Community Dental Health (2007) **24,** 194-197

Editorial

The STROBE initiative and its implications for dental public health research

In recent years, an armoury of guidelines and checklists has been produced to assist researchers in the conduct and reporting of biomedical studies. The information contained in these publications has been of considerable benefit to the scientific community and interested clinicians, and has greatly enhanced the drive towards evidence-based clinical practice.

The material is the outcome of deliberations by groups of leading authorities, closely involved in the various types of experimental, descriptive and review studies covered, and in their reporting and publication. It represents a repository of expertise that only existed fragmentarily hitherto. The intention of the producers of the various documents, and the leading biomedical journals to which many of them are affiliated, is to ensure that published material is the soundest and most reliable available, and to refine and codify the reporting of the advances taking place continually in the biomedical field. The principles put forward, aimed at setting standards and determining courses of action in the conduct and reporting of research, are making a major contribution to systematising investigations into the distribution of diseases and disorders in human populations and their determinants.

Important among the collection of manuals, working documents and guidelines hitherto published are those concerned with the reporting of randomised controlled trials (RCT), exemplified by the CONSORT statement (Altman, 1996); those dealing with the conduct and reporting of meta-analyses, notably the QUORUM statement – Quality of Reporting of Meta-analyses (Clarke, 2000); those covering systematic reviews of RCTs, specifically the guidelines and training manuals (including the QUORUM statement) promulgated by the Cochrane Collaboration, www.cochrane.org, (named after the celebrated medical epidemiologist who inspired its creation); and those advising on the conduct and reporting of systematic reviews of both RCTs and observational studies (NHS Centre for Reviews and Dissemination, 2001). The authors of these documents have an ongoing interest in revisiting and updating them. They come together regularly to review feed back from interested outside parties and to reconsider and modify their published guidance. A revised version of the CONSORT statement for the reporting of parallel-group randomised trials, for example, appeared in 2000 and it is understood a further version is in development.

Lately the list has been joined by STROBE – the statement on Strengthening the Reporting of Observational Studies in Epidemiology. (Elm et al., in press) - which promises to further improve the reporting of clinical research and perhaps broaden the range of material acknowledged as being of intrinsic scientific worth. As its name implies, the outcome of the initiative consists essentially of an inventory and checklist of items that should be addressed in the reporting of analytical epidemiological studies. It represents the first set of guidelines of this type devoted exclusively to the reporting of observational studies as opposed to RCTs or systematic reviews. Taking into account empirical evidence and theoretical considerations, the authors have concentrated on the three main study designs: cohort, case-control and cross-sectional investigations. The group of European and North American methodologists, researchers and editors responsible for developing STROBE initially approached the task by obtaining funding for an inaugural workshop in 2004 and the setting up of a website, www.strobe-statement.org. Their *modus operandi* thereafter involved the searching of textbooks, bibliographic databases, reference lists and personal files for relevant material, including previous recommendations, empirical studies of reporting and articles describing pertinent methodological research. The statement is underpinned by a 69 page back-up document providing explanation and elaboration of the group's recommendations (Vandenbroucke *et al*., 2007) with supporting examples from published studies illustrating good practice in various aspects of the reporting process. The authors noted that much clinical or public health knowledge comes from observational research with about nine of ten papers published in clinical specialty journals being devoted to observational research. However, in their opinion, the current reporting of observational studies is often of insufficient quality. They believe furthermore that poor reporting hampers the assessment of the strengths and weaknesses of a study and the generalisability of its results.

The workshop and the subsequent iterative process of consultation and revision resulted in a checklist of 22 items (the STROBE statement) that relate to the title, abstract, introduction, methods, results and discussion sections of articles. Eighteen items are common to cohort studies, case-control studies and cross-sectional studies and four are specific to each of these study designs. The authors stress that the intention is solely to provide guidance on how to report observational research well. The recommendations are not to be seen as prescriptions for designing or conducting studies, and while clarity of

Correspondence to: Professor Martin Downer, 16A Westbury Park, Westbury Park, Bristol BS6 7JA, UK.
E-mail martindowner@btinternet.com.

reporting is a prerequisite to evaluation, the STROBE checklist is not to be seen as an instrument for evaluating the quality of observational research. Moreover, they state that the recommendations are not prescriptions for setting up or conducting studies, nor do they dictate methodology or mandate a uniform presentation. Nevertheless, in the opinion of the present writer, the very act of tightening up the reporting process should have the effect of rendering the formulation of study protocols and the conduct of investigations more scientifically rigorous from the outset.

It seems that one of the key features of the STROBE initiative and the expansion in the scope of the evidence base for clinical practice that it implies, is the unequivocal acknowledgement that well-conducted cross-sectional studies can make an important and valuable contribution to the sum of our knowledge on the distribution and determinants of health and disease. The concepts of the systematic review, and the checklist system of ensuring that the reporting of clinical studies is comprehensive and transparent, has greatly assisted the movement towards evidence-based clinical practice. STROBE constitutes another major step forward.

The previous instruments, referred to above, have laid stress on the ultimate supremacy of the randomised controlled trial, or on systematic reviews specifically of RCTs, in the hierarchy of scientific evidence. While this position is arguably unassailable it has, nevertheless, led at worst to the virtual discounting of evidence gained from other types of study design - in particular cross-sectional studies - or at best an acceptance that cohort and case-control studies lower down the hierarchy of scientific evidence, can make a contribution, but of lesser value. Thus, for example, potentially useful evidence on the benefit of population strategies for the improvement of the health of communities which are simply not amenable to the RCT approach but which nevertheless are highly effective therapeutic population measures, has been at risk of rejection.

With regard in particular to systematic reviews of RCTs, it is evident that these are by no means infallible. Readers may recall the sharp controversy over an important systematic review of screening for breast cancer with mammography by Olsen and Gotzsche (2001) from the Nordic Cochrane Centre. The authors stated that their review found that 'currently available evidence does not show a survival benefit from mass screening for breast cancer (and the evidence is inconclusive for breast cancer mortality), whereas it has shown that mass screening leads to increased use of aggressive treatment.' These findings were challenged, with the office of the NHS cancer screening programmes in the United Kingdom disputing the association between mammography and treatment (Mayor, 2001). In commenting on what became known as the 'breast cancer screening row' the editor of the *Lancet* took the view that it seemed the Cochrane process, like any other human enterprise, was not perfect (Horton, 2001).

From our own field of oral and dental research there are examples of independent researchers undertaking systematic reviews in the same general subject area but admitting substantially different inventories of research reports to analysis. In two systematic reviews by separate groups of investigators on the longevity of dental restorations, albeit with somewhat different terms of reference, Chadwick *et al*. (2001) included 195 studies in their analysis whereas Downer *et al*. (1999) identified only six longitudinal studies that they considered methodologically adequate. It is known that the findings of a systematic review can be influenced by many factors including the completeness of the search, the relevance criteria applied (level of recall and precision), the correctness of the identification of study methodology and characteristics, and the validity of the inclusion criteria. Jokstad (2002) noted that systematic reviews occasionally end up with disparate conclusions even if they focus on exactly the same topic, the main reason for differences in outcome being the inclusion and exclusion criteria. This is again illustrated by two more recent systematic reviews of the effectiveness of oral cancer screening. Kujan *et al*. (2005), in a review conducted according to Cochrane Collaboration standards, examined 100 citations but identified only one study (Ramadas *et al*., 2003) as conforming with the guidelines' recommended stringent inclusion criteria. In another review on the same topic, conducted as far as practicable within guidelines promulgated by the NHS Centre for Reviews and Dissemination (2001), Downer *et al*. (2006) used more liberal inclusion criteria based on markers of screening effectiveness identified by Chamberlain (1993) as being necessary – though not of themselves sufficient – for proving the effectiveness of screening. In this latter review, agreement was reached for the inclusion of 28 studies. Which of these reviews contributed the greater amount to our understanding of the issues surrounding the evaluation of an oral cancer screening programme, and to judging the likelihood of such a programme being effective, is a matter for debate.

There is little doubt that an element of subjectivity can enter even the most rigorous systematic review process and the same, of course, applies to both experimental and observational studies, the brickwork from which systematic reviews are constructed. Thus reports of clinical trials, which *prima facie* have been conducted meticulously and have ticked the majority of boxes in, for example, the CONSORT checklist, may still be of doubtful provenance. Unless the critical reviewer was actually present on site or in the field to observe what really took place during the course of a study, it is ultimately impossible to pronounce unequivocally on its validity and reliability. There are notable instances of fudged and concocted results being uncovered retrospectively in research reports that were accepted at the time by the scientific community as being *bona fide*. The incidence of scientific fraud should not be under-estimated.

A zealous reliance on prescribed methods carries the risk that misrepresentation and downgrading of the important role which a public health strategy can have in promoting disease prevention and control, may occur. Instances can be cited, again from our own field of oral and dental research, where a rigid imposition of possibly unachievable standards in the conduct and reporting of studies has cast doubt on, or threatened to deny, the benefit of a valuable therapeutic measure to the community. The systematic review of public water fluoridation carried out under the auspices of the University of

York (McDonagh *et al*., 2000) considered three levels of evidence and rejected a third. Level A (the highest level) included prospective studies started within a year of the initiation of fluoridation; randomised studies or studies addressing at least three possible confounding factors; and studies where the investigators were blinded to the place of residence of the participants. Level B were studies with moderate risk of bias, including those started within three years of the initiation of fluoridation; those that adjusted for at least one confounding factor; and non-blinded studies where other provision was made to prevent measurement bias. Level C, the lowest (discounted) level of evidence, consisted *inter alia* of prospective or retrospective, and cross-sectional observational studies – including the very types of design now acknowledged within the STROBE initiative. With respect to effectiveness of fluoridation in preventing dental caries, a total of 26 level B (but no level A) studies were found. The authors stated that a large number of studies were excluded because they were *cross-sectional and did not meet the inclusion criteria* (my italics). On the basis of the reports of the 26 disparate, clinically heterogeneous investigations from the United Kingdom, Germany, the Netherlands, North and South America, and Taiwan that were admitted, the authors concluded 'with extreme caution' that the pooled estimate of the risk difference between the fluoridated and non-fluoridated control localities of the change in the proportion of caries-free children was 15 % (it is not entirely clear whether this somewhat convoluted sentence implies 15% or 15 percentage points) with a mean difference in dmft/DMFT of 2.3. These median values were derived from children of different ages living in vastly different localities, using combined data on the primary and permanent dentitions. Logically these estimates (or any other estimates which are simple constants) could not apply to populations where, for example, the baseline mean dmft (or DMFT) is less than 2.3 or (we assume) where 85% of children are caries-free.

On the other hand, by accepting high quality level C evidence and using, as a good example, the cross-sectional data generated by the British Association for the Study of Community Dentistry (BASCD) epidemiological survey programme, arguably more reliable and certainly more pragmatic demonstrations of the effectiveness of fluoridation in English populations could have been achieved. The BASCD epidemiology programme is of high quality. As readers will know, it employs large population samples that are examined by trained and calibrated clinicians, and verified every 10 years against national population samples. Utilising linear regression analysis of these cross-sectional survey data it is possible to compare age-specific caries prevalence and severity at local health authority (PCT) level according to fluoridated and non-fluoridated status of their water supplies, and socio-economic status of their populations. The overall benefit of fluoridation and the greater benefit accruing to the more deprived areas can be demonstrated graphically and quantitatively, beyond reasonable doubt. (British Fluoridation Society, 2006; Jones *et al*., 1997; Riley *et al*., 1999).

There is an ever present hazard with the evidence-based approach of metaphorically 'throwing the baby out with the bathwater'. Applied over-zealously it can give the reader of, for instance, the report of a systematic review the impression that there is no evidence. A systematic review should surely be about presenting the best *available* evidence? One particular database publishes critical commentaries on published systematic reviews on its website and acts as a kind of guardian of the purity of the evidence base. Its reviewers are liable to downgrade a review, albeit carried out conscientiously and methodically, and subsequently published in a peer-reviewed journal with a high impact factor, typically as suffering from flaws in various aspects of the methodology and presentation. This begs the question, is it therefore better when confronted with a whittled down handful of less than perfect research reports to simply say 'there is no available evidence' and produce nothing?

Authors have a natural tendency, confronted with a negative commentary on their work, to mutter - with a degree of understandable bitterness - that it is all very well for some person to sit on the sidelines demolishing the credibility of the work, whilst knowing that the same reviewer would find it a great deal harder to take on the task of trying to make a better fist of the same challenge. However, these sorts of carping complaints about critics have probably been voiced for as long as the written word, the pictorial image and the performance arts, let alone the scientific article, have existed. It is to be hoped that the STROBE initiative will at least contribute to widening the spectrum of what evidence the scientific consensus regards as acceptable and, like CONSORT, guide researchers in its presentation. In this and many other respects, STROBE is a most important development.

Martin Downer,
Honorary Professor, UCL Eastman Dental Institute and
Honorary Professor, University Dental School of Manchester,

References

Altman, D.G. (1996): Better reporting of randomized controlled trials: the CONSORT statement. *British Medical Journal* **313**, 570-571.

British Fluoridation Society. (2006): *Reducing inequalities in children's dental health*. http://www.bfsweb.org/Briefing_on_inequalities_2006.pdf.

Chadwick, B.L., Dummer, P.M.H., Dunstan, F., Gilmour, A. Jones, R. Phillips, C. *et al*. (2001): *A systematic review of the longevity of dental restorations*. Report 19. York: NHS Centre for Reviews and Dissemination, University of York.

Chamberlain, J. (1993): Evaluation of screening for cancer. *Community Dental Health* **10 (Supplement 1)**, 5-11.

Clarke, M. (2000): The QUORUM statement. *Lancet* **355**, 756-757.

Downer, M.C., Azli, N.A., Bedi, R., Moles D.R. and Setchell, D.J. (1999): How long do routine dental restorations last? A systematic review. *British Dental Journal* **187**, 432-439.

Downer, M.C., Moles, D.R., Palmer, S. and Speight P.M. (2006): A systematic review of measures of effectiveness in screening for oral cancer and precancer. *Oral Oncology* **42**, 551-560.

Elm, E., von., Altman D.G., Egger, M., Pocock, S.J., Gotzsche, P.C. and Vandenbroucke, J.P. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *British Medical Journal* (in press).

Horton, R. (2001): Screening mammography – an overview revisited. *Lancet* **358**, 1284-1285.

Jokstad, A. (2002): How long do fillings last? *Evidence-Based Dentistry* **3**, 96-99.

Jones, C.M., Taylor, G.O., Whittle, J.G., Evans, D. and Trotter, D.P. (1997): Water fluoridation, tooth decay in 5-year-olds, and social deprivation measured by Jarman score: Analysis of data from British dental surveys. *British Medical Journal* **315**, 514-517.

Kujan, O., Glenny, A.M., Duxbury, A.J., Thakker, N., Sloan, P. (2005): Screening programmes for the early detection and prevention of oral cancer (Review). The *Cochrane Library*, **Issue 2**:1-17.

Mayor, S. (2001): Row over breast cancer screening shows that scientists bring 'some subjectivity into their work'. *British Medical Journal* **323**, 956.

McDonagh, M., Whiting, P., Bradley, M., Cooper, J., Sutton, A., Chestnutt, I., *et al*. (2000): *A systematic review of public water fluoridation*. York: NHS Centre for Reviews and Dissemination, University of York.

NHS Centre for Reviews and Dissemination. (2001): *Undertaking systematic reviews of research on effectiveness*. **4**. University of York: York Publishing Services.

Olsen, O. and Gotzsche, P.C. (2001): Cochrane review on screening for breast cancer with mammography. *Lancet* **358**, 1340-1342.

Ramadas, K., Sankaranarayanan, R., Jacob, B.J., Thomas, G., Somanathan, T., Mahe C. *et al*. (2003): Interim results from a cluster randomized oral cancer screening trial in Kerala, India. *Oral Oncology* **39**, 580-588.

Riley, J., Lennon, M. and Ellwood, R. (1999): The effect of water fluoridation and social inequalities on dental caries in 5-year-old children. *International Journal of Epidemiology* **28**, 300-305.

Vandenbroucke, J.P., Elm, E. von., Altman, D.G., Gotzsche, P.C., Mulrow, C.D., Pocock, S.J. *et al*: (2007): *Strengthening the Reporting of Observational Studies in Epidemiology (STROBE): Explanation and elaboration*. Bern: Institute for Social and Preventive Medicine (ISPM). strobe@ispm.unibe.ch.