Self-evaluation in general practice

THE BIRMINGHAM RESEARCH UNIT OF THE ROYAL COLLEGE OF GENERAL PRACTITIONERS

NEVER has it been so necessary or so difficult to measure the quality of medical care. The need arises from the enormous and ever escalating cost of achieving and maintaining the present level of care. The difficulty arises, paradoxically, from the general effectiveness of much current medical care when measured against any objective measurement of outcome.

The probable factors underlying this paradox are expressed in the smoothed histogram (Figure 1) which shows that in schematic form the cost rises exponentially as the theoretical or ultimate goal achievement is approached.

In the commercial world, the art of managing a business is to achieve an economic balance between costs and the quality of performance of the product. Such an ideal is represented by point B on the curve. If the manufacturer cuts costs too drastically, for example to point A on the curve, his product will be markedly and soon obviously inferior to that of the manufacturer operating at point B, with only marginal savings in the costs imposed.

Any manufacturer operating between B and C on the curve is achieving marginal increases in quality (i.e. performance or goal achievement) for disproportionate increases in costs. Only Rolls Royce and Mercedes Benz, for example, as minority producers selling status as well as quality, can afford to operate in this way.

Medical care is not subject to the same naked forces of supply and demand. The limits have been set more by the resources available at any given time rather than by any theoretical total supply of money available to individuals to pay for it. In other words, the system of medical care has been operating somewhere between C and D on the curve. The variations in quality between systems operating at say, points C and D, while real, are marginal and either immeasurable or difficult to measure. However, costs may vary slightly. Studies of cost effectiveness would show that the system operating at point C was the better 'buy' and not measurably worse than the system operating at point D. Hence, objective measurement of the quality of medical care is very difficult.

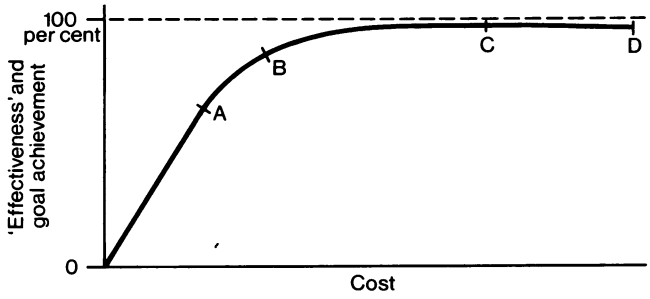
Defects of absolute measurements

In relatively primitive systems of medical care, the general mortality rates, as well as neonatal, perinatal, maternal mortality, and stillbirth rates can be used as efficient objective measures of 'outcome', but in the western world in recent years their usefulness has diminished as the general quality of medical care has risen.

All other rates, such as total and reported morbidity, absence from work and school, drug consumption, use of other therapeutic and diagnostic services, are all possible, but at the same time are indirect and blunt measures of the quality of care. These difficulties have led to a search for other possible objective measures of quality of care. These include the use of ideal protocols for critical areas of clinical and operational management as standards against which individual performance can be measured by comparison i.e. 'process audit'.

We may eventually establish consistently and generally agreed protocols for some forms of work, but, at present, this 'process auditing' has limited use outside hospitals. For example, the 'overuse' of antibiotics in the treatment of pharyngitis and the 'underuse' of throat swabs have been taken as a basis for a quantitative measure of quality of care (McFarlane and O'Connell, 1970). The presumptions and premises on which the measures of quality are based are debatable and an alternative set of measurements could be devised, using the same basic data, but where the implications about quality are almost the antithesis of those suggested.

Figure 1. *Schematic smoothed histogram showing cost rises in relation to quality.*



Spitzer *et al.* (1974) have shown that such ideal protocols for some indicator diseases may be used to infer a more general measure of the quality of clinical performance. However, the general use of this method is restricted not only by the limitations mentioned above, but by the need to keep secret the actual indicator conditions. The secrecy is dictated by the small number of suitable protocols.

Evaluation by peers in medical practice

Without techniques of evaluation for effectively identifying real advances and implementing them at the earliest possible moment, no physician can continue to provide the best medical care, however well he may have been trained initially (Hodgkin, 1973).

We suggest that there is at least one mechanism by which clinicians can avoid the traps of static unchanging inertia, blind acceptance of authority, or change for change's sake. This is by the development of a method of continuous *clinical and administrative self-audit or self-evaluation.* Self-evaluation by analysis of activities in the practice differs from conventional medical auditing in that the data used are not measured against any absolute (and therefore arbitrary) scales of quality, but are used solely to develop value judgements in the context of creative peer group discussions.

The essential basis for effective self-evaluation is a group of professionals who have enough time to discuss together to form a true peer group (Crombie, 1970). The essence of this self-evaluation is the ability of the members of such a peer group to accept the implicit and even explicit comments and criticism of their peers, without feeling threatened. At the same time each individual should have built up sufficient respect for each of his peers to ensure that he will curb the natural drive to point out indiscriminately idiosyncrasies, if not errors, in the performance of the others. Only under these circumstances will the individuals in the group feel able to reveal the details of their personal clinical performance. Without these basic conditions, no form of auditing can be implemented or fulfilled.

Physicians have traditionally worked in groups in all branches of medicine other than general practice. These groups include nurses and students as well as other doctors. It was the hospitals that until recently provided the basis for the creative activities of the medical profession, whether in pure research or in the evolution of clinical and teaching standards and methods (Crombie, 1963).

Not only does the peer group provide the basis for the negative functions of the imposition of minimum standards, but also it provides the basis for the creative activities which establish those standards in the first place (Crombie, 1975). The same need for the company and the approbation of their peers supplies part of the motivation for the creative activity (Darwin, 1874).

Group work provides multiple feedback through comment and rational criticism which leads to the rapid evolution of more effective methods in treatment and diagnosis, and better standards for judging the quality of these essential components of professional activity. This is simply the special form that creative, normative group behaviour takes in the context of medical care. In addition, any group discussing together for long enough will evolve its own general norms, attitudes, and values, as well as agreed plans for dealing with everyday problems of medical practice.

Need for factual information

In particular, for the evaluation of its own performance a peer group requires information, not in the form of an implicit or explicit directive, but ideally in a form which identifies and highlights differences. These may be differences between the way the group achieves or performs, compared with other comparable peer groups in the care system, or differences between the members of the group. The information must objectively and scientifically identify these differences but makes no value judgements about which is right or best.

Only occasionally can such value judgements be justified scientifically, but where they can their use is not only justified but mandatory, as ideal 'process'

Table 1. Estimated number of items per 1,000 population per year by group of drugs, for all prescriptions.

	Antidepressants	*Non-barbiturate sleeping tablets*	*Other sleeping tablets*	*Other psychotropic drugs*	*All other drugs*	*Total*
Doctor	All prescriptions					
A	478	13	70	297	3,186	4,044
B	289	77	153	527	5,878	6,926
C	278	113	251	468	3,353	4,463
D	189	94	163	417	3,379	4,245
All practice	324	71	158	407	3,615	4,575
Britain (1970)	142	158	291	382		5,600

Table 2. Estimated number of items per 1,000 population per year by group of drugs, for new and repeat prescriptions.

Doctor			New prescriptions			
A	268	—	—	51	1,724	2,043
B	136	—	34	170	3,471	3,812
C	165	41	110	234	1,951	2,501
D	86	34	17	172	2,032	2,342
All practice	176	21	42	151	2,083	2,473
Others	53	42	53	117	1,549	1,815
Total	141	27	45	141	1,932	2,287
Doctor			Repeat prescriptions			
A	210	13	70	246	1,462	2,001
B	153	77	119	357	2,407	3,114
C	113	72	141	234	1,402	1,962
D	103	60	146	245	1,347	1,903
All practice	148	50	116	256	1,532	2,102
Others	150	47	66	270	1,475	2,009
Total	149	49	97	261	1,510	2,067

protocols as suggested by McFarlane and O'Connell (1970). Occasionally some measure of performance or achievement can be immediately accepted by the whole group as evidence of quality of care. More often onerous criteria form the objective data for creative group discussion. If general practitioners were to insist that only those data which fulfil the first criterion would be used, then they would cut themselves off from the most powerful mechanism for the evolution of improved performance. All advances have their origins in such creative procedures.

A practical example

A practical example, which took place in our general practice was a programme designed to establish value judgements about the best way of using psychotropic drugs. Each doctor kept a carbon copy of each new prescription issued during a period of one week and each repeat prescription issued during a four-week period. Any prescriptions containing barbiturate sleeping tablets were separated from the rest and counted. The prescriptions containing non-barbiturate sleeping tablets were counted, along with those containing an antidepressant drug, those containing any other psychotropic drug (largely tranquillizers), and all other drugs. The number in each category was then converted to a prescription rate per 1,000 patients at risk for each doctor, if the total number of patients at risk was known or could be estimated.

The doctors then met to discuss the significance of differences, if any, between themselves and any other group who had carried out the same analysis.

The distribution of new and repeat prescriptions for the four partners in one group practice, expressed as rates per 1,000 patients at risk each year, is shown in the tables. Equivalent estimates, but only for new and repeat prescriptions combined, are included from national figures (Parish,1971). These figures were discussed by the four partners in the practice.

The session began with a consideration of the more general findings (Table 1). The comparison of the practice as a whole with the national figures suggested that the partners were using fewer barbiturates and more antidepressants than general practitioners as a whole, but the total rate for prescription of psychotropic drugs was remarkably similar. These totals hid two important sources of variation, interpartner differences and the differences between the new and repeat prescriptions.

Continuing with the principle of examining first the general, least personal, and therefore also the least threatening data, the group then turned its attention to the comparison of the new and repeat prescriptions (Table 2). Immediately this highlighted an unexpected finding. The total rates for prescriptions were evenly divided between new and repeat prescriptions. Half of the partnership prescribing load, and therefore costs, was for repeat prescriptions. In particular, this ratio applied to psychotropic drugs as a whole. Not only was this unexpected, it was also unsatisfactory.

There were other unsatisfactory features. While the rate for the initial prescription of barbiturates in particular, and all hypnotics and tranquillizers in general, was low and for antidepressants high, this was counterbalanced by the reverse situation among the

Table 3. Items per prescription.

Doctor	New prescription	Repeat prescription
A	1.2	1.7
B	1.0	2.3
C	1.4	1.8
D	1.2	1.2
All practice	1.2	1.6

repeat prescriptions. This differential was as marked for the individual doctors with low initial prescription rates for the antidepressants, as it was for those with the higher rates. The inference here was that vigilance and possibly an over-Draconian attitude to the problem of initiating treatment with psychotropic drugs did not prevent the accumulation, however slow, of an unacceptably high number of patients on long-term therapy.

These results are based on the patients registered with each doctor. On the whole each doctor was responsible for his own registered patients, but Dr B, who had recently joined the practice, had a small list of registered patients. It is obvious from Table 5 that Dr B was probably seeing some of the patients registered with the others and this had to be taken into account in interpreting the figures. In any extended practice analysis estimates of the workload of each doctor would already have been made systematically.

The debate which followed eventually led the partners to a radical reconsideration of the whole policy for management of repeat prescriptions. In particular, a much more intense but selective review of all repeat non-antidepressant psychotropic drug prescriptions was started at once.

The initial purpose of such a practice analysis had been to highlight differences between partners in the use of antidepressants on the one hand, and hypnotics and tranquillizers on the other. Such differences might have reflected different attitudes to and interpretation of depressive illness. In the event, this was the least important part of the audit, though it fulfilled its purpose. There was no previous judgement about who is right, but value judgements were evolved in the ensuing discussions. This is just one possible programme of practice analysis which can be mounted on prescriptions. A similar series of analyses can be carried out from a knowledge of differential or comparative prescription rates for: oral diuretics, hypotensives, hypoglycaemics, vitamin B_{12}, steroids (local and systemic), systemic antibiotics (in particular chloramphenicol), and asthma inhalers.

Table 4. Items per consultation.

Doctor	New prescription	Repeat prescription
A	0.8	0.9
B	0.7	0.9
C	0.9	0.9
D	0.8	0.7
All practice	0.8	0.8

Similarly, the referral pattern to social worker and other non-medical colleagues for admissions or outpatient care, use of clinicopathology, radiography, and other specialist diagnostic services, is available for this systematic analysis within a practice.

Family doctors, like their specialist colleagues, are increasingly using the case conference and review seminar as a basis for indirectly examining their different patterns of clinical care. In the best partnership practices there has aways been an equivalent tradition of constructive enquiry and criticism of the handling of shared clinical problems, but on a day-to-day rather than a formal basis. It is, in effect, a clinical audit and is one form of learning through a peer group.

The case conference uses clinical problems in general as the agenda. The practice analyses described here are systematically directed to highly specific aspects of clinical and operational work. They are mutually interdependent and complementary. The case conference has tended to emphasize the clinical problem from the point of view of disease processes. However, in recent years, under the influence of the family doctor, this emphasis has shifted to the needs of the patient, the family, and the community. The analysis of clinical care in a practice, as elaborated here, is primarily concerned with the performance of the doctor.

The potential of this system is greatly extended by including *comparative* performance and not confining it to conventional auditing against some arbitrary absolute scale of excellence set up by others outside the group. This does not preclude the development of suitable procedures by others for the use of such groups. If we are to extract the maximum advantage, the development of more procedures for analysis which can be constantly improved and extended by feedback and criticism from other peer group users is essential. The aim should be to develop gradually a library of accessible and appropriate practice analyses, constantly updated as the system of medical care evolves.

The results from such an analysis in a general practice carried out in a small group can be consolidated into an ever extending baseline, which each group will have in future for reference.

The characteristics of these programmes, including restraints as well as opportunities, are as follows:

1. They must measure consistently and in a standardized and reproducible form some clearly defined aspect of medical care.
2. The findings should be the basis for inferences about

Table 5. Items per 1,000 people at risk.

Doctor	New prescriptions per week	Repeat prescriptions per week
A	39.3	153.9
B	73.3	239.5
C	48.1	150.9
D	45.0	146.4
All practice	47.6	161.7

clinical performance such as the practice analysis of prescribing habits.

3. It must be possible to carry out the analysis in a service practice with minimum disturbance to day-to-day work.

4. Ideally, they should not involve additional recording by the practitioner, and the minimum additional load on ancillary staff. The simplest programmes are those which use data already generated for service purposes, for example, prescriptions, the 'billing forms' used in Canada, and appointment books.

5. They should each be completed in a reasonably short time, ideally within one week or one month.

6. The analysis of results should be within the competence of practice ancillary staff.

7. All definitions, instructions and criteria for the conduct of the study and the analysis should be available on short explicit instruction sheets.

8. The record should be short and as explicit as possible.

9. The consolidated results from various studies should be available to all users as results accumulate.

A systematic set of such practice analyses or programmes has been developed by the Birmingham Research Unit of the Royal College of General Practitioners in conjunction with the Department of Engineering Production of the University of Birmingham.

In the development of a systematic programme of practice analyses, the College can fulfil two important functions. It can design the systematic standardized programme in the first place, so that any group of doctors using it can be assured that it can be implemented economically and reliably in general practice with minimal involvement of the practitioner's own time. For example, the audit of prescriptions involved one of the practice secretarial staff in 17 hours of sorting, analyzing and preparation of tabulations, at a marginal cost to the practice of under £10 in 1975. The College can receive the results of the analyses in several practices and consolidate them to form a baseline of information.

The last function carries with it the implication that the College staff must ensure absolute confidentiality of such material and will disclose this information to a third party only with the express permission of the individual doctor.

Teaching, auditing and learning

Teaching and research ought to go hand in hand, and such comparisons between colleagues are a basic example of this symbiosis. It clearly demonstrates that curiosity, organized by these procedures, is basically a process of self-education.

There is also a need to find some basis from which general practitioners can teach one another. There is a need to break down the inhibitions which prevent general practitioners teaching. These inhibitions have also resulted from the continuing intellectual isolation of general practitioners compared with specialists. There is reasonable hope that these programmes will provide a basis for initiating teaching by general practitioners by stimulating their participation in a continuing process of debate, which will strengthen their confidence in their own work and standards.

These restraints do not apply to the postgraduate entering family practice after a training in any of the modern departments of general practice or family medicine. He should have been conditioned to systematic, organized, and effective evaluation of his clinical performance by others and the next step to the peer group discussion is taken easily.

In England and Wales the postgraduate centre in the district hospital has become a part of the general practitioner's way of life. The opportunity exists for the establishing of clinical auditing sessions to supplement the more traditional case conference. This could be of particular importance to the single-handed practitioner who otherwise has restricted opportunities for involvement in the peer group activities described here. Professional isolation is still probably the greatest single defect of the general-practitioner system (Crombie, 1963).

Such analyses, particularly if based on disease indexing systems and age-sex registers, must surely become an essential component of undergraduate and immediate postgraduate training. The systematic exploitation of these possibilities is now under way in the Department of Family Medicine in London, Ontario.

There are also the current political arguments about the control of the cost of medical care and whether doctors should be paid by salary, item of service, capitation fees or some other mixture. Without prejudging any of these issues, it is surely obvious that a system of self-evaluation by the doctors themselves, which could include operational and financial factors, as well as purely clinical factors, could meet many of the current legitimate complaints. In any case, no operational or financial audit is without some element of clinical evaluation, and vice versa.

Conclusions

Self-evaluation can be carried out in general practice by analyzing and comparing activity in the practice. This system is in contrast to ideas imported largely from North America. The North American model is *explicit.* It is *authoritarian,* and is based on *previously establishing an 'ideal' procedure* elaborated by teams of 'experts'. It is imposed from above on the rank and file and is backed up by the external threat of public loss of privileges.

This approach will always fail because:

1. The most important components of clinical performance and judgement, under the impact of ever changing ideas and techniques, are dynamic.

2. Quality of care, however measured, will always be relative. There are very few measures of quality of care which are universally agreed and applicable.

3. Individuals learn best by intrinsic auditing, that is, personal conviction following self-evaluation carried out in true peer groups through the motivation of peer group esteem or approbation.

4. Individuals learn least of all from imposed externally derived standards coupled with public threats to their self-esteem.

References

Crombie, D. L. (1963). *Lancet,* **i,** 209-211.
Crombie, D. L. (1970). *Journal of the Royal College of General Practitioners,* **19,** 66-78.
Crombie, D. L. (1975). *Update,* **11,** 517-522.
Darwin, C. (1874). *The Descent of Man.* 2nd edition. London: Murray.
Hodgkin, K. (1973). *Towards Earlier Diagnosis—a Guide to General Practice.* 3rd edition. Edinburgh: Churchill Livingstone.
McFarlane, A. H. & O'Connell, B. P. (1970). *Canadian Family Physician,* **16,** 53-55.
Parish, P. (1971). *Supplement No. 4,* **21,** No. 92. London: *Journal of the Royal College of General Practitioners.*
Spitzer, W. O., Sackett, D. L., Sibley, J. C., Roberts, R. S., Gent, M., Kergin, D. J., Hackett, B. C. & Olynich, A. (1974). *New England Journal of Medicine,* **290,** 251-256.

Social and financial effects of malignant reticuloendothelial disease

Eighty-five patients with lymphoma having active outpatient care replied to an oral questionnaire to determine the social and financial effects on their lives of their disease and its treatment. Difficulties included transport, inadequate explanation of their disease, and some loss of the general practitioner/patient relationship.

Reference

Laird, H. J. & Dawson, A. A. (1977). *Health Bulletin,* **35,** 25-29.