleading. Demonstrating the validity and reliability of quality rating was not addressed by the MOOSE proposal or elsewhere.

Briss et al⁵ have developed a methodology for an evidence-based guide to community preventive services. We now await the release in late 2001 of the chapter on promoting oral health (see <http://www.thecommunityguide.org/guide/oral%20health%20page.htm>). It is to include community preventive strategies for dental caries, oral and pharyngeal cancers, oral and facial injuries and educational interventions. We should not expect simplistic and misleading single recommendations, but insightful progress.

1. Black N. Why we need observational studies to evaluate the effectiveness of health care. *BMJ* 1996; 312:1215-1218.
2. Stroup DF, Berlin JA, Morton SC, Olkin I, Williamson GD, Rennie D, et al. Meta-analysis of observational studies in epidemiology: a proposal for reporting. Meta-analysis of Observational Studies in Epidemiology (MOOSE) group. *JAMA* 2000; 283:2008-2012.
3. Egger M, Schneider M, Davey Smith G. Spurious precision? Meta-analysis of observational studies. *BMJ* 1998; 316:140-144.
4. Shapiro S. Meta-analysis/Shmeta-analysis. *Am J Epidemiol* 1994; 140:771-778.
5. Briss PA, Zaza S, Pappaioanou M, Fielding J, Wright-De Aguerro L, Truman BI, et al. Developing an evidence-based guide to community preventive services - methods. The Task Force on Community Preventive Services. *Am J Prev Med* 2000;18 (suppl):35-43.