Eur J Cancer, Vol. 27, No. 5, pp. 525–526, 1991.
Printed in Great Britain
0277–5379/91 $3.00 + 0.00
Pergamon Press plc

The Value of Consensus

DURING the past decade consensus conferences on health issues have been a major growth area. The concept was first popularised in the USA [1], where the National Institutes of Health have organised getting on towards 100 consensus conferences, and is now firmly entrenched in Britain and other Western Countries.

What are the aims and aspirations of such conferences, how do they work and are they worthwhile? Skrabanek [2] has suggested that "uncertainty in medicine, as in theology, is intolerable and a consensus conference like a synod of bishops, is convoked to settle the matter". Put this way the reasons for consensus conferences and their potential usefulness are easy to understand. If Skrabanek is right in his choice of metaphor, as I believe he is, then such devices are destined to fail. Religion relies on faith and this is why convocations are needed—though it is striking how, despite this, every major religion seems to undergo schism as bishops fail to arrive at universally accepted views. Similarly, consensus conferences are generally built on opinion (faith) and not fact. As Dr Arnold Relman [3], one time editor of the *NEJM* observed, "an assessment of current knowledge, no matter how sophisticated and rigorous, cannot go beyond that knowledge and rarely generates new information. What was unknown before the review remains unknown afterwards." Consensus conferences rely on experts melding their views to reach a medical formula said to express current truths. However, choice of experts may be carefully controlled, the data presented is often highly selected and the "organisers" may have a particular goal that they wish to achieve.

Most importantly, vital data are usually not available. The very need to find a consensus surely suggests that there is no persuasive data to tell us about optimal treatments. Consensus conferences are thus likely to pontificate about the unknown and then with due ceremony and whilst blindfolded try to affix the tail onto the donkey. One way around this approach is to conduct a systematic overview (meta-analysis) of all published and unpublished randomised trials relevant to the questions being asked. In the consensus conferences on adjuvant therapy of breast cancer (National Institutes of Health, Washington 1985; Kings Fund London 1986) the most impressive data supporting adjuvant therapy came from Richard Peto's group who performed such a meta-analysis [4, 5]. These data showed a small but highly significant benefit from adjuvant therapy. Since the differences were small, inclusion of all trials was very important as publication bias may result in spurious treatment differences when only published trials are included [6]. Clearly the collection of *individual* patient data from *all* published and unpublished trials is a major undertaking, but if we are to have a reliable result with narrow confidence limits such an approach is essential. Only then can clinicians judge for themselves whether the treatment being assessed is worthwhile for their patients. One weakness of overviews is that it is not possible to be sure that all unpublished studies have been recognised, though the omission of small studies in a large patient population may not be particularly damaging.

Received 21 Jan. 1991; accepted 4 Feb. 1991.

However, despite the inclusion of overview data in the Breast Consensus conferences, they did not make an immediate impact on British practice, perhaps because these data were devalued by failure to publish a peer review paper [7] for several years. Gazet and his colleagues [8] and others have found that a considerable number of premenopausal British women are not offered adjuvant chemotherapy even though all consensus conferences on the topic have recommended such therapy. One reason for this state of affairs may be reluctance to accept the method and findings of overviews which combine the results of many individual trials. Far more persuasive would be a positive result from a really large believable trial. The use of fibrinolytics after infarction is an excellent example of how one trial can really sway medical opinion. Fibrinolytics have been tested in patients undergoing myocardial infarction for over 30 years. Following a series of small inconclusive studies the general medical consensus was that this approach did not improve survival—indeed many physicians thought it positively dangerous though a formal overview did not support this, showing instead a survival benefit for treated patients [9]. The ISIS-2 trial, however, conclusively showed that a short course of therapy caused a relatively small but highly significant reduction in deaths in the first month after infarction [10]. Within a few years this one large (more than 17 000 patients) trial has profoundly changed medical attitudes. No consensus conference was necessary—the data spoke for itself. Similarly, penicillin did not require a consensus for physicians to be persuaded to adopt it for the treatment of bacterial infections; indeed randomised trials were not even needed. Such dramatic treatment differences as those achieved by the early antibiotics are unfortunately rare. Most new treatments produce *modest* benefits at best and current clinical trials are an order of magnitude too small to reliably demonstrate treatment effectiveness.

Rather than asking consensus conferences to gaze into the crystal ball of incomplete data from inadequate trials, we need careful assessment of hard facts. Meta-analysis of the type described above gives the best available answer from trials run to data and may suggest appropriate questions for new large trials. For instance, the overview of chemotherapy in advanced ovarian carcinoma has spawned two international cancer trials [11]. This is important since the best solution to the uncertainty Skrabanek refers to above is clear and unequivocal data that is applicable to real patients.

Indeed, this very uncertainty is the cornerstone of current thinking on large scale trials. Rather than strictly defining entry criteria and selection of patients, trials based on the *uncertainty principle* chose patients by whether the clinician is substantially uncertain if an *individual* patient will benefit from the treatment being tested. Patients who the clinician is fairly certain will or will not benefit are excluded. This freedom of individual choice allows the trial to reflect what is likely to happen in the community if the new treatment is introduced. In contrast, consensus conference recommendations are blanket prescriptions which often fail to allow the individual physician to exercise his judgement and which threaten to become restrictive and to be used as a measure of competence [3, 12].

Thus, where consensus conferences produce restrictive *opinion* appropriate large scale clinical trials (and to a lesser extent, overviews) produce reliable data which allow the clinicians to decide what is best for the individual patient. Similarly, where Skrabanek deplores the tendency for inflated claims (Jumbo jet statistics) of enthusiasts at consensus conferences designed to start a bandwagon rolling [2], large scale clinical trials can define

treatment efficacy so that health economists or other decision makers can decide if the benefit is worth the cost to the community as well as the individual patient.

Consensus conferences are becoming the opium of the physician and they may be a dangerous tool—it would have taken a brave person to bet that between 1900 and 1980 *any* consensus conference would have recommended wide local excision and radiotherapy as the treatment of choice for most women with breast cancer. Our present acceptance of this approach is founded on the results of moderately large trials and it would be an ill day if we allow consensus conferences to replace the need for even larger clinical studies.

C. Williams
CRC Wessex Regional Medical Oncology Unit
CF99 Southampton General Hospital
Southampton SO9 4XY, U.K.

1. Perry S, Kalberer JT. The NIH Consensus–Development Program and the assessment of health-care technologies the first two years. *N Engl J Med* 1980, **303**, 169–172.
2. Skrabanek P. Nonsensus consensus. *Lancet* 1990, **i**, 1446–1447.
3. Relman AS. Assessment of medical practices: a simple proposal. *N Engl J Med* 1980, **303**, 153–154.
4. Lipman ME. National Institutes of Health Consensus development conference on adjuvant chemotherapy and endocrine therapy for breast cancer. NCI Monographs 1985, 1–5.
5. Anonymous. Consensus development conference: treatment of primary breast cancer. *Br Med J* 1986, **293**, 946–947.
6. Begg CB, Berlin JA. Publication bias and dissemination of clinical research. *J Natl Cancer Inst* 1989, **81**, 107–115.
7. Early Breast Cancer Trialists' Collaborative Group. Effects of adjuvant tamoxifen and cytotoxic therapy on mortality in early breast cancer: an overview of 61 randomized trials among 28,896 women. *N Engl J Med* **319**, 1681–1692.
8. Gazet J-C, Rainsbury RM, Ford HT, Powles TJ, Coombes RC. Survey of treatment of primary breast cancer in Great Britain. *Br Med J* 1985, **290**, 1793–1795.
9. Yusuf S, Collins R, Peto R, *et al.* Intravenous and intracoronary fibrinolytic therapy in acute myocardial infarction: overview of results on mortality, reinfarction and side effects from 33 randomized controlled trials. *Eur Heart J* 1985, **6**, 556–585.
10. ISIS-2 (Second International Study of Infarct Survival) Collaborative Group. Randomised trials of intravenous streptokinase, oral aspirin, both, or neither among 17,187 cases of suspected myocardial infarction. *Lancet* 1988, **1**, 349–360.
11. Guthrie D, Parmar M, Stewart L, Williams CJ. An overview in the treatment of advanced ovarian cancer. *Br J Cancer* 1990, **61**, 495–496.
12. Farrington JF, Felch WC, Hare RL. Quality assessment and quality assurance: the performance-review alternative. *N Engl J Med* 1980, **303**, 154–156.

Eur J Cancer, Vol. 27, No. 5, pp. 526–528, 1991.
Printed in Great Britain
0277–5379/91 $3.00 + 0.00

Magnetic Resonance Spectroscopy as a Probe of Tumour Metabolism

As a result of parallel developments in imaging and spectroscopy, nuclear magnetic resonance (NMR) is now extensively used in diagnostic radiology and in biomedical research. Magnetic resonance imaging (MRI), based primarily on the detection of signals from water and fats, has become an established radiological technique, with over 3000 systems currently installed worldwide. Magnetic resonance spectroscopy (MRS) provides a method of studying metabolism, with an emphasis on research applications rather than diagnosis (with one or two notable exceptions that are mentioned below). Here, we are concerned with the possible roles of MRS in oncology.

MRS can be used to prove a wide variety of systems, ranging from body fluids, tissue extracts and cell cultures to non-invasive studies of tissue metabolites in man, which all have a role to play in relation to our understanding of cancer. One of the more controversial topics arose out of the proposal that water-suppressed proton NMR spectroscopy of plasma might provide an approach to the detection of cancer and the monitoring of therapy [1]. This proposal, based on the linewidth properties of the resonances from plasma lipoprotein lipids, has provoked a great deal of further work and correspondence, the strong consensus of opinion being that there are far too many false positives and false negatives for the linewidth measurement to provide a useful test for cancer [2].

It has also been reported that ^{1}H MRS can distinguish between normal and malignant tissue by the detection of neutral lipids in or attached to the membrane protein [3]. On the basis of these and subsequent observations [4], further analysis of this crowded region of the ^{1}H spectrum certainly appears to be justified, but perhaps the emphasis should be not so much on a test for cancer but on understanding in more detail the biochemical abnormalities that are associated with malignant disease. It should be stressed that NMR is not a sensitive technique—relatively high concentrations (typically 0.1–1 mmol/l or greater) are required in order to produce a detectable signal—so that if there is a marker for cancer, techniques that are more sensitive than NMR are likely to be more appropriate.

The prime advantage of MRS is its ability to probe tissue metabolites non-invasively *in vivo*. ^{31}P MRS permits the study of energy metabolism through the signals from ATP, phosphocreatine (PCr) and inorganic phosphate (P_i). Since the frequency of the P_i signal is sensitive to pH, information is available not only about the relative concentrations of these metabolites but also about intracellular pH (the assumption being that most of the P_i is intracellular). In view of the enhanced lactate production that is commonly associated with tumour cells, an interesting observation that has emerged from both animal models and

Received 28 Jan. 1991; accepted 4 Feb. 1991.