Why the impact factor of journals should not be used for evaluating research

Per O Seglen

Institute for Studies in Research and Higher Education (NIFU), Hegdehaugsveien 31, N-0352 Oslo, Norway
Per O Seglen, *professor*

BMJ 1997;314:498–502

Evaluating scientific quality is a notoriously difficult problem which has no standard solution. Ideally, published scientific results should be scrutinised by true experts in the field and given scores for quality and quantity according to established rules. In practice, however, what is called peer review is usually performed by committees with general competence rather than with the specialist's insight that is needed to assess primary research data. Committees tend, therefore, to resort to secondary criteria like crude publication counts, journal prestige, the reputation of authors and institutions, and estimated importance and relevance of the research field,[1] making peer review as much of a lottery as of a rational process.[2 3]

On this background, it is hardly surprising that alternative methods for evaluating research are being sought, such as citation rates and journal impact factors, which seem to be quantitative and objective indicators directly related to published science. The citation data are obtained from a database produced by the Institute for Scientific Information (ISI) in Philadelphia, which continuously records scientific citations as represented by the reference lists of articles from a large number of the world's scientific journals. The references are rearranged in the database to show how many times each publication has been cited within a certain period, and by whom, and the results are published as the *Science Citation Index (SCI)*. On the basis of the *Science Citation Index* and authors' publication lists, the annual citation rate of papers by a scientific author or research group can thus be calculated. Similarly, the citation rate of a scientific journal—known as the journal impact factor—can be calculated as the mean citation rate of all the articles contained in the journal.[4] Journal impact factors, which are published annually in *SCI Journal Citation Reports*, are widely regarded as a quality ranking for journals and used extensively by leading journals in their advertising.

Since journal impact factors are so readily available, it has been tempting to use them for evaluating individual scientists or research groups. On the assumption that the journal is representative of its articles, the journal impact factors of an author's articles can simply be added up to obtain an apparently objective and quantitative measure of the author's scientific achievement. In Italy, the use of journal impact factors was recently advocated to remedy the purported subjectivity and bias in appointments to higher academic positions.[5] In the Nordic countries, journal impact factors have, on occasion, been used in the evaluation of individuals as well as of institutions and have been proposed, or actually used, as one of the premises for allocation of university resources and positions.[1 6 7] Resource allocation based on impact factors has also been reported from Canada[8] and Hungary[9] and, colloquially, from several other countries. The increasing awareness of journal impact factors, and the possibility of their use in evaluation, is already changing scientists' publication behaviour towards publishing in journals with maximum impact,[9 10] often at the expense of specialist journals that might actually be more appropriate vehicles for the research in question.

Given the increasing use of journal impact factors—as well as the (less explicit) use of journal prestige—in research evaluation, a critical examination of this indicator seems necessary (see box).

Summary points

- Use of journal impact factors conceals the difference in article citation rates (articles in the most cited half of articles in a journal are cited 10 times as often as the least cited half)
- Journals' impact factors are determined by technicalities unrelated to the scientific quality of their articles
- Journal impact factors depend on the research field: high impact factors are likely in journals covering large areas of basic research with a rapidly expanding but short lived literature that use many references per article
- Article citation rates determine the journal impact factor, not vice versa

Is the journal impact factor really representative of the individual journal articles?

Relation of journal impact factor and citation rate of article

For the journal's impact factor to be reasonably representative of its articles, the citation rate of individual articles in the journal should show a narrow distribution, preferably a Gaussian distribution, around

Problems associated with the use of journal impact factors

- Journal impact factors are not statistically representative of individual journal articles
- Journal impact factors correlate poorly with actual citations of individual articles
- Authors use many criteria other than impact when submitting to journals
- Citations to "non-citable" items are erroneously included in the database
- Self citations are not corrected for
- Review articles are heavily cited and inflate the impact factor of journals
- Long articles collect many citations and give high journal impact factors
- Short publication lag allows many short term journal self citations and gives a high journal impact factor
- Citations in the national language of the journal are preferred by the journal's authors
- Selective journal self citation: articles tend to preferentially cite other articles in the same journal
- Coverage of the database is not complete
- Books are not included in the database as a source for citations
- Database has an English language bias
- Database is dominated by American publications
- Journal set in database may vary from year to year
- Impact factor is a function of the number of references per article in the research field
- Research fields with literature that rapidly becomes obsolete are favoured
- Impact factor depends on dynamics (expansion or contraction) of the research field
- Small research fields tend to lack journals with high impact
- Relations between fields (clinical *v* basic research, for example) strongly determine the journal impact factor
- Citation rate of article determines journal impact, but not vice versa

the mean value (the journal's impact factor). Figure 1 shows that this is far from being the case: three different biochemical journals all showed skewed distributions of articles' citation rates, with only a few articles anywhere near the population mean.[11]

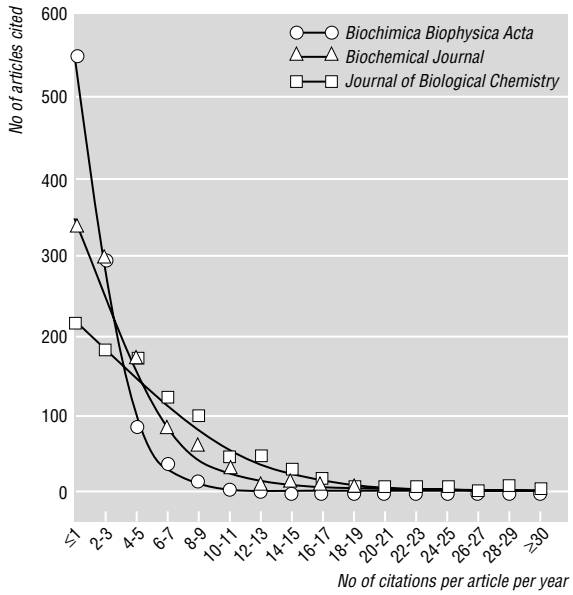


Fig 1 Citation rates in 1986 or 1987 of articles published in three biochemical journals in 1983 or 1984, respectively[11]

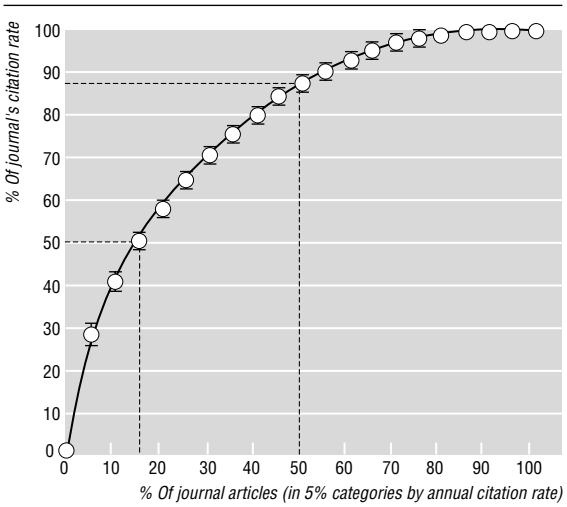


Fig 2 Cumulative contribution of articles with different citation rates (beginning with most cited 5%) to total journal impact. Values are mean (SE) of journals in fig 1; dotted lines indicate contributions of 15% and 50% most cited articles[11]

The uneven contribution of the various articles to the journal impact is further illustrated in figure 2: the cumulative curve shows that the most cited 15% of the articles account for 50% of the citations, and the most cited 50% of the articles account for 90% of the citations. In other words, the most cited half of the articles are cited, on average, 10 times as often as the least cited half. Assigning the same score (the journal impact factor) to all articles masks this tremendous difference—which is the exact opposite of what an evaluation is meant to achieve. Even the uncited articles are then given full credit for the impact of the few highly cited articles that predominantly determine the value of the journal impact factor.

Since any large, random sample of journal articles will correlate well with the corresponding average of journal impact factors,[12] the impact factors may seem reasonably representative after all. However, the correlation between journal impact and actual citation rate of articles from individual scientists or research groups is often poor[9 12] (fig 3). Clearly, scientific authors do not necessarily publish their most citable work in journals of the highest impact, nor do their articles necessarily match the impact of the journals they appear in. Although some authors may take journals' impact factors into consideration when submitting an article, other factors are (or at least were) equally or more important, such as the journal's subject area and its relevance to the author's specialty, the fairness and rapidity of the editorial process, the probability of acceptance, publication lag, and publication cost (page charges).[13]

Journal impact factors are representative only when the evaluated research is absolutely average (relative to the journals used), a premise which really makes any evaluation superfluous. In actual practice, however, even samples as large as a nation's scientific output are far from being random and representative of the journals they have been published in: for example, during the period 1989-93, articles on general medicine in Turkey would have had an expected citation rate of 1.3 (relative to the world average) on the

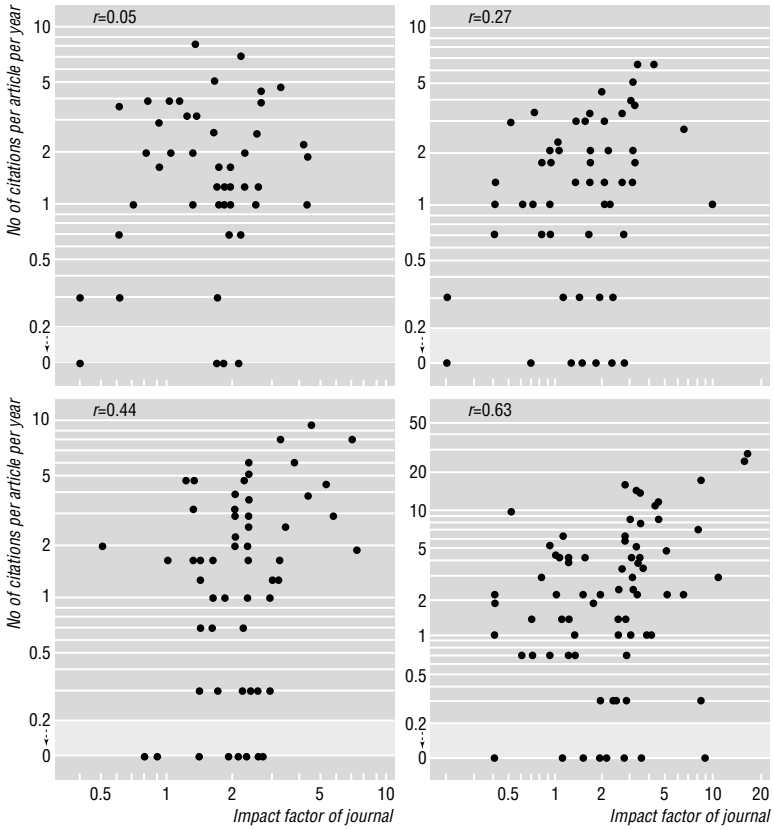


Fig 3 Correlation between article citation rate and journal impact for four authors[12]

basis of journal impact, but the actual citation was only 0.3.[14] The use of journal impact factors can therefore be as misleading for countries as for individuals.

Journal impact factors are calculated in a way that causes bias

Apart from being non-representative, the journal impact factor is encumbered with several shortcomings of a technical and more fundamental nature. The factor is generally defined as the recorded number of citations within a certain year (for example, 1996) to the items published in the journal during the two preceding years (1995 and 1994), divided by the number of such items (this would be the equivalent of the average citation rate of an item during the first and second calendar year after the year of publication). However, the Science Citation Index database includes only normal articles, notes, and reviews in the denominator as citable items, but records citations to all types of documents (editorials, letters, meeting abstracts, etc) in the numerator; citations to translated journal versions are even listed twice.[15-17] Because of this flawed computation, a journal that includes meeting reports, interesting editorials, and a lively correspondence section can have its impact factor greatly inflated relative to journals that lack such items. Editors who want to raise the impact of their journals should make frequent reference to their previous editorials, since the database makes no correction for self citations. The inclusion of review articles, which generally receive many more citations than ordinary articles,[17 18] is also recommended. Furthermore, because citation rate is roughly proportional to the length of the article,[19] journals might wish to publish long, rather than short, articles. If correction were made for article length, "communications" journals like *Biochemical and Biophysical Research Communications* and *FEBS Letters* would get impact factors as high as, or higher than, the high impact journals within the field, like *Journal of Biological Chemistry*.[20 21]

The use of an extremely short term index (citations to articles published only in the past two years) in calculating the impact factor introduces a strong temporal bias, with several consequences. For example, articles in journals with short publication lags will contain relatively many up to date citations and thus contribute heavily to the impact factors of all cited journals. Since articles in a given journal tend to cite articles from the same journal,[22] rapid publication is self serving with respect to journal impact, and significantly correlated with it.[23] Dynamic research fields with high activity and short publication lags, such as biochemistry and molecular biology, have a correspondingly high proportion of citations to recent publications—and hence higher journal impact factors—than, for example, ecology and mathematics.[23 24] Russian journals, which are cited mainly by other Russian journals,[25] are reported to have particularly long publication lags, resulting in generally low impact factors.[26] Pure technicalities can therefore account for several-fold differences in journal impact.

Limitations of the database

The Science Citation Index database covers about 3200 journals[8]; the estimated world total is about 126 000.[27] The coverage varies considerably between research fields: in one university, 90% of the chemistry faculty's publications, but only 30% of the biology faculty's publications, were in the database.[28] Since the impact factor of any journal will be proportional to the database coverage of its research field, such discrepancies mean that journals from an underrepresented field that are included will receive low impact factors. Furthermore, the journal set in the database is not constant but may vary in composition from year to year.[24 29] In many research fields a substantial fraction of scientific output is published in the form of books, which are not included as source items in the database; they therefore have no impact factor.[30] In mathematics, leading publications that were not included in the Science Citation Index database were cited more frequently than the leading publications that were included.[31] Clearly, such systematic omissions from the database can cause serious bias in evaluations based on impact factor.

The preference of the Science Citation Index database for English language journals[28] will contribute to a low impact factor for the few non-English journals that are included,[32] since most citations to papers in languages other than English are given by other papers in the same language.[25 27 33] The Institute for Scientific Information's database for the social sciences contained only two German social science journals, whereas a German database contained 542.[34] Specifically, American scientists, who seem particularly prone to citing each other,[33 35] dominate these databases to such an extent (over half of the citations) as to raise both the citation rate and the mean journal impact of American science 30% above the world average,[14] the rest of the world then falling below average. This bias is

aggravated by the use of a short term index: for example, in American publications within clinical medicine, 83% of references in the same year were to other papers by American scientists (many of them undoubtedly self citations), a value 25% higher than the stable level reached after three years (which would, incidentally, also be biased by self citations and citations of other American work).[33] Thus, both the apparent quality lead of American science and the values of the various journal impact factors are, to an important extent, determined by the large volume, the self citations, and the national citation bias of American science,[27] in combination with the short term index used by the Science Citation Index for calculating journal impact factors.

Journal impact factors depend on the research field

Citation habits and citation dynamics can be so different in different research fields as to make evaluative comparisons on the basis of citation rate or journal impact difficult or impossible. For example, biochemistry and molecular biology articles were cited about five times as often as pharmacy articles.[33] Several factors have been found to contribute to such differences among fields of research.

The citation impact of a research field is directly proportional to the mean number of references per article, which varies considerably from field to field (it is twice as high in biochemistry as in mathematics, for example).[24] Within the arts and humanities, references to articles are hardly used at all, leaving these research fields (and others) virtually uncited,[36] a matter of considerable consternation among science administrators unfamiliar with citation kinetics.[37]

In highly dynamic research fields, such as biochemistry and molecular biology, where published reports rapidly become obsolete, a large proportion of citations are captured by the short term index used to calculate journal impact factors, as previously discussed[38] —but fields with a more durable literature, such as mathematics, have a smaller fraction of short term citations and hence lower journal impact factors. This field property combines with the low number of references per article to give mathematics a recorded citation impact that is only a quarter that of biochemistry.[24]

In young and rapidly expanding research fields, the number of publications making citations is large relative to the amount of citable material, leading to high citation rates for articles and high journal impact factors for the field.[39 40]

In a largely self contained research field, the mean article (or journal) citation rate is independent of the size of the field,[41] but the absolute range will be wider in a large field, meaning higher impact factors for the top journals.[42] Such differences become obvious when comparing review journals, which tend to top their field (table 1). Leading scientists in a small field may thus be at a disadvantage compared with their colleagues in larger fields, since they lack access to journals of equally high citation impact.[43]

Most research fields are, however, not completely self contained, the most important field factor probably being the ability of a research field to be cited by adjacent fields. The relation between basic and clinical medicine is a case in point: clinical medicine draws heavily on basic science, but not vice versa. The result is that basic medicine is cited three to five times more than clinical medicine, and this is reflected in journal impact factors.[42 44 45] The outcome of an evaluation based on impact factors in medicine will therefore depend on the position of research groups or institutions along the basic-clinical axis.[33]

Table 1 Journal impact factors and research field

Journal	1986	1987
Annual Review of Biochemistry	31.6	35.1
Annual Review of Immunology	26.5	25.2
Annual Review of Cell Biology	14.1	22.8
Annual Review of Genetics	14.0	14.3
Annual Review of Neuroscience	15.4	13.7
Annual Review of Pharmacology	10.1	9.9
Annual Review of Physiology	7.8	9.1
Annual Review of Biophysics	7.2	7.7
Annual Review of Microbiology	4.9	6.4

In measures of citation rates of articles, attempts to take research field into account often consist of expressing citation rate relative to some citation impact specific to the field.[46] Such field corrections range from simply dividing the article's citation rate by the impact factor of its journal[28] (which punishes publication in high impact journals) to the use of complex, author specific, field indicators based on reference lists[47 48] (which punishes citations to high impact journals). However, field corrections cannot readily be applied to journal impact factors, since many research fields are dominated by one or a few journals, in which case corrections might merely generate relative impact factors of unit value. Even within large fields, the tendency of journals to subspecialise with certain subjects is likely to generate significant differences in journal impact: in a single biochemical journal there was a 10-fold difference in citation rates in subfields.[19]

Is the impact of an article increased by publication in a high impact journal?

It is widely assumed that publication in a high impact journal will enhance the impact of an article (the "free ride" hypothesis). In a comparison of two groups of scientific authors with similar journal preference who differed twofold in mean citation rate for articles, however, the relative difference was the same (twofold) throughout a range of journals with impact factors of 0.5 to 8.0.[12] If the high impact journals had contributed "free" citations, independently of the article contents, the relative difference would have been expected to diminish as a function of increasing journal impact.[49] These data suggest that the journals do not offer any free ride. The citation rates of the articles determine the journal impact factor (a truism illustrated by the good correlation between aggregate citation rates of article and aggregate journal impact found in these data), but not vice versa.

If scientific authors are not detectably rewarded with a higher impact by publishing in high impact journals, why are we so adamant on doing it? The answer, of course, is that as long as there are people out there who judge our science by its wrapping rather than by its contents, we cannot afford to take any chances. Although journal impact factors are rarely

used explicitly, their implicit counterpart, journal prestige, is widely held to be a valid evaluation criterion[50] and is probably the most used indicator besides a straightforward count of publications. As we have seen, however, the journal cannot in any way be taken as representative of the article. Even if it could, the journal impact factor would still be far from being a quality indicator: citation impact is primarily a measure of scientific utility rather than of scientific quality, and authors' selection of references is subject to strong biases unrelated to quality.[51 52] For evaluation of scientific quality, there seems to be no alternative to qualified experts reading the publications. Much can be done, however, to improve and standardise the principles, procedures, and criteria used in evaluation, and the scientific community would be well served if efforts could be concentrated on this rather than on developing ever more sophisticated versions of basically useless indicators. In the words of Sidney Brenner, "What matters absolutely is the scientific content of a paper, and nothing will substitute for either knowing or reading it."[53]

Funding: None.

Conflict of interest: None.

1 Hansen HF, Jørgensen BH. *Styring af forskning: kan forskningsindikatorer anvendes?* Frederiksberg, Denmark: Samfundslitteratur, 1995.
2 Cole S, Cole JR, Simon GA. Chance and consensus in peer review. *Science* 1981;214:881-6.
3 Ernst E, Saradeth T, Resch KL. Drawbacks of peer review. *Nature* 1993;363:296.
4 Garfield E. Citation analysis as a tool in journal evaluation. *Science* 1972;178:471-9.
5 Calza L, Garbisa S. Italian professorships. *Nature* 1995:374:492.
6 Drettner B, Seglen PO, Sivertsen G. Inverkanstal som fördelningsinstrument: ej accepterat av tidskrifter i Norden. *Läkartidningen* 1994;91:744-5.
7 Gøtzsche PC, Krog JW, Moustgaard R. Bibliometrisk analyse af dansk sundhedsvidenskabelig forskning 1988-1992. *Ugeskr Læger* 1995;157:5075-81.
8 Taubes G. Measure for measure in science. *Science* 1993;260:884-6.
9 Vinkler P. Evaluation of some methods for the relative assessment of scientific publications. *Scientometrics* 1986;10:157-77.
10 Maffulli N. More on citation analysis. *Nature* 1995;378:760.
11 Seglen PO. The skewness of science. *J Am Soc Information Sci* 1992;43:628-38.
12 Seglen PO. Causal relationship between article citedness and journal impact. *J Am Soc Information Sci* 1994;45:1-11.
13 Gordon MD. How authors select journals: a test of the reward maximization model of submission behaviour. *Social Studies of Science* 1984;14:27-43.
14 Braun T, Glänzel W, Grupp H. The scientometric weight of 50 nations in 27 science areas, 1989-1993. Part II. Life sciences. *Scientometrics* 1996;34:207-37.
15 Magri M-H, Solari A. The SCI Journal Citation Reports: a potential tool for studying journals? I. Description of the JCR journal population based on the number of citations received, number of source items, impact factor, immediacy index and cited half-life. *Scientometrics* 1996;35:93-117.
16 Moed HF, Van Leeuwen TN. Impact factors can mislead. *Nature* 1996;**381**:186.
17 Moed HF, Van Leeuwen TN, Reedijk J. A critical analysis of the journal impact factors of Angewandte Chemie and the Journal of the American Chemical Society. Inaccuracies in published impact factors based on overall citations only. *Scientometrics* 1996;37:105-16.
18 Bourke P, Butler L. Standard issues in a national bibliometric database: the Australian case. *Scientometrics* 1996;35:199-207.
19 Seglen PO. Evaluation of scientists by journal impact. In: Weingart P, Sehringer R, Winterhager M, eds. *Representations of science and technology*. Leiden: DSWO Press, 1992;240-52.
20 Sengupta IN. Three new parameters in bibliometric research and their application to rerank periodicals in the field of biochemistry. *Scientometrics* 1986;10:235-42.
21 Semenza G. Editorial. *FEBS Lett* 1996;389:1-2.
22 Seglen PO. Kan siteringsanalyse og andre bibliometriske metoder brukes til evaluering av forskningskvalitet? *NOP-Nytt (Helsingfors)* 1989;15:2-20.
23 Metcalfe NB. Journal impact factors. *Nature* 1995;**77**:260-1.
24 Moed HF, Burger WJM, Frankfort JG, Van Raan AFJ. The application of bibliometric indicators: important field-and time-dependent factors to be considered. *Scientometrics* 1985;8:177-203.
25 Lange L. Effects of disciplines and countries on citation habits. An analysis of empirical papers in behavioural sciences. *Scientometrics* 1985;8:205-15.
26 Sorokin NI. Impact factors of Russian journals. *Nature* 1996;380:578.
27 Andersen H. ACTA Sociologica på den internationale arena-hvad kan SSCI fortælle? *Dansk Sociologi* 1996;2:72-8.
28 Moed HF, Burger WJM, Frankfort JG, Van Raan AFJ. *On the measurement of research performance: the use of bibliometric indicators*. Leiden: Science Studies Unit, LISBON-Institute, University of Leiden, 1987.
29 Sivertsen G. *Norsk forskning på den internasjonale arena. En sammenlikning av 18 OECD-lands artikler og siteringer i Science Citation Index 1973-86*. Oslo: Institute for Studies in Resesearch and Higher Education, 1991.
30 Andersen H. Tidsskriftpublicering og forskningsevaluering. *Biblioteksarbejde* 1996;17:19-26.
31 Korevaar JC, Moed HF. Validation of bibliometric indicators in the field of mathematics. *Scientometrics* 1996;37:117-30.
32 Bauin S, Rothman H. "Impact" of journals as proxies for citation counts. In: Weingart P, Sehringer R, Winterhager M, eds. *Representations of science and technology*. Leiden: DSWO Press, 1992;225-39.
33 Narin F, Hamilton KS. Bibliometric performance measures. *Scientometrics* 1996;**6**:293-310.
34 Artus HM. Science indicators derived from databases. *Scientometrics* 1996;37:297-311.
35 Møller AP. National citations. *Nature* 1990;348:480.
36 Hamilton DP. Research papers: who's uncited now. *Science* 1991;251:25.
37 Hamilton DP. Publishing by—and for?—the numbers. *Science* 1990; 250:1331-2.
38 Marton J. Obsolescence or immediacy? Evidence supporting Price's hypothesis. *Scientometrics* 1985;**7**:145-153.
39 Hargens LL, Felmlee DH. Structural determinants of stratification in science. *Am Sociol Rev* 1984;49:685-97.
40 Vinkler P. Relationships between the rate of scientific development and citations. The chance for a citedness model. *Scientometrics* 1996;35:375-86.
41 Gomperts MC. The law of constant citation for scientific literature. *J Documentation* 1968;24:113-7.
42 Seglen PO. Bruk av siteringsanalyse og andre bibliometriske metoder i evaluering av forskningsaktivitet. *Tidsskr Nor Laegeforen* 1989;31:3229-34.
43 Seglen PO. Evaluering av forskningskvalitet ved hjelp av siteringsanalyse og andre bibliometriske metoder. *Nordisk Medicin* 1989;104:331-5.
44 Narin F, Pinski G, Gee HH. Structure of the biomedical literature. *J Am Soc Inform Sci* 1976;27:25-45.
45 Folly G, Hajtman B, Nagy JI, Ruff I. Some methodological problems in ranking scientists by citation analysis. *Scientometrics* 1981;3:135-47.
46 Schubert A, Braun T. Cross-field normalization of scientometric indicators. *Scientometrics* 1996;36:311-24.
47 Lewison G, Anderson J, Jack J. Assessing track records. *Nature* 1995;377:671.
48 Vinkler P. Model for quantitative selection of relative scientometric impact indicators. *Scientometrics* 1996;36:223-36.
49 Seglen PO. How representative is the journal impact factor? *Research Evaluation* 1992;2:143-9.
50 Martin BR. The use of multiple indicators in the assessment of basic research. *Scientometrics* 1996;36:343-62.
51 MacRoberts MH, MacRoberts BR. Problems of citation analysis: a critical review. *J Am Soc Information Sci* 1989;40:342-9.
52 Seglen PO. Siteringer og tidsskrift-impakt som kvalitetsmål for forskning. *Klinisk Kemi i Norden* 1995;7:59-63.
53 Brenner S. Cited [incorrectly] in: Strata P. Citation analysis. *Nature* 1995;375:624.

(Accepted 9 January 1997)

ONE HUNDRED YEARS AGO

Samson's riddle

Within the last few days two great sums of money have been given for the advancement of science. Dr Alfred Nobel, the inventor of dynamite, has left nearly £2,000,000 as a prize fund for the most important discoveries in physics, chemistry, physiology, and medicine; and the Baroness Hirsch has promised £80,000 to the Pasteur Institute. We who have just received Dr Mond's splendid gift of the Davy-Faraday Laboratory need not envy the good fortune of other nations, nor does help given to Science benefit one country, but all. The work of the Pasteur Institute goes forth over the whole world; as Dr Roux says of it, "We are perpetually carrying on a struggle against death, and we can only express our deep gratitude to all benefactors who help us to lighten the load of suffering humanity." The prizes given by Dr Nobel are open to all the nations; and he has added yet another prize for him who has done most to promote the cause of peace. Truly, as in Samson's riddle, out of the eater has come forth meat, out of the strong has come forth sweetness. (*BMJ* 1897;i:161.)

The rationing debate

Defining a package of healthcare services the NHS is responsible for

The case for

Bill New

As part of its mission to raise the level of debate about rationing of healthcare resources, the Rationing Agenda Group commissioned six pairs of articles debating specific propositions to do with rationing. The BMJ plans to publish these at roughly fortnightly intervals over the next few weeks. The first appears below.

There is a need for a clearly defined package of healthcare services which is relevant to, and the responsibility of, the NHS

King's Fund Policy Institute, London W1M 0AN
Bill New,
senior research officer,

BMJ 1997;314:503–5

The use of tattoo removal as an example of NHS rationing is now so common that it is in danger of trivialising an important debate. Behind such questions as "Should the NHS be devoting resources to tattoo removal?" lies a more fundamental issue: what *kinds* of benefit should the NHS provide?

Most readers will assume that defining a healthcare "package" is a means of rationing healthcare resources. In other words, faced with the task of managing the limited NHS budget one option is to exclude some services altogether. But my case rests on a different interpretation of a package and involves asking a preliminary question. Before deciding how to ration, we need to know what to ration: What is the range of services relevant to the role of the NHS? What "business" is the NHS in? Is it the NHS's job to provide fertility treatment, physiotherapy for sports injuries, long term nursing care, gender reassignment, adult dentistry, and cosmetic surgery? Or should these services be provided by local authorities, voluntary agencies, or the private sector? The question does not rely on clinical judgment. It is about the boundary of a public institution's responsibilities. And it is a question which has been muddled up with issues of rationing proper.

Defining the boundaries

The need to address this question derives from a growing sense of confusion and uncertainty about what it is reasonable to expect from the NHS. For example, where one lives can have a decisive effect on whether or not NHS treatment is available. The 1991 NHS reforms were an important catalyst in this process: purchasing authorities now concentrate on commissioning health care for their resident populations, rather than on management issues. Wishing to be seen to be making the best use of financial allocations, some took the view that certain services were not a priority and therefore not worth purchasing. For example, the availability of fertility treatment depends on the whim of purchasing health authorities—and increasingly general practice fundholders. In addition to this uncertainty over regional variations is the apparent removal of some primary care services from NHS provision altogether. Adult dentistry is subsidised by central government, but only for a fraction of its cost. In some areas it is hard to find dentists who offer even this minimal NHS cover. There has been no explicit national debate about why dentistry is apparently not an NHS responsibility.

It is inequitable that one's place of residence should determine access to care. Levels of service provision will inevitably vary from one part of the country to another, in response to varying need or because some providers are more efficient than others. But this is different from removing availability altogether: infertility is not at zero levels in those areas where in vitro fertilisation treatment it is not available. Ad hoc developments such as these can serve only to promote uncertainty and a sense of unfairness quite out of proportion to the quantity of resources at stake. Furthermore, the vigilance of the news media has had a significant impact on the public's perception of health delivery. Activity, and inactivity, in the NHS is now scrutinised and reported daily. This is welcome, but awkward: old issues, once hidden, must now be tackled if the NHS is not to fall into disrepute.

Nothing to do with saving money

To be clear about what devising a healthcare package would seek to achieve, we must be clear about what it is not trying to do. Firstly, it is not (necessarily) about saving money. The case for a centrally defined package has been associated with easing the pressure on resources. However, this is not the purpose of the proposals outlined here—the desire is to promote equity, collective understanding, and reassurance. The package considered relevant to NHS business is just as likely to be more extensive than that available now. The point is that it should not vary from one area to another and that it should be derived as the result of an explicit, democratic process.

> *"Before deciding how to ration, we need to know what to ration"*

Secondly, drawing up such a package is not an attempt to avoid additional rationing. In the well known Oregon initiative in the USA, all those services which might possibly be provided collectively are ranked and the line drawn where resources allow. Above the line everyone has access; below no one does (unless privately financed). The line moves up and down depending on the availability of resources. Rationing health care is therefore a centrally

The BMJ, BMA, King's Fund, College of Health, and RAG (Rationing Agenda Group) will be holding a conference entitled "Rationing in the NHS: Time to get real" on 10 and 11 July at Kensington Town Hall. For further information ring Jane Lewis 0171 383 6605 (Email: 106005.2356@compuserve.com).

undertaken activity, specifying a package to which everyone has access.

But the approach presented here is a preliminary to rationing. It is about deciding what should appear on Oregon's list in total, not about where the line should be drawn. Of course, there will be resource implications from this decision. For example, if long term nursing care was considered an NHS responsibility then resources might need to be reallocated from private households to the NHS through taxation. But specifying a package to promote reassurance and geographical fairness does not involve deciding how NHS resources should be distributed between individuals who make a claim on them. Defining a package does not imply a right to treatment.

Cost effectiveness not relevant to establishing "the package"

However, if the need, in principle, for a package is accepted then there must be a coherent and practical means of establishing it. For this to be successful the focus must shift from criteria which guide rationing decisions to criteria which help establish the range of relevant services which are to be rationed. Trying to do both at once results in doing very little at all, as international experience testifies.

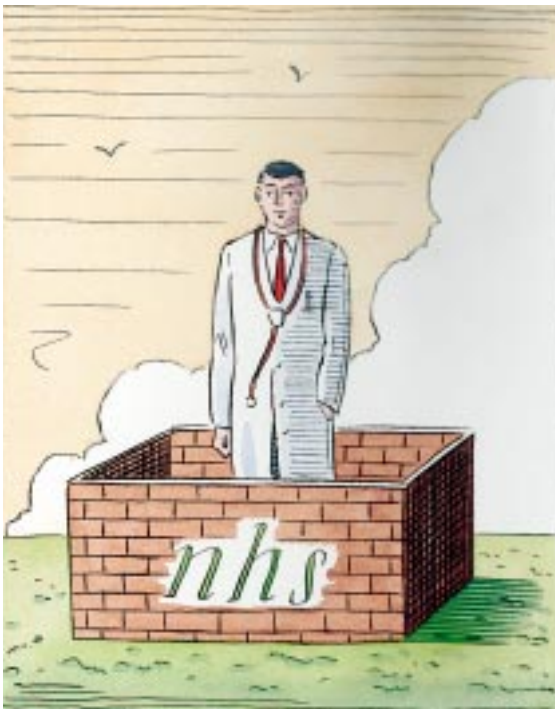
> *"There has been no explicit national debate about why dentistry is apparently not an NHS responsibility"*

New Zealand and Holland have both tried to establish packages of healthcare services, and both have had little success. In New Zealand, "core" services were intended both to clarify what the population could expect and limit the financial burden on the state. Ultimately, though, the planners were forced to concede that everything that was currently provided would form the core—hardly the result they were looking for. In Holland, four criteria—necessity, effectiveness, efficiency, and individual responsibility—were used to define a package. The Dutch also found it difficult to specify precisely which services ought to be excluded. Why so little success?

These strategies were trying to do too much at once. Issues of equity, reassurance, and clarity about the responsibility of the state were mixed up with a desire to contain costs and ration more systematically. The root of the difficulties lies in the inclusion of effectiveness and cost effectiveness as criteria for establishing a package.

Whether a service is relevant to a healthcare system has nothing to do with effectiveness. For example, cosmetic surgery for enhancement where there is no severe psychological distress is rarely supported for collective healthcare provision. But no one suggests that the plastic surgeons who work miracles on Hollywood stars are not effective. So questions of effectiveness and cost effectiveness should be left aside at this stage of the debate.

Their importance comes when deciding how to allocate resources between all the services that are relevant and between the people who can benefit from them. The result of such deliberations may also lead to



SIMON STERN/THE INKSHED

a package, but it will be of an entirely different kind. In Oregon, cost effectiveness was the basic principle at work (although with many refinements and alongside other criteria). But the resulting package did not address whether fertility treatment, residential care, dentistry, and so on were relevant to Oregon's public healthcare provision.

Reliance on measures of cost and effectiveness has meant that most commentators believe strategies for defining a package are doomed to failure, or at least likely to disappoint those who promote them.[1] The reasons given are now reasonably well accepted: health interventions are extremely variable in their effects on individuals. Just about every treatment therefore does some good for someone, even if it is "ineffective" in general. Making blanket exclusions on this basis will inevitably be a blunt instrument and will antagonise doctors, who feel their clinical freedom is curtailed. As a consequence, the Oregon experiment has proved to be extremely controversial and has generally not been considered relevant to the NHS.

A qualitative approach

But the following approach for establishing relevance to the NHS does not rely on effectiveness. Instead it proposes a qualitative approach which avoids the difficulties of variable individual response to treatment. The approach has been described in more detail elsewhere,[2] but the central proposition is this: those characteristics which define healthcare's special nature, and which in general terms make it unsuitable for economic exchange, should determine whether or not individual services are relevant to the NHS.

Taken together, three characteristics set health care apart: fundamental importance, information imbalance, and uncertainty. Health is clearly of fundamental importance; there is little certainty about how our health will develop in the future; and typically we know

little about the nature of our needs for health care or of the likely effects of the treatments available to us. In short, we are not good consumers. We must trust doctors who, under free market conditions with insurance markets, have an incentive to provide as much care as possible.

"Whether a service is relevant to a healthcare system has nothing to do with effectiveness"

It is the combination of characteristics which makes health care special. Fundamental importance is not sufficient on its own—food, after all, is important but no one suggests that we should have a National Food Service because people are perfectly able to choose what and how much food they need. Neither are information imbalance and uncertainty sufficient on their own. Car repair services display these characteristics, but no one is suggesting a National Car Repair Service—because cars are not of fundamental importance.

However, not all services which are related to health care display these characteristics in combination. Residential care for the elderly is fundamentally important but does not suffer from significant information imbalances; cosmetic surgery for enhancement is not generally considered to be of fundamental importance. Services of this kind should not be an NHS responsibility. On the other hand, curative dentistry, fertility treatment, and intensive nursing care do seem to satisfy the criteria; consequently, they should form part of the NHS's range of available services. All health authorities and GP fundholders should be required to provide some level of these services. Note that defining this package does not imply any judgment about effectiveness or whether any individual service represents good value for money. It simply clarifies what the NHS should be in the business of doing.

Meeting the objections

There are a several practical objections to such a proposal. The first argues that such an approach necessitates drawing up a long list of individual services which would need to be continually updated. Every existing and new treatment for cancer, for example, would need specification. In fact, such an enterprise would be unnecessary. Specification is required only at a general level: treatment for cancer, AIDS, infertility, or whatever, not individual drugs or surgical interventions. Indeed, it may be sufficient simply to concentrate on exclusions rather than long lists of inclusions. Furthermore, whether or not a particular treatment is effective is irrelevant—the NHS may decide not to purchase a drug to treat AIDS until it proves its effectiveness, but it would be clear that drugs of this kind are relevant to NHS business.

The second objection argues that any attempt to centralise decision making will fail to accommodate individual cases and exceptional circumstances. Clearly, defining the range of relevant services will restrict clinicians' ability to use their skills to certain ends. But this will not be as inflexible as critics suggest.

A typical example cites an individual with extreme psychological distress caused by a tattoo mistakenly purchased when young. In circumstances where cosmetic surgery for enhancement is outside NHS responsibilities, surely such central codification would deny legitimate treatment? Not if treatment is correctly focused on the nature of the condition. In the case of psychological distress the correct course of action is referral to a psychiatrist (psychiatry, let us assume, is within NHS responsibilities). The specialist might well provide appropriate treatment herself or may recommend removal of the tattoo as the preferred means of treating the distress. In this way tattoo removal could still be undertaken quite legitimately on the NHS. But at the same time some restriction has been placed on clinical freedom: cosmetic procedures for conditions that do not involve psychological distress are outside the scope of the NHS—no exceptions. If a real improvement in the clarity of the NHS's role and equity in the availability of its services is to be achieved then such restrictions are inevitable.

"It is the combination of characteristics which makes health care special"

The final objection asserts that introducing legal specifications of what the NHS should provide will simply allow the nominal provision of services. So, for example, if infertility were to be included it would probably be sufficient for a health authority to provide one round of treatment per year to satisfy its legal obligations. Thus health authorities could pay lip service to the policy but continue to decide the range of responsibilities for themselves. This may indeed turn out to be how purchasers act. But it is equally reasonable to suppose that health authorities simply want clear guidance on the range of their responsibilities and have no wish to indulge in gamesmanship with policy directives. In any event, individual purchasers would be clear about their role—they would still decide on the extent of provision depending on local circumstances, but would now provide services secure in the knowledge that they were in step with all other purchasers in the NHS.

One final point. Criteria such as those suggested here for guiding the specification of a healthcare package are just that: guidance. They cannot replace debate and political compromise. So it is not possible simply to read off a list of relevant healthcare services—people will inevitably disagree over the degree to which fertility treatment, for example, satisfies information imbalance and fundamental importance. But once a decision has been made, openly and with reference to coherent criteria, what we stand to gain from clarity and equity will surely outweigh the awkward processes involved. The alternative—allowing the NHS increasingly to ignore the principles on which it was founded—risks losing mass popular respect for a successful and valued public institution.

1 Klein R. Can we restrict the health care menu? *Health Policy* 1994;27:103-12.
2 New B, Le Grand J. *Rationing in the NHS: principles and pragmatism.* London: King's Fund, 1996.

The case against

Rudolf Klein

There is a need for a clearly defined package of healthcare services which is relevant to, and the responsibility of, the NHS

King's Fund Policy Institute, London W1M 0AN
Rudolf Klein, *professorial fellow*

BMJ 1997;314:506–9

Traditionally the National Health Service has relied on implicit rationing by clinicians within budgetary constraints set by central government. Neither the founding fathers, nor any of their successors, defined the scope and limits of the NHS's responsibilities. The statutory responsibility of ministers of health is to provide an "adequate" service. But the frontiers of adequacy have never been defined, and the courts have resolutely refused to rule on what should be provided to whom. The package of health care offered has thus varied, in terms of its composition and its generosity, both over time and geographically. It is for individual health authorities to decide what package of healthcare services to provide for their populations and for individual clinicians to decide between the competing claims on the resources available to them.[1]

Unsatisfactory state of affairs

In many ways this is an unsatisfactory state of affairs and not surprisingly has increasingly come under challenge. There are two main grounds for criticism. Firstly, the present situation allows ministers to duck responsibility for the consequences of their decisions when setting the NHS's budget. If "adequacy" remains an elastic and fuzzy notion, there is no way of establishing whether the budget is sufficient to meet the NHS's commitments. Without any definition of what those commitments are in the first place, the debate about whether or not the NHS is "underfunded" becomes a meaningless dialogue of the deaf and accountability is fudged. Secondly, the lack of any defined package means that in practice there can be no equity, if by equity is meant that everyone should have the same opportunity of treatment for any given degree of need for a particular healthcare intervention. A considerable degree of arbitrariness remains in the chances of getting treated in the NHS: even if equity were achieved in terms of ensuring that all health authorities have the same command over resources, relative to the needs of their populations, there would still be no assurance that they would necessarily buy the same set of services.

Developments since the introduction of the 1991 NHS reforms have reinforced these general considerations. The medical profession has become increasingly restive about having to carry the ultimate responsibility for rationing. If resources are short, many doctors now argue, ministers should accept the burden of determining what can or cannot be provided. Further, the purchaser-provider split has given visibility to decisions by health authorities about what care to buy—or not to buy—for their populations. The visibility may be less than complete, but enough has been revealed to cause disquiet about the differences of policy that have emerged. For example, it would seem absurd that the chances of getting in vitro fertilisation treatment should depend on where people live, yet health authorities differ sharply on whether or not to buy this treatment.[2]

SIMON STERN/THE INKSHED

The case for defining the package of healthcare services to be delivered by the NHS is therefore strong. I will argue, however, that the case against moving in this direction is even stronger—for four main reasons. Firstly, no consensus exists about the principles or criteria that should be used in designing such a package. Secondly, any decisions to restrict the NHS menu are difficult to implement, given patient heterogeneity. Thirdly, by concentrating on rationing by denial we risk ignoring other (probably more important) dimensions of rationing. Fourthly, defining an NHS package would probably not achieve the declared objectives of promoting equity and accountability, and the all out pursuit of these objectives would in any case lead to a damaging degree of centralised rigidity.

A perplexity of criteria

The case for defining the NHS menu assumes the feasibility of developing criteria acceptable to both public and professionals for determining what should be included and excluded. And indeed many such criteria are on offer. The growing international interest in limiting the open ended financial commitments of healthcare systems has, in turn, produced a series of attempts to develop principles for defining the limits of national packages. The result of all this activity is, however, discouraging. A cacophony of criteria is on offer, embodying competing (and sometimes conflicting) views about how to define the limits of public responsibility for health care. And when seemingly uncontentious criteria are proposed, it turns out that their acceptability depends crucially on their level of abstraction: acceptable as general propositions, they become contentious when applied to particular cases.

International experience

The problem of conflicting criteria is well illustrated by the international experience.[1] Oregon's much cited exercise in defining a package of care for those not covered by medical insurance was originally based on ranking different conditions-treatments according to their cost-benefit ratio. It thus embodied the economist's notion that any package of care should be designed to maximise the community's return on any resources invested in health care. In the outcome this was effectively abandoned, partly because adequate data were lacking, partly because the exercise produced some counterintuitive results—for example, appendectomy ranked lower than tooth capping. The final rankings that appeared, after repeated massaging, seem to reflect judgments about "reasonableness" taken in the light of community values. In other words, the attempt to apply clear cut, transparent criteria was abandoned.

"Attempts at defining a package of health care stub their toes against the rock of patient heterogeneity"

The Dunning committee, which sought to develop criteria for defining a package of health care for the Netherlands, proposed four criteria for including any intervention or services: necessity, effectiveness, efficiency, and whether it is a matter of individual, rather than community, responsibility. Necessary services were those which "guarantee normal functions as a member of the community or simply protect existence as a member." It thus recognised both economic considerations and the "rule of rescue"—that conditions that threaten survival or the capacity to function must be included in any package. So, for example, in vitro fertilisation failed to pass the necessity test: "Undesired childlessness in the Netherlands poses no danger to the community, and it cannot be said that childlessness interferes with normal function in our society."

In contrast, a Swedish commission rejected outright the efficiency principle—that is, the economistic approach to defining a basic package. Instead, it endorsed the rule of rescue by giving priority to the treatment of life threatening conditions. It also invoked social solidarity—the principle that the commitments of a healthcare system should be shaped by a sense of collective responsibility for the wellbeing of its members, especially the most vulnerable, such as the chronically and terminally ill. And the New Zealand committee on core services declined the task of defining the contents or limits of a healthcare package.

Defining characteristics of health care

The list of possible criteria could be extended further. Consider, for example, a recent, highly ingenious attempt to cut through the confusion by deriving the principles for defining the basic package from the arguments used to justify public intervention in the provision of health care in the first place. Public provision or financing is justified, New and Le Grand argue,[3] by three defining characteristics of health care that distinguish it from other goods: the unpredictability of need, information asymmetry between patients and providers, and its fundamental importance to people's ability to achieve their life goals. Unpredictability and asymmetry, as the authors recognise, also characterise many other transactions. So we are left with "fundamental importance"—health care as the key to functioning in society—as the key criterion.

This, of course, is unexceptionable: who could disagree? Indeed this criterion is first cousin to the Dunning committee's criterion of "necessity." But how are we to define fundamental importance? The Dunning committee excluded in vitro fertilisation; New and Le Grand consider the inability to have a child to be of fundamental importance. The criterion thus turns out to be vacuous to the extent that it provides no guidance on how conflicting interpretations can be resolved. This leads to a general conclusion: principles that incorporate a semiautomatic formula for implementing them (like maximising health benefits) tend to be highly contentious, while uncontentious principles owe their acceptability to the fact that there is ambiguity about their implementation.

On one point only there appears to be widespread agreement. This is that "ineffective" health care should be excluded from any healthcare package or menu. Again who could disagree? This turns out to be a rather blunt criterion. Most interventions are effective for someone, just as most services may be of fundamental importance to someone. Inevitably attempts at defining a package of health care stub their toes against the rock of patient heterogeneity: a point explored further in the next section.

Problems of practice

Given that it is so difficult to devise coherent criteria or principles, it is not surprising that most attempts to define the healthcare menu have been tentative and somewhat incoherent. The issue has, in effect, been approached backwards: by listing exclusions rather than by defining what is to be included. Oregon remains unique in explicitly setting out what will be provided, as well as what will be excluded—not surprisingly perhaps since the whole venture started as an exercise in extending coverage for the uninsured. Otherwise, however, the problem has been defined, in practice, as an exercise in limiting what is to be available. About a quarter of the 100 health authorities in England explicitly set out, in their purchasing plans for 1996-7, procedures which will not be included in their contracts.[4]

"Rationing...is about the exercise of judgment, not the drawing up of lists"

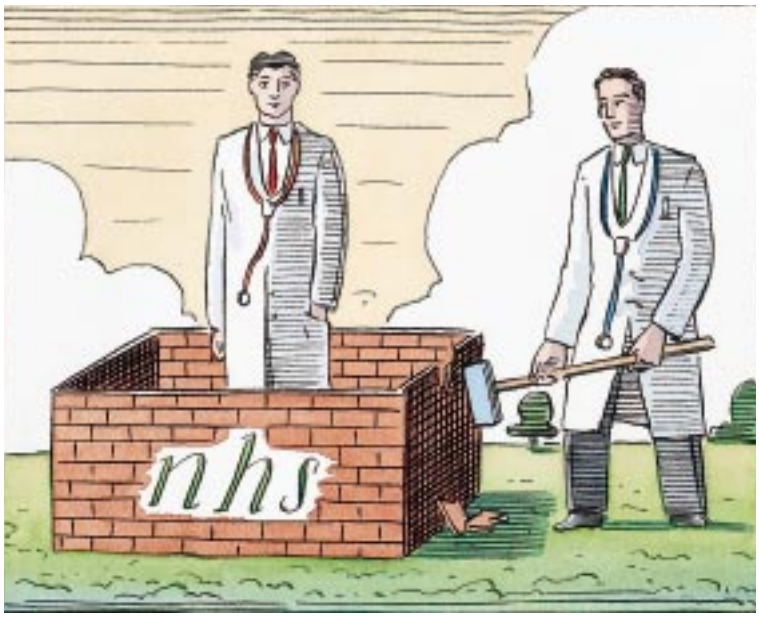
The lists of exclusions are dominated by various forms of cosmetic surgery. These range from tattoo removal to buttock lift, from breast augmentation to procedures for pinning back ears. The reversal of sterilisation and of vasectomy also feature frequently. Also included is in vitro fertilisation. Interestingly, the roll call of exclusions is not some peculiar English eccentricity. Very similar lists have been produced in Ontario and in Spain when attempts have been made to define entitlements to healthcare.

Nibbling at the edges

The above catalogue of exclusions is not complete. Health authorities have also begun to exclude some procedures, such as dilatation and curettage for women under 40, where the evidence suggests that they are not clinically effective. But, overall, the exclusions affect only the small change of NHS activity. The implicit criterion appears to be that it is no part of the NHS's responsibilities to deal with conditions that are self inflicted or that can be seen as not being "medical"—that is, as not impairing the ability to function. In effect, it is not the NHS's job to help people to look good.

As a criterion for defining the healthcare menu, this does not take one very far. But the experience of applying even this principle helps to explain why attempts to restrict the menu have been limited to nibbling at the edges—and why, indeed, the whole enterprise is flawed in conception. The pattern is clear. No sooner does a health authority announce that it is proposing not to buy a particular procedure than there is an outcry. In part, this is the reflex reaction of a medical profession which claims that only its members are qualified to pronounce on what patients need. But, more importantly, it reflects the fact of patient heterogeneity. For some patients, if only a very few, even a buttock lift may be crucial for their ability to function socially. Blanket exclusions ignore the fact of patient heterogeneity. Health authorities have therefore tended to retreat from such blanket exclusions to a more flexible position: such procedures are normally not bought unless a clinical need can be demonstrated.

If it is difficult to sustain a policy of blanket exclusions even in the case of marginal procedures it is not surprising that there has been little or no attempt to extend it to services which make larger demands on the NHS budget and which are more central to people's conception of health care. Much the same conclusion follows if we look at the list of medicines which general practitioners are not permitted to prescribe: these range from various cold relief remedies to toothpaste.[3] Again, the exclusions are conspicuous for their marginality. Any attempt to go further in this direction is therefore likely to mean that there would have to be a disproportionate investment of energy—in terms of persuading professional and public opinion—with a trivial yield in terms of the effects on the NHS's budgetary commitments. And it might well divert attention from the real challenge of managing scarcity in the NHS: which is not whether to provide specific services but how much to provide and how to decide who should be treated and how.



SIMON STERN/THE INK SHED

The dimensions of rationing

Explicit rationing through the exclusion of specific procedures or services from the healthcare menu has the great appeal of giving visibility to collective decisions about how the resources devoted to the NHS should be used. But, as argued above, in practice the inevitable price of visibility is triviality. Moreover, any such strategy fails to address the various dimensions of rationing. If rationing is defined as giving patients less demonstrably effective health care than might be desirable in the absence of resource constraints (and from which they would benefit), then it is a pervasive characteristic of the NHS and of other healthcare systems. There is rationing by dilution: offering less intensive care or spending less time with a patient than might be ideal. There is rationing by termination: discharging patients from hospital earlier than desirable. There is rationing by delay: waiting lists are the obvious example.

The following two quotations illustrate the point. The first is from an American commentator on the Oregon experiment[5]:

> It takes no great talent to realise that appendectomy is worth funding, at least for a clear cut diagnosis of appendicitis. The real issue is not whether to perform the appendectomy; it is whether to fund countless marginal interventions that are potentially part of the procedure—marginal blood tests and repeat tests; precautionary preventive antibiotic therapy before surgery; the number of nurses in the operating room; and the backup support on call or in the hospital. Even more decisions about marginal elements will arise during the recovery phase; exactly how many days of hospital stay are permitted, how often the physician should make rounds, how many follow up tests there should be, and so on. Many of these are predicted to offer more benefits than harm, but with margins so small that one could argue that resources should be used elsewhere.

The second comes from the Swedish commission on priorities[1]:

> If resources are limited then in certain circumstances it may be reasonable to opt for the second best treatment. In hip surgery, for example, a steel prosthesis is less expensive than a titanium one but less durable. It must be considered acceptable for a physician, as is often the case, to choose a steel prosthesis for a patient aged over 80 while giving a titanium one to a patient who is 70 years old and might perhaps need renewed surgery after a few years. In dealing with pronounced coronary strictures involving a risk of stroke, one can choose between surgery and medication. Surgery is a good deal more expensive, involves a short term risk but is in the long term a more effective method of averting stroke. If resources are limited, it may be justifiable to refrain from expanding surgical activities and to stick to medication—which is simple, inexpensive but less effective—instead.

Countless day to day decisions

This, then, is the reality of rationing: countless, day to day decisions by clinicians and others taken in the light of the resources available and the particular circumstances of the patient concerned. Rationing, in effect, is a continuous attempt to reconcile competing claims on

limited resources, a balancing act between optimising and satisfying treatment. It is about the exercise of judgment, not about the drawing up of lists of what should or should not included in the NHS's menu.

Nor is the phenomenon of rationing limited to the acute sector of care. At present there is much debate about the extent to which the NHS should provide long term care and how the demarcation line between social and medical care should be defined. One central fact tends to be overlooked: the NHS has always rationed—to the point of scandal—the care provided in the long term sector. An analysis of the reports of the Health Advisory Service on NHS provision for the elderly mentally ill showed that, of the hospitals inspected, 77% had poor sanitary conditions, 66% dilapidated buildings, and 60% overcrowded wards.[6] This is rationing of a very different kind from that discussed previously: the poor quality care flows directly from decisions about the allocation of resources to different parts of the NHS and does not involve clinical decisions. But it is rationing. Dilution could hardly go further.

"Rather than worrying about drawing up an NHS menu, we should concentrate on what is going on in the kitchen"

There are many problems about the kind of implicit rationing that characterises the NHS. It may involve the use of arbitrary and unacceptable criteria, such as the age of the patient. It may mean that the allocation of resources reflects as much the idiosyncrasies of individual clinicians as the characteristics of patients. Clinician heterogeneity is as much of a problem as patient heterogeneity. From this, of course, flows the equity case for defining the care that the NHS should provide. However, it is far from self evident that defining such a package or menu would resolve the equity issue.

Designing a straitjacket

A strict construction of the equity principle would require not only that all health authorities should provide the same range of services but that individual patients should have the same opportunity of getting treatment for any given need. In other words, the argument would move from specifying the range of services to be provided to specifying also the quantity of services to be provided. And indeed there have been some moves in this direction. In the 1980s the Department of Health started to set health authorities targets for carrying out certain procedures such as coronary artery bypass grafts and hip replacements. For example, the 1990 target was a rate of 300 coronary artery bypass grafts and 1950 hip replacements per million population.[7] Implicit in this strategy was the objective that health authorities should, in effect, deliver a particular quantity of specified packages of health care to their populations. This strategy appears to have been largely abandoned as the department's focus has switched to outcomes, but it shows that, in principle at least, it would be possible to define both the menu and the number of meals to be served.

Indeed the equity case for designing a basic package of care would seem to require taking this further step. For without it "inconsistency and arbitrariness in the rationing of health care," as New and Le Grand put it, would surely persist. In vitro fertilisation provides an illustration. Many health authorities have decided against excluding this from their provision but have, instead, adopted a strategy of limiting the number of procedures to be made available. They therefore ration by the selection of patients, with different health authorities using different criteria (the age of the prospective mother, the stability of the partnership, etc) for selecting candidates. The service actually provided could therefore be extremely sparse and largely symbolic in many areas. And while in vitro fertilisation may be a special case, differences in the way in which health authorities interpret their responsibilities are in no way exceptional. Throughout the whole range of NHS provision variation in the level of services provided is the norm.

Pursuing the logic

However, advocates of specifying the range of service to be provided by the NHS tend to flinch from the logic of their own arguments. Thus New and Le Grand (to quote, again, the most sophisticated exponents of this approach) conclude that while every health authority would have to provide at least some level of service for the specified range "the actual level of provision, highly contingent on local circumstances, would be left to local discretion." So "inconsistency and arbitrariness" come in by the back door of local discretion.

There is, of course, a case for local discretion. The scope for substitution in health care is great: a deficit in one kind of service may be compensated for by provision elsewhere. Health care is in a continuous process of technological and organisational evolution and to specify particular levels of service would put the NHS in a straitjacket, inhibiting adaptation and innovation. The concept of "need" is, itself, highly elusive, and flexibility in its interpretation is essential. However much we may chafe at the way in which local discretion is often exercised, it still seems preferable to imposing a national template on the design and delivery of health care.

But if this line of argument is accepted, there is little left of the case for devising an NHS menu. If its proponents refuse to accept the full logic of their own case—if they rightly recoil from the notion of imposing a national template on the NHS—there would seem little point in travelling half way down the road with them. Rather than worrying about drawing up an NHS menu, we should concentrate on what is going on in the kitchen: we should accept the inevitability and indeed desirability of leaving rationing decisions to clinicians and concentrate on ways of making the profession collectively more accountable for the way in which they carry out this onerous task.

1 Klein R, Day P, Redmayne S. *Managing scarcity: priority setting and rationing in the NHS.* Buckingham: Open University Press, 1996.
2 Redmayne S, Klein R. Rationing in practice: the case of in vitro fertilisation. *BMJ* 1993;306:1521-4.
3 New B, Le Grand J. *Rationing in the NHS: principles and pragmatism.* London: King's Fund, 1996.
4 Redmayne S. *Small steps, big goals: purchasing policies in the NHS.* Birmingham: NAHAT, 1996.
5 Veatch RM. The Oregon experiment: needless and real worries. In: Strosberg MA, Wiener JM, Baker R, Fein IA, eds. *Rationing America's medical care: The Oregon plan and beyond.* Washington, DC: Brookings Institution, 1992.
6 Day P, Klein R, Tipping G. *Inspecting for quality: services for the elderly.* Bath: Centre for the Study of Social Policy, 1988.
7 Department of Health and Office of Population Censuses and Surveys. *Departmental Report.* London: HMSO, 1992 (Cm 1913).

Primary care: opportunities and threats

Deregulating primary care

Angela Coulter, Nicholas Mays

This is the second of a series of six articles discussing the imminent reforms in primary care

King's Fund, London W1M 0AN
Angela Coulter, *director, King's Fund Development Centre*
Nicholas Mays, *director of health services research, King's Fund Policy Institute*
Correspondence to: Dr Coulter.

BMJ 1997;314:510–3

The 1996 primary care white paper, *Choice and Opportunity*, offers scope for a wide range of new organisations to enter the market for NHS primary care. If the new law is implemented it will sweep away many of the existing legislative, budgetary, and procedural barriers to innovation, initially through a range of pilot schemes. Full implementation could end general practitioners' monopoly of general medical services and permit experimental alternatives to the single national contract for general practitioners. Such pilots have the potential for creating better primary care, especially in inner cities, through new employment options, such as salaried posts. There are, however, considerable risks. For example, the legislation will allow piloting of new forms of vertically integrated provider organisations, thereby eroding the purchaser-provider separation established by the NHS changes of 1991. The proposals in the white paper also require local health authorities to develop primary care and regulate the pilot schemes at a time when they have lost many staff. To ensure equity of innovation pilots must emerge where primary care most needs improving and pilots must be evaluated thoroughly before they are applied widely.

Big bang or damp squib?

The launch of the primary care white paper, *Choice and Opportunity*,[1] last October was accompanied by a rash of newspaper headlines heralding a revolution in primary care provision. Reports said that pharmacy chains such as Boots and UniChem, American managed care organisations, and even the supermarket chains Tesco, Asda, and Safeway were considering offering primary care services.[2-5] Since October, concern has been growing about the consequences of the imminent break up of the general practitioners' professional monopoly and about the risks of allowing hospitals or commercial organisations to provide primary care.[6-8] Despite this, ministers have attempted to downplay the headlines and have concentrated on the likelihood of developments among existing primary care providers in the NHS.

ADRIAN TAYLOR/THE INKSHED

If the primary care bill (foreshadowed by *Choice and Opportunity*) receives parliamentary approval it will sweep away many of the existing legislative, budgetary, and procedural barriers to innovation. It will pave the way for experimenting with a variety of new organisational forms for the delivery of primary care, including general medical, dental, and community pharmaceutical services. For the first time, health authorities will be able to try suspending the national contract for general practitioners and directly commissioning general medical and other primary care services in ways that respond to local needs (box 1).

The government emphasises that some things will not change. The NHS will remain free at the point of use and will still be funded out of taxation. The right to enrol with an NHS family doctor will continue (see box 2). Opinions differ on whether this enabling legislation will change the pattern of service provision radically. Some commentators say that it could lead to something like the "big bang" deregulation of 1986 which dramatically transformed financial institutions in the City of London.[9] Others predict only minimal change because there will be no compulsion and little incentive for primary care professionals to join or form new types of organisation.[7] Stephen Dorrell, the secretary of state for health, hopes that around 7% of practices in the United Kingdom will be piloting new forms of organisation from April 1998.[10] This prediction, based on the experience of general practitioner fundholding, could be blown off course by an early general election. The Labour party, however, is not opposing the majority of the proposals, although it is unhappy about allowing commercial organisations to run primary care services.[11]

There are signs that a number of groups are keen to seize the opportunity to innovate. Some experiments are already up and running. In this sense the white paper is simply the recognition of established trends towards organisational and budgetary mergers. For example, the Lyme Community Care Unit, founded in April 1992, is a limited company set up and run by general practitioners in Lyme Regis, Dorset. The unit employs a wide range of community staff, including nurses, midwives, therapists, counsellors, care managers, social workers, and general practitioners, who provide integrated health and social care services, run a hospital at home scheme, and purchase secondary care.[12] The Andover Health Consortium links six fundholding practices and a community trust to provide and purchase health services for a population of around 100 000.[13 14] In Newcastle, a primary care trust already employs salaried general practitioners.[15] Other plans mooted before the white paper include plans by some community health trusts to employ

Box 1—New arrangements for organising delivery of primary care which could be piloted

- Whole practice contracts for primary care. Through these contracts a health authority could negotiate locally with individual general or dental practices for them to provide a specified range of general medical services and selected specialist services. Practices might employ community nurses, therapists, and hospital specialists on a sessional basis
- Joint working by general practices and community trusts (on contract to local health authorities) to provide a full range of primary care services, possibly including community hospital or hospital at home schemes. Community trusts and general practices might merge to form single, jointly managed organisations
- Community trusts contracted by health authorities to provide all primary care services. Trusts might employ salaried general practitioners or dentists, or subcontract to selected general practices
- Acute trusts contracted by health authorities to provide primary and secondary services—either the full range or specified services such as mental health
- Private or voluntary sector organisations contracted with the health authority to provide primary care services. These might include major retailers, drug companies, insurance companies, provident societies such as BUPA, or national charities
- Private or voluntary sector subcontractors providing disease management services to other primary or secondary care providers—for example, for asthma care
- Multiprofessional, not for profit, primary care organisations bidding for contracts to provide primary care. These might be set up and managed by health professionals, such as nurses or therapists, or general managers, within or outside the NHS
- Consortiums of general practices already in total purchasing pilot projects or locality commissioning groups which could take on more fully integrated budgets for providing or purchasing the complete range of primary and secondary care services. Such budgets could cover general medical services and hospital and community health services budgets, and include resources for fundholding

salaried general practitioners in inner city areas where recruitment has been difficult[16]; the purchase of a community hospital by a group of practices in south London[17]; joint ownership of practice premises by a cooperative of general practitioners, nurses, dentists and pharmacists[18]; management of an out of hours cooperative in Dorset by the local ambulance trust[19]; establishment of an American-style health maintenance organisation by a group of fundholders in Leicestershire[20]; and the proposed establishment by Premier Health Care NHS Trust of a subsidiary agency to supply salaried general practitioners, practice nurses, district nurses, and health visitors to practices all over Britain.[21]

Need for change

In their foreword to the white paper, the secretaries of state said that they wanted local people to develop services which will fit local needs and circumstances more closely: "We have no template in mind." [1] What they did have in mind was the need to deal with primary care's problems identified during the Department of Health's "listening exercise" (box 3).[22] This investigation emphasised that primary health care should be continuous, comprehensive, and properly coordinated to meet patients' needs; should be the gatekeeper to secondary care; and should respond to the health needs of communities as well as individuals. The catalogue of criticisms in box 3 does not present a balanced picture of the current state of primary care which, in many ways, is the great success story of the NHS. The criticism will be important, however, for those planning pilot projects, and it will remind them of the reasons why experimentation was considered necessary.

Risks of change

The risks involved in replacing small scale partnerships in general practice with larger, more bureaucratic organisations are immense, although many practices have now grown so large that informal teamworking is difficult. Community nurses will probably resist increased medical domination and direct employment by general practitioners, preferring to work for community trusts. Professional rivalries could intensify when the divisions inherent in the current system become more permeable—for example, if new primary care organisations encourage internal competition between doctors and nurses to undertake overlapping roles.

It is not clear whether any of the new services would be funded to augment existing services rather than replace them. For example, would patients be able to remain registered with their conventional general practitioner and also use the convenience of the NHS general practitioner or nurse service at the supermarket? If so, could the NHS afford the costs and ensure continuity and coordination of care? Administrative costs in primary care will probably increase, particularly if a range of for-profit providers enters the market. Profit-making managed care organisations in the United States are now reporting management and marketing costs in excess of 20% of turnover.[23]

On the other hand, change may bring considerable benefits. For example, wider choice in financing general practice might enable health authorities in inner cities to improve the availability and quality of services in places that have not benefited from earlier initiatives such as fundholding.

Monitoring and regulation

The white paper's rhetoric calls for diversity and local flexibility, stating that new arrangements should emerge from the field.[1] In comparison with the disastrous imposition of the revised contract for general practitioners in 1990, the new consultative approach seems more likely to gain acceptance among health professionals. The changes must be tightly regulated, not least to safeguard the values espoused in the other recent primary care white paper[24]—quality, fairness, accessibility, responsiveness, and efficiency—which imply the development of national standards in primary care.

Potential conflicts of interest could arise from group practices owning or controlling the community hospitals or specialist facilities to which they refer. Pharmaceutical companies offering disease manage-

Box 2—Key principles to guide future legislation on primary care provision in the NHS[1]

- Participation in pilots will be voluntary and existing contractual arrangements will continue if practitioners wish
- Patients will still have the right to register with a general practitioner of their choice to ensure continuity of care and personal service
- New arrangements for providing primary care will be piloted and evaluated before being implemented more widely
- The new arrangements will combine local flexibility with national safeguards for patients and practitioners
- Pilots will not be imposed from the centre but will emerge from experience in the field
- Public accountability for the use of funds and provision of services should be transparent and clearly maintained
- Reverting to previous arrangements for provision of primary care will be possible if pilots fail

ment services are unlikely to promote or use the products of rival firms. In addition, the new forms of vertically integrated organisation (providing both primary and secondary care) may, in the longer term, inhibit attempts to test out new and different sources and forms of provision. Acute trusts running primary care facilities could use their controlling interests to generate demand for their specialist services and may be tempted to squeeze resources out of primary care when the going gets tough in hospitals. It is hard to see how the claimed advantages of the purchaser-provider split can be maintained when the two functions are vested in the same primary care organisation. The government has refused to allow health authorities to employ general practitioners directly, presumably to safeguard purchaser-provider separation, but has not applied the same logic elsewhere.

Finally, what will happen to patient choice if merged organisations control all primary and community facilities in an area? For example, the Wakefield total purchasing pilot now covers the entire district; and proposals are currently being mooted, if the NHS (Primary Care) Bill becomes law, for the establishment of very large franchises of primary and secondary care organisations that would use private capital investment and control services throughout an entire urban area.

It looks as if much of the burden of accrediting, monitoring, and regulating the new arrangements will fall on local health authorities. They may not be able to do this adequately, especially if a large number of pilots are allowed to proceed. Health authorities are currently preoccupied with the annual contracting round and the pressure to reduce waiting times for hospital services. Most of the top jobs in the new merged and slimmed down authorities went to people with experience of secondary rather than primary care. Even those people who liaise with general practitioner fundholders or locality commissioning groups have been concerned mainly with developing the purchasing role of primary care rather than its providing role.

Box 3—Common problems in NHS primary care identified through the "listening exercise"[22]

- Variations in quality of care
- Failures of coordination between different agencies and professional groups
- Weak teamworking
- Lack of responsiveness of services to local needs
- Gaps in the provision of information for patients
- Barriers to developing the roles of nurses and other non-medical primary care professionals
- Inequitable distribution of primary care resources
- Poor quality premises and infrastructure in some areas, particularly inner cities
- Low morale and recruitment difficulties, particularly in inner cities
- Inflexibility in service provision due to rigid national practitioner contracts
- Limited opportunities for research and career development
- Lack of incentives to shift resources and services out of hospitals

Evaluation

The government has provided no details on how the secretary of state will approve proposed pilot schemes, or on how the pilot projects will be evaluated and by whom. It will be important to guard against inequity of innovation, with new initiatives happening only in better off areas where primary care is stronger (a feature of the early days of general practitioner fundholding). One guiding principle could be that the pilots should assist in levelling up the quality of primary care in disadvantaged areas. Pilots should be selected to allow the maximum opportunity to learn about the costs, risks, and benefits of alternative arrangements. When general practitioner fundholding was launched the government did nothing to encourage independent evaluation of its effects. As a result, many of the published studies were descriptive accounts of the process of becoming a fundholder which failed to ask searching questions about the value of the scheme or about the risks as well as the benefits.[25] With the extension of fundholding to total purchasing pilots, the government made a commitment to evaluation.[26] This was not sufficient, however to curb the rapid expansion of the number of pilot sites, which threatened to overwhelm scientific inquiry.[27]

Evaluation should assess the extent to which specified objectives have been achieved. Some type of quasi-experimental design with reference populations will be needed to assess the impact of the change or innovation, and explicit comparisons should be made between different ways of achieving the same ends. At the same time, the overall effects of deregulating and diversifying primary care should be monitored nationally. Recent years have seen major upheavals in the organisation of the NHS; isolating the effects of specific innovations is especially difficult in such turbulent times. Innovators and evaluators will have to agree criteria against which success or failure will be judged. They will also have to agree on a timescale for evaluation. Too short a timescale will run the risk of missing the true effects, too long will increase the likelihood that the innovation will diffuse before the results are available.

Conclusion

Supporting improvements in the new primary care agencies will require a "hands on" developmental approach, very different in nature from the arm's length approach encouraged by negotiating annually with secondary care providers. This will be especially necessary in inner cities, where the pressures of dealing with multiple health and social problems leave primary care professionals little time for dreaming up creative new approaches to organising their tasks. The latest white papers should serve to refocus attention on improving primary care services—not before time—but the success of the strategy will depend heavily on strong, imaginative, and responsive leadership by local health authorities, and this will be costly.

We thank Lucy Johnson, research librarian at the King's Fund, for providing summaries of reactions to *Choice and Opportunity* from the professional press and David Wilkin, Alan Maynard, and Jennifer Dixon for helpful comments on the draft paper.

1 Secretaries of State for Health, Scotland, and Wales. *Choice and opportunity. Primary care: the future.* London: Stationery Office 1996. (Cm 3390.)
2 Elliot A, Winchester K. How private companies reacted. *Pulse* 1996 October 26:7.
3 Hollinger P, Timmins N. UniChem in plan to run doctor services for NHS. *Financial Times* 1996 December 9:1.
4 Stirling A. Private firms scent GP bonanza. *Pulse* 1996 December 14:1.
5 Gould M. Why let reality spoil the cottage hospital dream? *Pulse* 1996 October 26:8.
6 Gosden T, Maynard A. He who pays the piper. *Health Service Journal* 1996 October 31:18.
7 Elliot A, Winchester K. Fundholding could be undermined. *Pulse* 1996 October 26:7.
8 Linden M, Mardell L. Opportunity or new danger? *Pulse* 1996 October 26:15.
9 Gould M. £100m big bang. *Pulse* 1996 October 26:1.
10 Elliot A. Dorrell lights the GP fuse. *Pulse* 1996 November 9:1.
11 Allan G, Webster V. Primary care: white paper and parliamentary debate. *Healthcare Parliamentary Monitor* 1996 November 4.
12 Robinson B. Primary managed care: the Lyme alternative. In: Meads G, ed. *Future options for general practice.* Oxford: Radcliffe Medical Press, 1996:201-10.
13 Edmonds R, Sloane R. The primary care agency. In Meads G, ed. *Future options for general practice.* Oxford: Radcliffe Medical Press, 1996:109-21.
14 Butler P. Trust and GPs may merge to form unique care agency. *Health Service Journal* 1996 October 17:5.
15 Crail M. Pilot light. *Health Service Journal* 1996 October 24:10-1.
16 Trusts to pilot salaried contracts [editorial]. *Pulse* 1996 November 2:6.
17 GPs want to buy cottage hospital [editorial]. *Pulse* 1996 November 9:2.
18 O'Dowd A. Nurses could be set to join GPs' co-op. *Pulse* 1996 November 9:15.
19 Ambulance trust to manage GP co-op [editorial]. *Pulse* 1996 November 9:15.
20 Elliot A. GPs are set to pilot US-style managed care deals. *Pulse* 1996 November 30:20.
21 Salaried GPs to be supplied by trust? [editorial]. *Pulse* 1996 November 9:4.
22 NHS Executive. *Primary care: the future.* Leeds: NHSE, 1996.
23 Reinhardt UE. Comment on the Jackson Hole initiative for a twenty-first century American health care system. *Health Econ* 1993;2:7-14.
24 Secretaries of State for Health, Scotland, and Wales. *Primary care: delivering the future.* London: Stationery Office, 1996. (Cm 3512.)
25 Coulter A. Evaluating general practice fundholding in the UK. *Eur J Public Health* 1995;5:233-9.
26 Department of Health. *Developing NHS purchasing and GP fundholding.* Leeds: NHSE, 1994. (EL(94)79.)
27 Mays N, Goodwin N, Bevan G, Wyke S, on behalf of the Total Purchasing National Evaluation Team. *Total purchasing: a profile of national pilot projects.* London: King's Fund Publishing, 1996.
28 Salmon JW. The corporatization of medicine: introduction. *Int J Health Services* 1987;17:1-6.

Pushing teeth

Fifty years ago when I was a medical student, I had a card with a long list of procedures to perform under supervision and get signed up by clinical tutors: reducing and plastering a Colles' fracture, performing a lumbar puncture, passing a catheter, and so on. At the end of the list was "extract two teeth." Why two? I never discovered the answer. But, like some of my classmates, I extracted a great many more. The dentist who supervised us had a sensationally beautiful assistant, and we went back again and again to gaze at her longingly and attempt to invite her to our parties (to no avail; she disdained medical students suffering from arrested adolescence). But thanks to her, I soon excelled at extracting teeth. The dentist knew well why I was there, and ensured that my time was not entirely wasted.

The secret of dental extractions is not to pull the tooth, but to push it out: push the wedge shaped points of the dental forceps well down alongside the roots of the tooth so the wedge loosens the roots; extraction then is usually easy.

This skill came in handy a few years later. I was hitching a ride across the world as ship's doctor on a Blue Star freighter. Shortly after we had rounded the Cape of Good Hope en route from London to Adelaide, the chief engineer, a curmudgeonly tyrant, bit savagely into a bread roll and broke the cusp off a bicuspid. He was in agony. I had to deal expeditiously and well with this acute dental emergency, or my life in this small, closed, and highly critical community would not have been worth living. The surgery on the ship was well equipped; I could have done a craniotomy or a destructive operation on a fetus. And of course there was a full set of dental forceps.

I premedicated the chief with a tumbler of brandy. The hard part was injecting local anaesthetic; even semistupefied, the chief did not like this part at all. The heavy seas of the Roaring Forties were no help, nor was my choice of a dental chair without adequate support for his head and neck.

The rest was easy. I carefully positioned the dental forcep blades beside the broken tooth, and pushed down as hard as I possibly could. The tooth popped up and out like a pea out of a pod. It was so simple I felt like clearing the rest of that side of his mouth while it was numb, but self restraint prevailed. For the rest of the voyage perhaps it was as well that my reputation for competence was not tested further. I have never again been called on to extract teeth. Pity, really. I think that I could have become a master of the art, thanks to those hours of unfulfilled dreams and useful experience in the dental clinic behind our teaching hospital.

John Last *is emeritus professor of epidemiology at the University of Ottawa*

We welcome filler articles of up to 600 words on topics such as *A memorable patient, A paper that changed my practice, My most unfortunate mistake,*or any other piece conveying instruction, pathos, or humour. If possible the article should be supplied on a disk.