



Kathryn Fitch



Pablo Lázaro

# Using appropriateness criteria to improve health care

*“only around 15 per cent of medical decisions are based on scientific evidence about their outcomes”*

The basic mission of health services is to improve the health of both the individual and society. As is well known, however, health services in recent decades have been faced with enormous challenges, making it ever more costly to fulfill this mission. These include the introduction of increasingly complex services, the rapid innovation and diffusion of medical technologies and procedures, and pressures on the demand for services from both patients and health professionals. These phenomena are part of the reason health spending has grown so quickly in the industrialised nations over recent decades. In 1965 European Union countries spent on average 4.3 per cent of their gross domestic product on health care, a proportion that rose to 6.3 per cent in 1975, 7.0 per cent in 1985 and 7.7 per cent in 1995.<sup>1</sup> Clearly, this proportion cannot continue to grow indefinitely.

What is driving this growth in health care expenditures? Studies have shown that it is mostly due to increases in the “volume and

intensity of services.”<sup>2</sup> To curtail the volume of services without negative effects on the health status of the population, we need to find ways to assure that our health expenditures are used for effective services – that is, those that have demonstrated value. Yet it has been estimated that only around 15 per cent of medical decisions are based on scientific evidence about their outcomes.<sup>3</sup> If this proportion is even approximately correct, it is not surprising that such wide variations have consistently been shown to exist in the rates of use of medical procedures – differences that cannot be explained by patient characteristics. For example, a graph of carotid endarterectomy rates plotted as a function of the number of surgeons in three areas of the United States (Figure 1) shows that the site with the highest number of surgeons performed more than three times the number of surgical interventions as the site with the lowest number.<sup>4</sup> Does this mean that the area with more surgeons is performing procedures that are inappropriate, or that the one with a lower number is not completely meeting the needs of the population? Based on these statistics alone, we cannot know. It may even be that both overuse and under-use are occurring simultaneously in any or all of the areas.

## The RAND appropriateness method

In an attempt to answer these kinds of questions, researchers from RAND and the University of California in Los Angeles developed in the mid 1980s what has come to be called the ‘RAND appropriateness method’. The concept of appropriateness, in the RAND method, refers to the relative weight of the benefits and harms of a medical intervention. An appropriate procedure is one in which the expected benefits outweigh the expected risks by a sufficient margin that the procedure is worth doing. The rationale behind the method is that randomised clinical trials – the ‘gold standard’ for evidence-based medicine – often either are not available or cannot provide evidence at a level of detail sufficient to

**Figure 1: Number of carotid endarterectomies by number of surgeons in three US sites**



Source: Leape LL, et al. 1989.<sup>4</sup>

apply to the wide range of patients seen in everyday clinical practice. Although robust scientific evidence about the benefits of many procedures is lacking, physicians must nonetheless make decisions every day about when to apply them. Consequently, it was believed a method was needed that would combine the best available scientific evidence with the collective judgment of experts to yield a statement regarding the appropriateness of performing a procedure at the level of patient-specific symptoms, medical history and test results.

The basic steps in the RAND method are described in the accompanying article by Kahan and van het Loo in this issue. The final product of the two-round 'modified Delphi' process is a list of highly specific clinical scenarios or 'indications', each of which is classified as 'appropriate', 'uncertain' or 'inappropriate' for the procedure in question based on the median panel rating and the level of agreement among panelists. This set of indications – which may number from hundreds to thousands – with the corresponding appropriateness ratings constitutes what are called the 'appropriateness criteria'.

### Using appropriateness criteria to measure performance

Appropriateness criteria have most often been used as a tool to measure performance retrospectively. This is done by reviewing the medical charts of a representative sample of patients who have undergone the procedure. A specially developed 'abstraction form' is used to collect sufficient data on each patient to permit assignment of an appropriateness rating in accordance with the list of indications. The proportion of patients who have received procedures done for 'appropriate', 'uncertain' and 'inappropriate' reasons can then be calculated. Procedures done for inappropriate reasons are considered *overuse* of the procedure. Early studies in the United States showed that a substantial number of procedures were judged to be inappropriate: 17 per cent of coronary angiographies, 32 per cent of carotid endarterectomies, 17 per cent of

upper GI endoscopies,<sup>5</sup> and 16 per cent of hysterectomies.<sup>6</sup>

Of particular interest in these U.S. studies was the finding that the volume of procedures was generally not related with levels of appropriateness. That is, areas where relatively few procedures were performed did not necessarily have lower rates of inappropriate use than those where intensity of use was higher. For example, the proportion of inappropriate use of coronary angiography was approximately the same in three U.S. areas (ranging from 15 to 18 per cent), even though the rate of utilisation of the procedure in one area was more than double that of the other two.<sup>5</sup> In one area of the United Kingdom – a country where physicians perform only about one-seventh the number of cardiac procedures as in the United States – 21 per cent of coronary angiographies were found to be performed for inappropriate reasons, in accordance with the criteria developed by a U.K. panel.<sup>7</sup> These figures suggest that just reducing the number of proce-

*“In one area of the United Kingdom ... 21 per cent of coronary angiographies were found to be performed for inappropriate reasons”*

dures performed will not necessarily reduce the rate of inappropriate use.

The RAND method can also be applied to measure the possible *underuse* of procedures. Some panels carry out a third round of ratings to determine which of the appropriate indications are also necessary. Necessity is a more stringent criterion than appropriateness and refers to procedures which must be offered to a patient fitting a particular clinical description. Necessity can be more difficult to measure than appropriateness, however, because it involves identifying a group of patients who might have benefited from the procedure, but did not receive it. For example, to measure the underuse of coronary revascularisation, data could be collected from the medical charts of patients who received coronary angiography to determine what pro-

portion of them did not receive a revascularisation procedure that the panel considered necessary for their clinical situation.

### Using appropriateness criteria as clinical decision aids

Appropriateness criteria can also be used prospectively, as the basis for developing different types of aids to clinical decision making. These might be in the form of guidelines or flowcharts, which summarise the criteria in a more 'user-friendly' way than the tables of indications with their appropriateness classifications. Such summary formats can be time consuming to develop, however, and may run the risk of losing the specificity provided by the complete indications list. The tables themselves can also be disseminated in different ways – through publication in a medical journal, distribution of a special report by the relevant medical society, for example – so that physicians can consult the recommendation of the expert panel when confronted with a particular

patient. The challenge is to make the criteria available to physicians in an easy-to-use format, while also devising a system to ensure that the appropriateness classifications are updated as new scientific evidence becomes available.

One possible solution being tested by the Swiss members of the concerted action group is to make the criteria available through a web-based page on the Internet. This has the advantage of allowing physicians to view the criteria in whatever way they prefer: in some cases they may only want to see the appropriateness classification, while in others they may be interested to see the complete panel ratings for a particular indication. Using an Internet-based system, it would also be possible to provide hyperlinks to the relevant publications supporting a particular appropriateness classification, and to

quickly modify criteria that have become outdated.

### Gaining physician acceptance

Enlisting the support of the relevant medical societies early on is an important step in obtaining physician acceptance of the appropriateness criteria. Specialist societies are usually asked to provide nominations of panel members and may sponsor dissemination of the final ratings. Physicians need to be assured that the criteria are not dogmatic rules to be followed reflexively, but rather carefully considered recommendations that will usually apply to a patient fitting the clinical indication, in the absence of other unusual circumstances. If they do not agree with the recommendation for a particular case, they may be asked to justify why it constitutes an exception. Providing feedback to physicians on their own performance as measured by the appropriateness criteria can be a helpful way of motivating them to consider the appropriateness criteria when making clinical decisions.

### Conclusions

Wide variations exist in clinical practice, and a substantial proportion of health interventions are thought to be performed for inappropriate reasons. Bureaucratic, administrative or economic solutions to rising costs may limit the quantity of health care provided, but will not necessarily improve the appropriateness and quality of care. The selective elimination of inappropriate care would free resources to deliver effective care to those who need it. One way to do this is by developing high quality, flexible appropriateness criteria, which can be used both to measure past performance and to guide clinical decision making.

### REFERENCES

1. OECD Health Data 97. *A Software for the Comparative Analysis of 29 Health Systems*. Paris: OECD, 1997.
2. Eddy DM. Broadening the responsibilities of practitioners. The team approach. *Journal of the American Medical Association* 1993;269:1849–55.
3. Black N. Research, audit and education. *British Medical Journal* 1992;304:698–700.
4. Leape LL, Park RE, Solomon DH, Chassin MR, Koseoff J, Brook RH. Relation between surgeons' practice volumes and geographic variation in the rate of carotid endarterectomy. *New England Journal of Medicine* 1989;321:653–57.
5. Chassin MR, Koseoff J, Park RE, Winslow CM, Kahn KL, Merrick NJ, Keesey J, Fink A, Solomon DH, Brook RH. Does inappropriate use explain geographic variations in the use of health care services? A study of three procedures. *Journal of the American Medical Association* 1987; 258:2533–37.
6. Bernstein SJ, McGlynn EA, Sui AL, Roth CP, Sherwood MJ, Keesey JW, Koseoff J, Hicks NR, Brook RH. The appropriateness of hysterectomy. *Journal of the American Medical Association* 1993;269:2398–402.
7. Gray D, Hampton JR, Bernstein SJ, Koseoff J, Brook RH. Audit of coronary angiography and bypass surgery. *The Lancet* 1990;335:1317–20.