Stress and peptic ulcer: life beyond helicobacter

Susan Levenstein

Gastroenterology Service, Nuovo Regina Margherita Hospital, Viale Trastevere 72, Rome, Italy

Susan Levenstein, *adjunct research physician*

Correspondence to: Dr Levenstein, Via del Tempio 1A, 00186 Rome, Italy
100424.3254@compuserve.com

BMJ 1998;316:538–41

The discovery that *Helicobacter pylori* is a cause of peptic ulcer has tempted many to conclude that psychological factors are unimportant. But this is dichotomised thinking. There is solid evidence that psychological stress triggers many ulcers and impairs response to treatment, while helicobacter is inadequate as a monocausal explanation as most infected people do not develop ulcers. Psychological stress probably functions most often as a cofactor with *H pylori*. It may act by stimulating the production of gastric acid or by promoting behaviour that causes a risk to health. Unravelling the aetiology of peptic ulcer will make an important contribution to the biopsychosocial model of disease.

For this review of the role of psychological stress in the aetiology of peptic ulcer disease, I undertook conventional journal tracking and reference tracing, supplemented by Medline searches using Paperchase. The important keywords used in this search included peptic ulcer; duodenal ulcer—psychology; stress; life change events; and personality.

Summary points

- Psychosomatic factors in the aetiology of peptic ulcer have become unfashionable since the discovery of *Helicobacter pylori*
- Most people harbour *H pylori* so the organism cannot serve as the sole explanation for ulcer disease
- Psychological stress has an impact on the onset and course of ulcer disease
- Psychological stress probably interacts with *H pylori* and other risk factors in causing ulcer disease
- Peptic ulcer is an important example of the biopsychosocial model of disease

H pylori is not enough

When *H pylori* burst on the scene a few years ago, it revolutionised views on the aetiology and treatment of peptic ulcer. Psychosocial factors were quietly but firmly escorted off the stage, and gastroenterologists in particular banished psychological considerations with something approaching relief.

While this surge of biological reductionism is understandable, it risks throwing the baby out with the bath water. *H pylori* is inadequate as a sole explanation for peptic ulcers. Most people who harbour the organism never have ulcers, while a few who have never been infected with it or taken non-steroidal anti-inflammatory drugs develop ulcer disease.[1] This testifies to the role of factors additional to infection in peptic ulceration (the situation is similar for *Mycobacterium tuberculosis* and, indeed, most infectious organisms). *H pylori* alone does not explain fully the epidemiological patterns of upper gastrointestinal disease.[2] In addition, Marshall (who first identified *H pylori* on the surface of the gastric antrum) found that Koch's postulates for establishing the aetiological relation between a micro-organism and disease are not met in the case of *H pylori* and ulcers.[3] Again, the same might be said for many other established causes of disease.

Explaining the aetiology of a disease is often like assembling a jigsaw puzzle. By the time *H pylori* had been discovered, many pieces of the ulcer puzzle, from cigarette smoking to type O blood, had already been found and fitted together, although we did not know exactly where they belonged in the larger picture. The subsequent discovery of an important and central piece may mean that other completed sections have to be moved around—not that they have to be discarded.

The discovery of *H pylori*, far from eliminating interest in that older assortment of physical risk factors for ulcer disease, has spurred some researchers to re-establish and reinterpret the importance of older factors in the new context. Clinical studies have examined these physical risk factors in relation to their association with or independence from *H pylori* in ulcer patients.[1 4] Other workers are combing large databases to see how these same factors relate to *H pylori* in the general population.[5 6]

Why are psychological factors being ignored?

Methodological inadequacies

Why has there been no corresponding flurry of articles on the relation between *H pylori* and psychology?[7] For one thing, researchers interested in psychosocial factors tend to have little interest or expertise in microbiology, and vice versa. But I suspect the reason is deeper. The current generation of gastroenterologists has lost patience with psychosomatic explanations, at

least where these apply to peptic ulcer. The methodological inadequacies of some of the older published reports on psychosocial factors in ulcer disease are partly to blame. These studies were often flawed by contamination from disease effects, diagnostic uncertainty, and recall bias; did not pay enough attention to adjustment for confounding factors; and sometimes left readers distracted by side issues such as the nature or the measurement of stress.

Resistance to an integrated approach

The main feeling, however, seems to be that psychosomatic reasoning can be discarded as soon as another explanation becomes available. This dichotomised thinking, focused on the possibility of moving peptic ulcer from a stigmatised "psychosomatic" cubbyhole into a more dignified "infectious" one, reflects an ingrained resistance to the difficult but essential task of examining disease aetiology in an integrated manner that incorporates both psychological and biomedical elements[8]—a task begun for peptic ulcer long ago.[9]

Negative case-control studies

In fact, while all references to psychology have been removed from the chapter on ulcers of the 20th (1996) edition of Cecil and Loeb,[10] this is not solely due to *H pylori*. In the 17th edition (1982), published before the discovery of helicobacter, 18 lines were allotted to psychology, but all were dismissive. Around that time a spate of negative case-control studies was sabotaging interest in psychosomatic factors in ulcer disease.[11] One reason for the difficulty these studies had in showing a causal role for psychosocial factors is that patients with peptic ulcer are a heterogeneous group. Stress is probably an active factor in only some patients—perhaps those with less exposure to ulcerogenic substances[12] or those with higher pepsinogen concentrations.[13]

Stress is fashionable elsewhere

Ironically, while the gastroenterological community seems to view those who have continued to support a psychosomatic aetiology for ulcer disease as stubbornly clinging to obsolete views, psychosocial factors have the glitter of novelty for researchers in other specialties, who are happily exploring whether you can die of fright or get the sniffles from stress.

Stress and ulcers: the evidence

Since the 1980s there has been a modest resurgence of research interest in the ulcer-stress question, and the resulting body of evidence has generally, though not always, been supportive.[14] A solid series of methodologically sound studies now supports an aetiological effect of aspects of "stress" (a convenient term to cover both life stressors and subjective distress) ranging from depression to war. Considerable prospective evidence has been gathered alongside studies that have found an excess of life stressors in ulcer patients compared with matched controls.[15] In one large longitudinal population study, the occurrence of self reported ulcer over a nine year period was more likely in subjects who reported any of several concrete life stressors or psychological distress at baseline.[16 17] In another study, self reported stress predicted the occurrence of ulcer disease (diagnosed by a doctor) over the next 13 years.[18] Family and job difficulties increased the risk of ulcer over five years in one sample of Israeli men.[19] Major societal disasters, including German air raids in London and economic collapse in Sofia, have been associated with documented increases in acute ulcers.[20 21] The ulcerogenic effects of stress have been shown to be robust enough to survive adjustment for behavioural and physical confounding factors.[16–18]

Stress and prognosis in ulcer disease

Prospective clinical studies have reported that psychological distress impedes ulcer healing (as seen with endoscopy) after H_2 agonist treatment—even when stringent recruitment criteria exclude confounding in relation to disease chronicity.[22 23] Life stress continued to worsen the prognosis over several years in two prospective case series.[23 24] This effect seems to be reversible, however: a psychologically stable person who develops an ulcer during a stressful period is likely to remain free of symptoms for years after a short course of treatment, even without medication to eradicate *H pylori*.[24]

Psychological stress and gastric acid secretion

Psychological stress is not only empirically associated with ulcers, but is a very plausible risk factor for ulcer disease. Gastric acid output is correlated with psychological distress in patients with and without ulcers,[25] and increased enormously during intense military training.[26] Compared with healthy people, patients with duodenal ulcers are particularly likely to respond to laboratory stressors by secreting more acid.[27] In two patients with duodenal ulcers, extraordinary life circumstances resulted in a 10-fold to 20-fold increase in the basal acid output.[28]

Under stress, the amount of acid reaching the duodenum may increase further because gastric motility has changed or meals have been missed. People affected by stress may also smoke more, sleep less, and take more non-steroidal anti-inflammatory drugs, thereby increasing their susceptibility to ulcer by mechanisms that are not related to acidity. Evidence that prospective epidemiological associations between life stressors or psychological characteristics and ulcer are reduced by adjusting statistically for patterns of eating, sleeping, and substance use support a mediatory role for these health risk behaviours in the psychogenesis of ulcers.[16 17]

DOLLY THOMPSETT

Stress and *H pylori* as cofactors in ulcer disease

- Stress could facilitate the evolution of *H pylori* infection into ulcer by producing gastric hyperchlorhydria[31]
- Stress could disturb the equilibrium between *H pylori* and its host via psychoneuroimmunological mechanisms
- Stress could reduce mucosal defences to *H pylori* invasion through behavioural mediators such as cigarette smoking[4]
- Stress could increase the chances of ulceration in duodenal mucosa that have already been weakened by the effects of *H pylori* infection simply by increasing the acid load which flows past
- Stress induced acid secretion could promote *H pylori* colonisation of the duodenal bulb by neutralising the inhibitory effect of bile[32]

An association between stress and *H pylori* infection?

Despite empirical support and biological plausibility, there could in theory still be a reason for abandoning psychogenic causes of ulcer disease. Stress might be associated strongly with *H pylori* infection, inducing the illusion that psychosocial factors can influence ulcer formation. For many of the other classic risk factors, this possibility has been examined directly and discarded. Smoking, alcohol consumption, non-steroidal anti-inflammatory drugs, and type O blood are not correlated with *H pylori* in populations without ulcers.[5 6] The story is somewhat different for socioeconomic status. An excess of *H pylori* infection in poor people[5] suggests that low social class may partly owe its reputation as an ulcer risk factor to confounding by *H pylori*.[29]

No direct data are available as yet for stress, but there is no reason to expect that infection with *H pylori* is related to any particular psychological state. It is true that life stress, psychological distress, and *H pylori* may all be associated with low socioeconomic status, but epidemiological evidence argues against socioeconomic status being a principal cause of the association between stress and ulcers.[17]

How do *H pylori* and psychosocial stress interact in ulcer disease?

If we conclude that the discovery of *H pylori* has not erased the role of psychosocial risk factors in ulcer disease, asking how the two interact becomes logical. In some cases there may be no interaction at all. Given the large increase in acid production under severely stressful conditions[26 28] and the high secretion of acid[30] and pepsinogen[1] in patients whose duodenal ulcers are not related to either *H pylori* or non-steroidal anti-inflammatory drugs, stress may be capable of causing peptic ulceration even in the absence of *H pylori*.[12] In most cases, however, stress probably functions as a cofactor with *H pylori*. Several possible mechanisms can be postulated (box).

Empirical evidence

Again, little empirical evidence exists on the relation between stress and *H pylori*, and that is limited to studies of psychological distress in gastroenterology patients. In subjects with undiagnosed dyspepsia, those who are depressed or anxious are less commonly infected with *H pylori*.[33] Similarly, in patients with documented peptic ulcer, the higher the absolute titre of *H pylori* IgG antibody (a rough measure of the intensity of bacterial colonisation of the antrum), the less anxious is the patient likely to be.[12] These findings suggest that *H pylori* infection and psychological stress promote ulcer pathogenesis via pathophysiological pathways that are largely additive and therefore independent and complementary rather than synergistic.[34] Two different mechanisms of facilitating acid damage to the duodenum would result in a statistically inverse relation in ulcer patients, though not necessarily in the general population

Inverse relation between psychological stress and *H pylori*

In the classic view of the pathogenesis of peptic ulcer disease—that it is the sum of factors that increase acid secretion and those that lower mucosal defences—psychological stress probably acts mostly on the side of increased aggression and *H pylori* on the side of weakened defences.[35 36] *H pylori* infection does not necessarily raise, and may even lower, acid secretion.[30 34] Thus, though it has been suggested that the organism contributes to increased acid in the duodenum[37] and to raised pepsinogen values,[38] the striking increases in gastric acid found in many patients with duodenal ulcer seem to result from mechanisms that are at least partly independent of *H pylori* infection.[1 30]

Other possible explanations exist for an inverse relation between anxiety and *H pylori* antibody titres in ulcer patients. Stress may, for example, suppress serum antibody titres by stimulating cortisol production. Local processes in the gastric mucosa could also contribute, since a low pH tends to cause antigen and antibody to dissociate. Gastric hyperacidity resulting from stress could therefore suppress mucosal immunity, including *H pylori* antibody formation.

The way ahead

But all this is speculation. Now that the helicobacter earthquake has passed, those of us interested in the effects of psychosocial factors on peptic ulcer must begin investigating again in the new context. This will require scrupulous use of a variety of research tools if we are to have any impact on the many colleagues who think our efforts are a waste of time. We need to be careful to use appropriate measures in examining psychosocial data and to embrace the complexities of the biopsychosocial model. We need to pay great attention to avoiding recall bias, contamination of epidemiologically defined "ulcer" groups by subjects with non-ulcer dyspepsia, and statistical confounding by socioeconomic status.

Far from being obsolete, the concept that psychosocial factors play a role in peptic ulcer presents exciting and varied research opportunities in the age of *H pylori*. We can study the clustering between stress indicators and *H pylori* in the general population, we can look at the relation between *H pylori* and stress in case-control and prospective studies of ulcer patients, we can bring people whose *H pylori* status is known into the physiology laboratory and look at their physical reactions to stress, we can extract data on

H pylori antibody titres from serum samples frozen for ongoing longitudinal studies, and we can study interactions of *H pylori* and stress in animal models. These kinds of investigations can serve not only to breathe new life into psychosomatic concepts of peptic ulcer, but also to develop a more general paradigm for applying the integrated biopsychosocial model to medical disorders both infectious and otherwise.

I am indebted to Douglas Drossman, Harold Bourne, and Cosimo Prantera for their encouragement and for many invaluable exchanges of ideas.

Funding: None.

Conflict of interest: None.

lp;01 McColl KEL, El-Nujumi AM, Chittajallu RS, Dahill SW, Dorrian CA, El-Omar E, et al. A study of the pathogenesis of Helicobacter pylori negative chronic duodenal ulceration. *Gut* 1993;34:762-8.
2 Sonnenberg A. The US temporal and geographic variations of diseases related to Helicobacter pylori. *Am J Public Health* 1993;83:1006-10.
3 Marshall BJ. Helicobacter pylori in peptic ulcer: have Koch's postulates been fulfilled? *Ann Med* 1995;27:565-8.
4 Martin DF, Montgomery E, Dobek AD, Patrissi GA, Peura DA. Campylobacter pylori, NSAIDs, and smoking: risk factors for peptic ulcer disease. *Am J Gastroenterol* 1989;84:1268-72.
5 Graham DY, Malaty HM, Evans DG, Evans DJ, Jr., Klein PD, Adam E. Epidemiology of Helicobacter pylori in an asymptomatic population in the US: effect of age, race, and socioeconomic status. *Gastroenterology* 1991;100:1495-501.
6 Loffeld RJ, Stobberingh E. Helicobacter pylori and ABO blood groups. *J Clin Pathol* 1991;44:516-7.
7 Melmed RN, Gelpin Y. Duodenal ulcer: the helicobacterization of a psychosomatic disease? *Isr J Med Sci* 1996;32:211-16.
8 Drossman DA. Gastrointestinal illness and the biopsychosocial model. *J Clin Gastroenterol* 1996;22:252-4.
9 Mirsky A. Physiologic, psychologic, and social determinants in the etiology of duodenal ulcer. *Am J Dig Dis* 1958;3:285-313.
10 Bennett JC, Plum F, eds. *Cecil textbook of medicine*. 20th ed. Philadelphia: Saunders, 1996.
11 McIntosh JH, Nasiry RW, McNeil D, Coates C, Mitchell H, Piper DW. Perception of life event stress in patients with chronic duodenal ulcer: a comparison of the rating of life events by duodenal ulcer patients and community controls. *Scand J Gastroenterol* 1985;20:563-8.
12 Levenstein S, Prantera C, Varvo V, Scribano ML, Berto E, Spinella S, et al. Patterns of biologic and psychologic risk factors for duodenal ulcer. *J Clin Gastroenterol* 1995;21:110-7.
13 Walker P, Luther J, Samloff IM, Feldman M. Life events stress and psychosocial factors in men with peptic ulcer disease. II Relationships with serum pepsinogen concentrations and behavioral risk factors. *Gastroenterology* 1988;94:323-30.
14 Jess P. Personality pattern in peptic ulcer disease: a cohort study. *J Intern Med* 1994;236:271-4.
15 Gilligan I, Fung L, Piper DW, Tennant C. Life event stress and chronic difficulties in duodenal ulcer: a case control study. *J Psychosom Res* 1987;31:117-23.
16 Levenstein S, Kaplan GA, Smith M. Sociodemographic characteristics, life stressors, and peptic ulcer: a prospective study. *J Clin Gastroenterol* 1995;21:185-92.
17 Levenstein S, Kaplan GA, Smith MW. Psychological predictors of peptic ulcer incidence in the Alameda County study. *J Clin Gastroenterol* 1997;24:140-6.
18 Anda RF, Williamson DF, Escobedo LG, Remington PL, Mast EE, Madans JH. Self-perceived stress and the risk of peptic ulcer disease: a longitudinal study of US adults. *Arch Intern Med* 1992;152:829-33.
19 Medalie JH, Stange KC, Zyzanski SJ, Goldbourt U. The importance of biopsychosocial factors in the development of duodenal ulcer in a cohort of middle-aged men. *Am J Epidemiol* 1992;136:1280-7.
20 Spicer CC, Stewart DN, Winser DMR. Perforated peptic ulcer during the period of heavy air raids. *Lancet* 1944;i:14.
21 Pomakov P, Gueorgieva S, Stantcheva J, Tenev T, Rizov A. Ulcères gastro-duodenaux pendant la periode d'une crise economique aigue. *J Radiol* 1993;74:265-7.
22 Levenstein S, Prantera C, Scribano ML, Varvo V, Berto E, Spinella S. Psychologic predictors of duodenal ulcer healing. *J Clin Gastroenterol* 1996;22:84-9.
23 Holtmann G, Armstrong D, Pöppel E, Bauerfeind A, Goebell H, Arnold R, et al. Influence of stress on the healing and relapse of duodenal ulcers. *Scand J Gastroenterol* 1992;27:917-23.
24 Levenstein S, Prantera C, Varvo V, Arcà M, Scribano ML, Spinella S, et al. Long-term symptom patterns in duodenal ulcer: psychosocial factors. *J Psychosom Res* 1996;41:465-72.
25 Feldman M, Walker P, Goldschmiedt M, Cannon D. Role of affect and personality in gastric acid secretion and serum gastrin concentration: Comparative studies in normal men and in male duodenal ulcer patients. *Gastroenterology* 1992;102:175-80.
26 Oektedalen O, Guldvog I, Opstad PK, Berstad A, Gedde-Dahl D, Jorde R. The effect of physical stress on gastric secretion and pancreatic polypeptide levels in man. *Scand J Gastroenterol* 1984;19:770-8.
27 Bresnick WH, Rask-Madsen C, Hogan DL, Koss MA, Isenberg JI. The effect of acute emotional stress on gastric acid secretion in normal subjects and duodenal ulcer patients. *J Clin Gastroenterol* 1993;17:117-22.
28 Peters MN, Richardson CT. Stressful life events, acid hypersecretion, and ulcer disease. *Gastroenterology* 1983;84:114-9.
29 US Department of Health and Human Services. *Current estimates from the national health interview survey: United States, 1983*. Washington, DC: US Department of Health and Human Services, 1986
30 Chandrakumaran K, Vaira D, Hobsley M. Duodenal ulcer, Helicobacter pylori, and gastric secretion. *Gut* 1994;35:1022-6.
31 Lee A, Dixon MF, Danon SJ, Kuipers E, Megraud F, Larwsson M, et al. Local acid production and Helicobacter pylori: a unifying hypothesis of gastroduodenal disease. *Eur J Gastroenterol Hepatol* 1995;7:461-5.
32 Han SW, Evans DG, el-Zaatari FA, Go MF, Graham DY. The interaction of pH, bile, and Helicobacter pylori may explain duodenal ulcer. *Am J Gastroenterol* 1996;91:1135-7.
33 Andersson SI, Hovelius B, Mölstad S, Wadström T. Dyspepsia in general practice: psychological findings in relation to Helicobacter pylori serum antibodies. *J Psychosom Res* 1994;38:241-7.
34 Peterson WL, Barnett CC, Evans DJ, Jr, Feldman M, Carmody T, Richardson C, et al. Acid secretion and serum gastrin in normal subjects and patients with duodenal ulcer: the role of Helicobacter pylori. *Am J Gastroenterol* 1993;88:2038-43.
35 Goggin PM, Marrero JM, Spychal RT, Jackson PA, Corbishley CM, Northfield TC. Surface hydrophobicity of gastric mucosa in Helicobacter pylori infection: effect of clearance and eradication. *Gastroenterology* 1992;103:1486-90.
36 Hogan DL, Rapier RC, Dreilinger A, Koss MA, Basuk PM, Weinstein WM, et al. Duodenal bicarbonate secretion: eradication of Helicobacter pylori and duodenal structure and function in humans. *Gastroenterology* 1996;110:705-16.
37 El-Omar E, Penman ID, Ardill JES, Chittajallu RS, Howie C, McColl KEL. Helicobacter pylori infection and abnormalities of acid secretion in patients with duodenal ulcer disease. *Gastroenterology* 1995;109:681-91.
38 Hunter FM, Correa P, Fontham E, Ruiz B, Sobhan M, Samloff IM. Serum pepsinogens as markers of response to therapy for Helicobacter pylori gastritis. *Dig Dis Sci* 1993;38:2081-6.

(Accepted 21 July 1997)

Fifty years ago
The new NHS

SIR,— A crucial question at this time which I think is being forgotten is, Do the public gain in health through the introduction of the Service at this stage?

The answer is obviously " No." The Service will call for more work among doctors owing to more visits by patients and the inevitable form-filling which is the necessary evil of any coordinated service. If we look upon the patients as the material with which we work, and whose medical well-being is our "finished job," then I'm afraid we're being forced on them at a time when workmen and material are of poor quality. We should have workshops (hospital beds)—the present numbers are inadequate; also assistants (nurses)—whose number is also inadequate.

A little variation in available foods may brighten a few, in mind if not in body. Tuberculous-infected milk can still take its toll of children, and tuberculous patients still linger at home waiting for hospital beds. However bad these are, and however good are the excuses for not remedying them now, the introduction of a health service, which to the patient promises improvement in medical care, can do nothing—in fact, we could be made the scapegoat for its failure.

Let us accept the role of heaven-sent advisers to Aneurin Bevan and tell him gently "No," or at least "Not yet."—I am, etc, H. J. Houghton, Newport, Mon.

(*Letter*, 17 January 1948, p 122. See also editorial by Gordon Macpherson, 3 January 1998, p 6.)

What are quality of life measurements measuring?

Matthew F Muldoon, Steven D Barger, Janine D Flory, Stephen B Manuck

Center for Clinical Pharmacology, University of Pittsburgh, Pittsburgh, PA 15260, USA
Matthew F Muldoon, *assistant professor*

Department of Psychiatry, School of Medicine, University of Pittsburgh
Steven D Barger, *postdoctoral fellow*

Department of Psychology, University of Pittsburgh
Janine D Flory, *research assistant professor*
Stephen B Manuck, *professor*

Correspondence to: Dr Muldoon
mfm10 + @pitt.edu

BMJ 1998;316:542–5

It is now widely acknowledged that the personal burden of illness cannot be described fully by measures of disease status such as size of infarction, tumour load, and forced expiratory volume. Psychosocial factors such as pain, apprehension, restricted mobility and other functional impairments, difficulty fulfilling personal and family responsibilities, financial burden, and diminished cognition must also be encompassed. The area of research that has resulted from this recognition is termed "health related quality of life." It moves beyond direct manifestations of illness to study the patient's personal morbidity—that is, the various effects that illnesses and treatments have on daily life and life satisfaction. Although quality of life assessment was almost unknown 15 years ago, it has rapidly become an integral variable of outcome in clinical research; over 1000 new articles each year are indexed under "quality of life."

Although the importance of quality of life is broadly acknowledged, scepticism and confusion remain about how quality of life should be measured and its usefulness in medical research. These responses may reflect important conceptual and methodological limitations of the current concept of quality of life. We offer a simple framework that describes the core elements of quality of life related to health and use this to evaluate quality of life measurement as it is currently conducted.

Summary points

- Measures of disease status alone are insufficient to describe the burden of illness; quality of life factors such as pain, apprehension, depressed mood, and functional impairment must also be considered
- Two operational definitions of quality of life are identified—objective functioning and subjective wellbeing
- Assessments of objective functioning and subjective wellbeing convey different information, they also present different problems in relation to validation
- Assessment of functioning derived from questionnaires must be validated against measures of directly observed behavioural performance
- Subjective appraisal of wellbeing may be influenced substantially by psychological factors unrelated to health or to changes over time in patients' criteria for appraising wellbeing
- Whether and how quality of life researchers respond to these obstacles and deficiencies will probably determine the quality of their work in the future

A simple classification scheme for measuring quality of life

Division into functional status and subjective wellbeing

While there is neither a precise nor agreed definition of quality of life, quality of life research seeks essentially two kinds of information, the functional status of the individual and the patient's appraisal of health as it affects his or her quality of life. In addition, current questionnaires used in quality of life assessments generally embody one or both of the following operational definitions—quality of life as an individual's behaviour or level of functioning or quality of life as an individual's perceived health status or wellbeing. Measuring someone's ability to perform common tasks or activities is putatively objective, while asking patients to rate the effects of health status on personal wellbeing is explicitly subjective. For example, the question "Are you able to carry two bags of groceries 20 yards?" seeks explicitly behavioural information, whereas "Does your health interfere with your enjoyment of life?" invites respondents to make subjective ratings.

Most early measures of health status,[2] as well as some contemporary quality of life instruments,[3] were designed to measure objectively the adequacy of individuals' functioning across life's various domains—physical, occupational, and interpersonal. Published reports describing these particular instruments often use the terms health status, functional status, and quality of life interchangeably. Other instruments define quality of life in an inherently subjective way; for example, they include questions that ask how disabled the patient feels.

Division of health into physical and mental domains

Dividing health into physical and mental domains provides some further structure for understanding the effects of health status on quality of life.[4] The figure shows that assessing physical functioning (top left) involves measuring the ability to perform specific tasks (for example, activities of daily living or climbing stairs) as well as less easily defined concepts that are related to role (for example, the ability to continue employment as a carpenter).[5] In many respects, measurement of physical functioning is similar to assessment of physical disability. Mental functioning (figure, bottom left) is reflected in the patient's ability to rise to life's cognitive and social challenges, ranging from specific tasks (for example, balancing a cheque book) to complex social interactions (such as presenting a departmental productivity report at a business meeting).

Importance of subjective appraisal of health

The alternative, or complementary, perspective on quality of life assigns central importance to an

individual's subjective appraisal of their state of health. This definition presumes that quality of life is at least partly independent of health status,[6] and "is a reflection of the way that patients perceive and react to their health status and to other non-medical aspects of their lives."[7] The subjective nature of this conceptualisation of quality of life is perhaps best understood as focusing on how ill or disabled patients say they feel in the context of their personal lives, as distinct from external attempts to quantify stage or degree of illness or disability. Physical wellbeing (figure, top right) concerns the sense of discomfort arising from a particular symptom (or freedom from such), and extends to vitality or general satisfaction with physical health. A patient's appraisal of his or her mental wellbeing (figure, bottom right) is usually interpreted as the absence of psychological distress (that is, anxiety, depression, anger, etc) and can also include emotional ties and social support.[8]

Objective functioning should be distinguished from subjective wellbeing

All quality of life questionnaires purport to assess objective functioning, subjective wellbeing, or both. However, investigators have been reluctant to deal with the distinction between objective functioning and subjective wellbeing, partly because of controversy about the relative importance of these two ways of looking at quality of life. We believe that these approaches are both important, and that applying the classification scheme described above would make their definition clearer and more precise. Naturally, precision and clarity are also served by the investigators specifying the domains of quality of life that are of interest in each study.[1] Confusion also arises because many quality of life instruments produce composite indices. These combine information from numerous questionnaire items that span various domains (for example, working compared with home or family life) and include ratings of both functioning and subjective wellbeing. Composite indices have been criticised for failing to recognise that quality of life is inherently multidimensional.[9] Furthermore, some questionnaire items concern well defined behaviour or levels of functioning while others focus on subjective health appraisal, and we believe that aggregating these kinds of information is essentially illogical. By analogy, in the study of heart disease, measures of coronary stenosis and exercise tolerance are important and closely related to one another, yet actually combining these measures makes little sense.

Questions of validity

Criterion validity

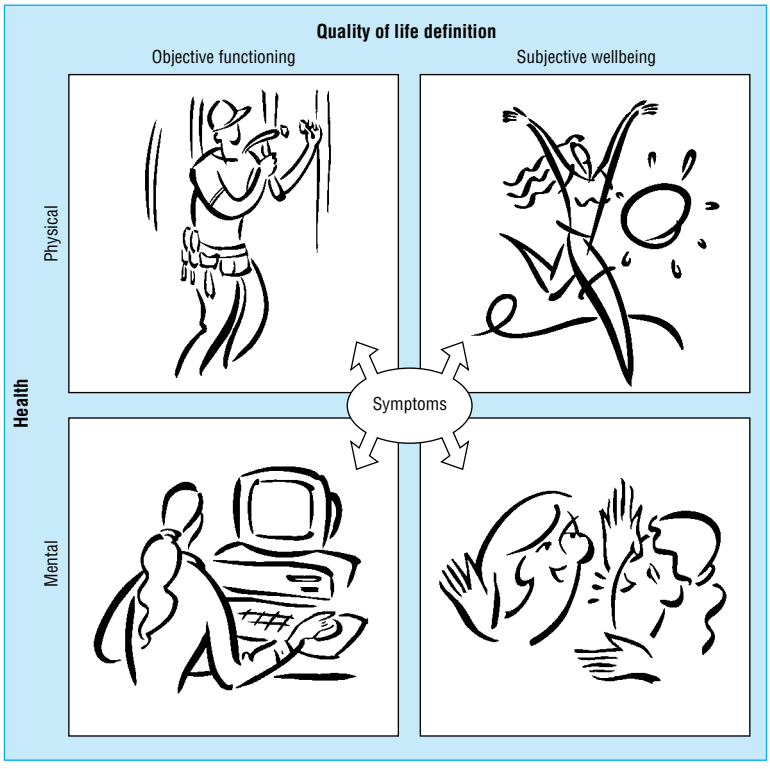
The value of quality of life questionnaires in medical research rests squarely upon their validity, and physicians cannot interpret quality of life measures until the instruments being assessed are adequately established. While validity can be examined in several ways, comparison with the best indicator available (criterion validity) is the preferred method. In evaluating quality of life measures of functioning, self reported physical abilities should correlate closely with behavioural performance that is defined objectively and measured directly. For example, in patients with Parkinson's disease, self reported scores for mobility should be compared with objective testing of walking, turning, and rising from the seated position. With few exceptions, however, little or no such validation exists for most quality of life measures of physical functioning.[10,11]

Construct validity

Once we move beyond physical functioning (figure, top left), yardsticks are generally not available. However, we can, and should, examine the construct validity of quality of life questionnaires using two complementary evaluations.[12] The first of these is for convergent validity—the degree to which questionnaire scores correlate with self report data from established instruments measuring similar things and with the same construct assessed with different methods (for example, rated by a doctor or spouse). Low scores on a quality of life scale of psychological wellbeing, for example, should predict high scores on a standard structured interview for depressive symptoms. Conversely, a questionnaire to assess health related quality of life should not correlate with measures that are unrelated to health, such as height or personality. In other words, the quality of life measure should have discriminant validity.

Accuracy of reporting

Quality of life assessments of mental functioning generally include questions on memory, job performance, sexual activity, and family role functioning. Self reported information in this area raises particular concern because neurological or psychological dysfunction can limit a patient's ability to report accurately.[13] In other words, we seek accurate information on cognitive



A classification scheme of quality of life measures (daily functioning and sense of wellbeing) related to health (mental and physical)

abilities when dysfunction in this area might make the patient's judgments unreliable. Alcoholism and other forms of psychopathology, for example, would present a problem in this regard.[14] Here, evaluating the convergent validity of a quality of life measure should be based upon agreement between the questionnaire scores and other measures of cognitive abilities, social behaviour, and job performance. However, this type of validation is virtually absent in published reports. Comparing how patients rate their driving abilities with performance during a driving test or in a driving simulator is an example of how self completed questionnaires could be validated (or found wanting).

Should perceived wellbeing and not functional assessment be used?

Much recent comment has maintained that quality of life is inherently subjective and that only perceived wellbeing, not functional assessment, should be used to determine quality of life.[7 15] This approach posits that the patient has privileged access to the quality of life outcomes of disease and treatment and that his or her assessment of wellbeing is of central importance. Subjective indices of quality of life correlate reliably with standard measures of psychiatric symptoms such as depression or anxiety, suggesting that in this sense they do measure subjective wellbeing (that is, have convergent validity).[9]

Effect on scores of extraneous factors

Ideally, subjective quality of life indices ideally should not be influenced by patient characteristics that are outside of the domain of disease and health care. These tests of discriminant validity are typically ignored or mischaracterised in quality of life validation. Patterns of response in questionnaires do vary with marital status, education, income, race, and geography, and, furthermore, are influenced by a variety of extraneous psychological factors.[16-19] For example, some people have response biases that lead them to give the answers they think are most socially acceptable or cast them in a favourable light.[20]

Influence of personality characteristics

Subjective quality of life scores can also be influenced by personality factors. Scores are therefore affected by enduring dispositional characteristics that predate the illness and treatment.[21 22] For example, a single item rating recommended as a suitable expression of quality of life—"Rate your overall quality of life as poor, fair, good or excellent"[7]—inadvertently measures personality characteristics such as the propensity to report negative affect, as well as hypochondriasis and somatisation.[23 24] The 36 item health survey of the medical outcomes study is a popular quality of life instrument that includes several subscales related to functioning as well as perceived wellbeing.[25] In a community sample of 348 generally healthy volunteers, we found that eight of the nine medical outcome study subscales correlated significantly with neuroticism, as measured by the NEO personality inventory (Muldoon MF et al, unpublished data). Other similar studies suggest that most subscales of the medical outcome study instrument vary with neuroticism and other dimensions of personality.[19 26] As the medical outcome study is a "mixed" instrument, this overlap suggests that self reported measures of functioning and perceived wellbeing lack optimal discriminant validity.

Confounding requires statistical adjustment

To protect against this confounding, investigators should report correlations between quality of life indices and characteristics that are unrelated to illness, and conduct statistical adjustments as indicated. For example, patients with mood or psychosomatic disorders in a primary care sample gave a lower rating for their general health than did patients with diabetes or pulmonary disorders.[27] On the surface, these findings indicate that mood or psychosomatic disorders reduce perceived health more than medical disorders do, but further analysis might suggest that personality factors lead to different response predispositions in various diagnostic groups.

Changes over time

How patients evaluate their quality of life may also change over time. For example, many cancer patients report benefits from their illness, ranging from an increased ability to appreciate each day to greater feelings of personal strength, self assurance, and compassion, such that they are sometimes more satisfied with their global quality of life than healthy comparison groups.[24 28-30] We might conclude that cancer improves quality of life. In fact, this paradox is now understood to reflect a psychological adaptation (a "response shift") that occurs in cancer patients as well as in patients with other chronic diseases such as diabetes, renal disease, and dermatological disorders.[31 32] The internal standard by which patients appraise their current state shifts and the same questionnaire items on wellbeing can elicit fundamentally different answers over time. To the extent that subjective wellbeing reflects psychological adaptation, the connection between subjective quality of life and disease course (or treatment response) weakens. Therefore, reported changes in quality of life over time[33] need not necessarily derive from actual changes in health or symptoms.

Conclusion

Assessment of the patient's experience of disease and treatment is now acknowledged as a central component of health care and healthcare research. Self reported information obtained from quality of life questionnaires is and will continue to be essential in this endeavour. However, conceptual and methodological issues that underlie this research—matters of definition, measurement objectives, and instrument validity—have received insufficient attention and thereby constrain permissible interpretation of the current medical literature.[33] In turn, implicit recognition of these deficiencies may partly account for the reluctance of many doctors to accept the legitimacy of quality of life research. Whether and how the quality of life "industry" responds to these obstacles and deficiencies will probably determine the future quality of research on quality of life.

Funding: This work was supported in part by National Institutes of Health grants HL46328, HL07560, and HL40962.

Conflict of interest: None.

1 Leplège A, Hunt S. The problem of quality of life in medicine. *JAMA* 1997;278:47-50.

2 Bergner M, Bobbitt RA, Carter WB, Gibson BS. The sickness impact profile: development and final revision of a health status measure. *Med Care* 1981;19:787-804.
3 Kaplan RM, Anderson JP, Patterson TL, McCutchan JA, Weinrich JD, Heaton RK, et al. Validity of the quality of well-being scale for persons with human immunodeficiency virus. *Psychosom Med* 1995;57:138-47.
4 Hays RD, Stewart AL. The structure of self-reported health in chronic disease patients. *Psychol Assess* 1990;2:22-30.
5 Katz S, Downs TD, Cash HR, Grotz RC. Progress in development of the index of ADL. *Gerontologist* 1970;10:20-30.
6 Guyatt GH, Feeny DH, Patrick DL. Measuring health-related quality of life. *Ann Intern Med* 1993;118:622-9.
7 Gill, TM, Feinstein AR. A critical appraisal of the quality of quality-of-life measurements. *JAMA* 1994;272:619-26.
8 Veit CT, Ware JE. The structure of psychological distress and well-being in general populations. *J Consult Clin Psychol* 1983;51:730-42.
9 McHorney CA, Ware JE, Raczek AE. The MOS 36-item short-form health survey (SF-36). II. Psychometric and clinical tests of validity in measuring physical and mental health constructs. *Med Care* 1993;31:247-63.
10 Hlatky HA, Boineau RE, Higginbotham MB, Lee KL, Mark DB, Califf RM, et al. A brief self-administered questionnaire to determine functional capacity (the Duke activity status index). *Am J Cardiol* 1989;64:651-4.
11 Kempen GI, Steverink N, Ormel J, Deeg DJ. The assessment of ADL among frail elderly in an interview survey: self-report versus performance-based tests and determinants of discrepancies. *J Geront Psychol Sci* 1996;51B:P254-60.
12 Campbell DT, Fiske DW. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol Bull* 1959;56:81-105.
13 Lopez OL, Becker JT, Somsak D, Dew MA, DeKosky ST. Awareness of cognitive deficits and anosognosia in probable Alzheimer's disease. *Eur Neurol* 1994;34:277-82.
14 Shedler J, Mayman M, Manis M. The illusion of mental health. *Am Psychol* 1993;48:1117-31.
15 Quality of life and clinical trials [editorial]. *Lancet* 1995;346:1-2.
16 Bucquet D, Condon S, Ritchie K. The French version of the Nottingham health profile: a comparison of item weights with those of the source version. *Soc Sci Med* 1990;30:829-35.
17 Hürny C, Bernhard J, Gelber RD, Coates A, Castiglione M, Isley M, et al for the International Breast Cancer Study Group. Quality of life measures for patients receiving adjuvant therapy for breast cancer: an international trial. *Eur J Cancer* 1992;28:118-24.
18 Centres for Disease Control. Quality of life as a new public health measure-behavioural risk factor surveillance system, 1993. *MMWR* 1994;43:375-80.
19 VanderZee KI, Sanderman R, Heyink JW, de Haes H. Psychometric qualities of the RAND 36-item health survey 1.0: a multidimensional measure of general health status. *Int J Behav Med* 1996;3:104-22.
20 Brooks WB, Jordan JS, Divine GW, Smith KS, Neelon FA. The impact of psychological factors on measurement of functional status. *Med Care* 1990;28:793-804.
21 Watson D, Pennebaker JW. Health complaints, stress, and distress: exploring the central role of negative affectivity. *Psychol Rev* 1989;96:234-54.
22 Duits AA, Boeke S, Taams MA, Passchier J, Erdman RA. Prediction of quality of life after coronary artery bypass graft surgery: a review and evaluation of multiple, recent studies. *Psychosom Med* 1997;59:257-68.
23 Barsky AJ, Cleary PD, Klerman GL. Determinants of perceived health status of medical outpatients. *Soc Sci Med* 1992;34:1147-54.
24 Tempelaar R, de Haes JCJM, de Ruiter JH, Bakker D, van den Heubel WJA, van Nieuwenhuijzen MG. The social experiences of cancer patients under treatment: a comparative study. *Soc Sci Med* 1989;29:635-42.
25 Stewart AL, Hays RD, Ware JE. The MOS short-form general health survey: reliability and validity in a patient population. *Med Care* 1988;26:724-35.
26 Kempen GI, Jelicic M, Ormel J. Personality, chronic medical morbidity, and health-related quality of life among older persons. *Health Psychol* 1997;16:539-46.
27 Spitzer RL, Kroenke K, Linzer M, Hahn SR, Williams JB, deGruy III FV, et al. Health-related quality of life in primary care patients with mental disorders. *JAMA* 1995;274:1511-7.
28 Fromm K, Andrykowski MA, Hunt J. Positive and negative psychosocial sequelae of bone marrow transplantation: implications for quality of life assessment. *J Behav Med* 1996;19:221-40.
29 Taylor SE, Lichtman RR, Wood JV. Attributions, beliefs about control, and adjustment to breast cancer. *J Personal Soc Psychol* 1984;46:489-502.
30 Danoff B, Kramer S, Irwin P, Gottlieb A. Assessment of the quality of life in long-term survivors after definitive radiotherapy. *Am J Clin Oncol* 1983;6:339-45.
31 Cassileth BR, Lusk EJ, Strouse TB, Miller DS, Brown LL, Cross PA, Tenaglia AN. Psychosocial status in chronic illness. *N Engl J Med* 1984;311:506-11.
32 Breetvelt IS, van Dam FS. Underreporting by cancer patients: the case of response-shift. *Soc Sci Med* 1991;32:981-7.
33 Hemingway H, Stafford M, Stansfeld S, Shipley M, Marmot M. Is the SF-36 a valid measure of change in population health? Results from the Whitehall II study. *BMJ* 1997;315:1273-9.

(Accepted 5 August 1997)

Continuing medical education

Recertification and the maintenance of competence

Philip G Bashook, John Parboosingh

Completion of postgraduate specialist training is a landmark event for most doctors. The award of a certificate is acknowledgment that a doctor has undergone a recognised training programme and been assessed as competent to practise as a specialist in his or her field. Specialists begin practice with a common knowledge base and similar clinical skills but go on to develop different areas of expertise in response to patients' needs. In time, the knowledge and skills of doctors within a specialty will vary appreciably.

Recertification in the United States

Recognition of the disparity in doctors' skills and the need to maintain common core standards have been a key factor behind the "recertification" movement in the United States.[1] The movement became established in 1969 when the American Board of Family Practice began issuing time limited certificates. Although recertification is nominally a voluntary process, doctors must get recertified every seven years if they want to retain the status of being "board certified."[2] The United States is currently the only country in which most trained specialists are expected to obtain recertification certificates at set intervals throughout

Summary points

Recertification should assess real performance in practice and competence to continue to learn

Recertification programmes in the United States use examinations and performance assessments as "snapshots" of competence taken every 7-10 years

In other countries most programmes evaluate documented participation in continuing education as evidence of continuing competence as a specialist

The proposed continuous recertification programme uses computer technology to document self directed learning from practice and to monitor performance

Poor performers could be recognised early, given focused assistance and additional periodic examinations at testing centres, and if necessary their certificates could be rescinded

This is the fourth in a series of seven articles looking at international trends and forces in doctors' continuing professional development

American Board of Medical Specialties, Evanston, IL 60201-5913, USA
Philip G Bashook, *director of evaluation and education*

continued over

BMJ 1998;316:545–8

Royal College of Physicians and Surgeons of Canada, Ottawa, Ontario, Canada K1S 5N8

John Parboosingh, *associate director, office of fellowship affairs*

Correspondence to: Dr Bashook pgb@abms.org

Series editors: Hans Asbjørn Holm and Tessa Richards

their professional lives. Twenty two of the 24 member boards of the American Board of Medical Specialties issue certificates, with expiry dates varying between seven and ten years. The incentive to get recertified is strong, for a valid board certificate has become essential for doctors in many communitites in order to admit patients to hospital and claim the top reimbursement fees and salaries of a specialist. Reports that doctors who do not have specialty certification are falsifying certificates or claiming specialty certification on their curriculum vitae are increasing.

What does recertification entail?

The recertification procedures set up by the member boards of the American Board of Medical Specialties aim to encourage doctors to continue learning and keep up to date; give recognition to doctors who continue to meet the specialty board's standards; and remove certification status from doctors holding time limited certificates who fail to apply for recertification.

Most of the boards use a snapshot assessment of knowledge, skills, and performance. Written examinations, usually in the form of multiple choice questions, are used by all boards, and 11 require set credit hours of continuing medical education (CME), typically 50 hours a year in the three years before recertification. Performance is measured indirectly by report of licensure status, letters of recommendation from chiefs of healthcare organisations and hospitals, attendance at CME programmes, and independent assessment by peers and other health professionals. Some boards allow specialists to select their own form of assessment.

Recertification is not cheap. The member boards of the American Board of Medical Specialties charge doctors between $533 and $1255 to sit the written examinations and up to $10 500 for a two day on-site visit. On-site review of practice has recently been discontinued,[2] and it is difficult and expensive to introduce more rigorous forms of assessment of clinical skills. Site visits, examinations using standardised patients, and case recall interviews[17] have been found to be too expensive or impossible to implement for large numbers of board certified doctors. Furthermore, obtaining hard evidence of the validity and reliability of such methods of assessment would entail extensive and expensive research—hence the reliance on written examinations.[12]

DAVID HITCH

Driven to extremes by competition

Medical care in the United States is a competitive marketplace. Doctors in fee for service practice have to compete with large health corporations that own hospitals and doctors' practices. These corporations use the number of affiliated "board certified specialists" as an indicator of the "quality" of service they provide. Certification has also been used by *Consumer Reports*, a respected consumer organisation engaged in quality assessment, as a criterion in ranking "best hospitals" and "best healthcare plans." Patients have routinely taken to consulting directories such as the *Official ABMS Directory of Board Certified Specialists* or calling a freephone number to determine a doctor's certification credentials.

This pressure on doctors to produce documented if purely nominal evidence that they are competent and up to date has had undesirable side effects. One has been the growth of self designated "certifying boards" set up by specialty societies and by entrepreneurs. Most of these, of which there are around 150, have adopted names which mimic the names of the member boards of the American Board of Medical Specialties. Doctors who obtain certification certificates from these organisations are required to pass an examination, take out membership, and pay annual fees to retain their "board" status. The standards for these qualifications vary widely, and the many different forms of "board" certification cause concern and confusion for both the profession and the public. A second development has been the launch of a new certification programme, the American Medical Accreditation Program, by the American Medical Association.[5] This programme allows non-certified doctors to obtain what the association terms "accreditation" as a specialist even if they have not completed recognised training programmes and obtained a certificate of satisfactory completion of specialist training. This move, which is likely to cause further confusion among the public, will need to be followed closely.

A third development has been the proliferation of CME programmes aimed at (and advertised as) teaching doctors how to pass board recertification exams. The essential question of whether these programmes provide education useful for practice is deemed to be of secondary importance.

Recertification and CME in other English speaking countries

Outside the United States, most postgraduate colleges have elected not to incorporate formal examinations into their recertification procedures. In many, the initial certification process amounts to more than a single exit examination, doctors being required to undergo frequent in-training evaluations over many years. The colleges then offer programmes for maintenance of competence, based largely on participation in formal

educational activities. Most postgraduate recertification or CME programmes simply require a set number of hours of attendance, usually 50 a year, at recognised CME courses. More recently, weighted credit systems have been introduced in Canada; these recognise that some forms of CME are more effective than others at changing practice. Thus the MOCOMP system (see box) awards credits on the basis of the educational quality of the programme: traditional didactic sessions are rated at 1 credit per hour while interactive workshops based on audits of practice with opportunities to interact with faculty members receive 2 credits per hour.

In Australia, the Royal Australasian College of Physicians has led the way in incorporating recertification criteria that relate more closely to doctors' performance than attendance at traditional CME courses. Participation in quality improvement initiatives such as audits of practice, as well as attendance of traditional CME courses, is required. The college also has a unique physician assessment programme in which peers, coworkers, and patients rate doctors on their clinical management and their "holistic" and personal skills with patients.[13] A recent pilot study in Canada showed that this method can provide reliable and meaningful assessments of doctors, and peer assessment may become a mandatory requirement for licensure in the province of Alberta.

Time limited certification is legally required of specialists in Australia and New Zealand, and in Canada it is required for membership of the College of Family Physicians of Canada.[14] In the United Kingdom the royal colleges and specialist associations are piloting credit systems that are similar to the Australian model except that participation is voluntary, not mandatory. In Canada, certification as a specialist by the Royal College of Physicians and Surgeons is life long. Although there are no plans for introducing recertification procedures, the college is experimenting with self directed learning programmes.

Continuous recertification: the way forward?

Snapshot assessments every 7-10 years are a crude form of assessment of competence. A more effective way to maintain professional knowledge and performance is to introduce a programme of continuous recertification. We propose a programme based on a combination of audit of practice data and documented evidence of continuous learning in practice.

Practice performance data

Medical records provide data on patient encounters, prescriptions, other treatment modalities and follow up visits. Four member boards of the American Board of Medical Specialties already use such data, requiring doctors to submit computerised summary reports on patients: family practice,[15 16] plastic surgery,[17] obstetrics and gynaecology,[18] and orthopaedic surgery.[19] Managed care corporations routinely use computer technology to monitor doctors' performance, patient outcomes, and patients' views of doctors' attitudes. A continuous recertification programme could build on this technology.

In addition to assessments of their knowledge, decision making skills, and technical expertise, doctors should be assessed on their abilities to communicate with both their patients and their peers, to share the process of decision making, to work as members of a team, and to break bad news with empathy. Modern information systems will facilitate this form of multiple assessment, which could be made annually or even more frequently as part of a cycle of continuous recertification.

The maintainence of competence program (MOCOMP)

- MOCOMP is a voluntary continuing education program by the Royal College of Physicians and Surgeons of Canada to help specialists manage their continuing education themselves
- The PCDiary software in MOCOMP is used by physician subscribers to define their learning needs and keep a portfolio of learning (pearls of wisdom) generated from practice, reflection on clinical experiences, CME meetings, journal reading, and "hallway consultations"
- PCDiary software contains powerful searching, sorting, and report generating capabilities to encourage reflection and appraisal of learning entries
- A searchable database is generated from entries into PCDiary to produce a "question library" available on the internet that allows physicians to compare with peers their leaning needs and practices. The question library also serves as a repository of identified medical education needs that is helpful for planners of CME
- MOCOMP has 10 000 specialists voluntarily registered out of the 30 000 specialists in Canada. Approximately 400 use PCDiary and 3000 use a paper version. Experienced users average 4-8 entries each month. All reports are easy to use and not time consuming. Some users report that MOCOMP motivates them to "take professional development seriously" and "to organise their learning"; others perceive they are more selective about attending educational conferences and meetings
- PCDiary provides summary reports that add a "living component" to the traditional curriculum vitae and have the potential to be used for renewing credentials of doctors
- MOCOMP contains the tools to enable doctors—including researchers, educators, and administrators—to move from the traditional medical school model of learning to self managed learning with reflection about practice experiences

Continuous learning in practice

The foundations of quality patient care begin during training, but with rapid developments in medical knowledge doctors have to learn continuously in practice if they are to maintain high quality care.[1] More than ever, doctors need support systems to help them use feedback on their performance to plan and implement effective and individual continuing education programmes. Systems, such as the computerised maintenance of competence program (MOCOMP) in Canada are being set up to help them meet these needs.[25]

The Royal Australasian College of Pathologists has piloted a similar software program to help its members to use learning portfolios as part of their maintenance of certification. Similar systems are also being explored in the United Kingdom.

Computer technology

Periodic examinations may be secondary to continuous evaluation of practice performance and keeping learning portfolios, but they undeniably have a place in continued medical education. Computer based examinations in particular are available at testing centres

worldwide through a combination of entrepreneurial companies and not for profit testing organisations.[23] The shift from paper tests to computer based tests has accelerated in recent years. For example, the recertification examinations of the American Board of Pediatrics and the American Board of Pathology are distributed on computer diskettes for use at home, and in 1997 the American Board of Orthopedic Surgery and the American Board of Anesthesiology ran their recertification programme only in computer testing centres. The American Board of Pathology has been operating such a centre for over two years and will double the centre's capacity by June.[24]

Advanced multimedia computer technology, such as virtual reality environments, is being developed to help train doctors to perform invasive surgical and endoscopic procedures. This technology may also be used to evaluate how well the doctors carry out these procedures and other patient-doctor interactions. These medical "flight simulators" are already available commercially.[26] Within a decade they are likely to be used widely by medical schools and hospitals, both as learning tools and to evaluate doctors' performance, and also to provide remedial training where there is evidence of deficiencies in practice. Certifying boards and colleges could use these centres as a second step for more in depth assessments of clinical skill.

Conclusion

In the future, recertification programmes could require specialists to provide certifying boards with computerised summary reports of their practice experience and learning portfolios every 3-5 years. Clicking a mouse button or touching the keyboard would generate the recertification report. Much of the scheduling could be automated, and specialists could have automatic reminders about what information is needed; where in their computer reports it is located; and how, when, and where to send it. Doctors who fail to meet set standards, or those who have not practised for some time, would have to undergo more in-depth educational assessment so that an educational "prescription" of continuing education could be drawn up to help improve their performance.

Continuing learning must be seen as a routine part of daily practice. Objective evidence of the quality of care can be obtained by integrating audit and self assessment programmes into routine clinical practice. Feedback on the results should be given on a regular basis and regarded not as a threat but as an opportunity to learn. Regular appraisal of practice, using multiple assessments, will also allow early recognition of doctors who are performing badly and need focused help or remedial education, or their licence removed.

The biggest obstacle to implementing continuous recertification is professional conservativism about learning methods and computer technology. These attitudes must change, for computer literacy will soon be essential for medical practice. At the same time it is increasingly being accepted that all medical students need to be taught about the concepts of adult learning so that as doctors they go on to become lifelong learners.[28]

It may take time to debate the merits of continuous recertification, but in our view this strategy is consistent with the evidence on how adults learn and keep up to date,[1] feasible and affordable with current technology, and crucial to the provision of high quality medical care.

The views expressed here are the authors' and do not represent either the American Board of Medical Specialties and its member boards or the Royal College of Physicians and Surgeons of Canada.

1 Fox RD, Mazmanian PE, Putman RW, eds. *Changing and learning in the lives of physicians.* New York: Praeger, 1989.
2 American Board of Medical Specialties. *1996 annual report and reference handbook.* Evanston, IL: ABMS, 1996.
3 American Board of Medical Specialties. *The official ABMS directory of board certified medical specialists.* New Providence, NJ: Marquis Who's Who, 1997.
4 Seifer SD, Troupin B, Rubenfeld GD. Changes in the marketplace demand for physicians: a study of medical journal recruitment advertisements. *JAMA* 1996;276:695-9.
5 American Medical Association. *American Medical Accreditation Program (AMAP), standards and criteria.* Chicago: AMA, 1997.
6 Kassirer JP. The new surrogate for board certification. *N Engl J Med* 1997;337:43-4.
7 Ramsey PG, Carline JD, Blank LL, Wenrich MD. Feasibility of hospital-based use of peer ratings to evaluate the performance of practising physicians. *Acad Med* 1996; 71:364-70.
8 Haladyna TM. Measuring higher order thinking. In: *Developing and validating multiple-choice test items.* Hillsdale, NJ: L Erlbaum Associates, 1994:87-107.
9 Page GG. Assessing reasoning and judgment. In: Mancall EL, Bashook PG, eds. *Assessing clinical reasoning: the oral examination and alternative methods.* Evanston, IL: American Board of Medical Specialties. 1995:9-27.
10 McGaghie WC. Evaluating competence for professional practice. In: Curry L, Wergin JF, eds. *Educating professionals.* San Francisco: Jossey-Bass, 1993:229-61.
11 Bashook PG. Planning a recertification program. In: Mancall EL, Bashook PG, eds. *Recertification: new evaluation methods and strategies.* Evanston, IL: American Board of Medical Specialties. 1994:15-27.
12 Smith FG. Computer-based examinations. In: Mancall EL, Bashook PG, eds. *Recertification: new evaluation methods and strategies.* Evanston, IL: American Board of Medical Specialties. 1994:33-42.
13 Newble D, Wakeford R. Primary certification in the UK and Australasia. In: Newble D, Jolly B, Wakeford R, eds. *The certification and recertification of doctors: issues in the assessment of clinical competence.* Cambridge: Cambridge University Press, 1991:13-8.
14 Violato C, Marini A, Toews J, Lockyer J, Fidler H. Feasibility and psychometric properties of using peers, consulting physicians, co-workers, and patients to assess physicians. *Acad Med* 1997;72:S82-4.
15 College of Family Physicians of Canada. MAINPRO™, Maintenance of Proficiency/Maintain de la Compétence Professionnelle. Toronto: CFPC, 1995.
16 Leigh TM. Computerized office record review—update. In: Lloyd JS, Langsley DG, eds. *Recertification for medical specialists.* Evanston, IL: American Board of Medical Specialties. 1987:103-20.
17 Young PR. Office record review. In: Mancall EL, Bashook PG, eds. *Recertification: new evaluation methods and strategies.* Evanston, IL: American Board of Medical Specialties. 1994:63-72.
18 Neale HW, Puckett CL. Clinical Case Review. In: Mancall EL, Bashook PG, eds. *Recertification: new evaluation methods and strategies.* Evanston, IL: American Board of Medical Specialties. 1994:57-62.
19 Merrill JA. Personal practice audit. In: Lloyd JS, Langsley DG, eds. *Recertification: for medical specialists.* Evanston, IL: American Board of Medical Specialties. 1987:97-102.
20 Kettelkamp DB. Practice review. In: Mancall EL, Bashook PG, eds. *Recertification: new evaluation methods and strategies.* Evanston, IL: American Board of Medical Specialties. 1994:53-5.
21 Davis DA, Fox RD. *The physician as learner: from research to practice.* Chicago: American Medical Association, 1995.
22 Campbell, CM, Parboosingh JT,Gondocz ST, Babitskaya G, Lindsey E, de Guzman RC, et al. Study of physicians' use of a software program to create a portfolio of their self-directed learning. *Acad Med* 1996;71(suppl):S49-51.
23 Parboosingh J. Learning portfolios: potential to assist health professionals with self-directed learning. *J Cont Educ Health Professions* 1996;16:75-81.
24 Rosen GA. Single sites, multi-sites and multi-day availability. In: Mancall EL, Bashook PG, Dockery JL, eds. *Computer-based examinations for board certification.* Evanston, IL: American Board of Medical Specialties. 1996:41-5.
25 Hartmann WH. Certifying boards and computer-based examinations. In: Mancall EL, Bashook PG, Dockery JL, eds. *Computer-based examinations for board certification.* Evanston, IL: American Board of Medical Specialties. 1996:91-7.
26 Satava RM, Jones SB. Virtual reality environments in medicine. In: Mancall EL, Bashook PG, Dockery JL, eds. *Computer-based examinations for board certification.* Evanston, IL: American Board of Medical Specialties. 1996:121-31.
27 Taggart WR. Certifying pilots: implications for medicine and for the future. In: Mancall EL, Bashook PG, eds. *Assessing clinical reasoning: the oral examination and alternative methods.* Evanston, IL: American Board of Medical Specialties. 1995:175-80.
28 Karle H, Nystrup J, Walton HJ. Medical specialization in Europe: the way forward. Promotion of educational quality. *Med Educ* 1993;27:299-303.
29 Manning PR, DeBakey L. *Preserving the passion.* New York: Springer-Verlag, 1987.

Statistics notes

Sample size in cluster randomisation

Sally M Kerry, J Martin Bland

Techniques for estimating sample size for randomised trials are well established,[1 2] but most texts do not discuss sample size for trials which randomise groups (clusters) of people rather than individuals. For example, in a study of different preparations to control head lice all children in the same class were allocated to receive the same preparation. This was done to avoid contaminating the treatment groups through contact with control children in the same class.[3] The children in the class cannot be considered independent of one another and the analysis should take this into account.[4 5] There will be some loss of power due to randomising by cluster rather than individual and this should be reflected in the sample size calculations. Here we describe sample size calculations for a cluster randomised trial.

For a conventional randomised trial assessing the difference between two sample means the number of subjects required in each group, n, to detect a difference of d using a significance level of 5% and a power of 90% is given by $n=21s^2/d^2$ where s is the standard deviation of the outcome measure. Other values of power and significance can be used.[1]

For a trial using cluster randomisation we need to take the design into account. For a continuous outcome measurement such as serum cholesterol values, a simple method of analysis is based on the mean of the observations for all subjects in the cluster and compares these means between the treatment groups. We will denote the variance of observations within one cluster by s_w^2 and assume that this variance is the same for all clusters. If there are m subjects in each cluster then the variance of a single sample mean is s_w^2/m. The true cluster mean (unknown) will vary from cluster to cluster, with variance s_c^2. The observed variance of the cluster means will be the sum of the variance between clusters and the variance within clusters—that is, variance of outcome $=s_c^2+s_w^2/m$. Hence we can replace s^2 by $s_c^2+s_w^2/\mathrm{m}$ in the formula for sample size above to obtain the number of clusters required in each intervention group. To do this we need estimates of s_c^2 and s_w^2.

For example, in a proposed study of a behavioural intervention in general practice to lower cholesterol concentrations practices were to be randomised into two groups, one to offer intensive dietary intervention by practice nurses using a behavioural approach and the other to offer usual general practice care. The outcome measure would be mean cholesterol values in patients attending each practice one year later. Estimates of between practice variance and within practice variance were obtained from the Medical Research Council thrombosis prevention trial[6] and were $s_c^2=0.0046$ and $s_w^2=1.28$ respectively. The minimum difference considered to be clinically relevant was 0.1 mmol/l. If we recruit 50 patients per practice, we would have $s^2=s_c^2+s_w^2/m=0.0046+1.28/50=0.0302$. The number of practices is given by $n=21\times0.0302/0.1^2=63$ in each group. We would require 63 practices in each group to detect a difference of 0.1 mmol/l with a power of 90% using a 5% significance level—a total of 3150 patients in each group.

Total number of practices required to detect a difference of 0.1 mmol/l cholesterol with 90% power at 5% significance level

No of patients per practice (*m*)	Standard deviation	No of practices	No of patients	Design effect
10	0.364	558	5 580	1.04
25	0.236	234	5 850	1.09
50	0.173	126	6 300	1.17
100	0.132	74	7 400	1.38
500	0.085	32	16 000	2.98
No needed with individual randomisation			5 364	1.00

It can be seen from the formula for the variance of the outcome that when the number of patients within a practice, m, is very large, s_w^2/m will be very small and so the overall variance is roughly the same as the variance between practices. In this situation, increasing the number of patients per practice will not increase the power of the study. The table shows the number of practices required for different values of m, the number of subjects per practice. In all situations the total number of subjects required is greater than if simple random allocation had been used.

The ratio of the total number of subjects required using cluster randomisation to the number required using simple randomisation is called the design effect. Thus a cluster randomised trial which has a large design effect will require many more subjects than a trial of the same intervention which randomises individuals. As the number of patients per practice increases so does the design effect. In the table, the design effect is very small when m is less than 10. This would involve recruiting a total of 558 practices, and the nature of the intervention and difficulties in recruiting practices made this impractical. Thus it was decided to recruit fewer practices. The design effect of using 126 practices with 50 patients from each practice was 1.17. This design requires the total sample size to be inflated by 17%. If the study involves training practice based staff it may be cost effective to reduce the number of practices even further. If we chose to use 32 practices then we would need 500 patients from each practice and the design effect would be 2.98. Thus the cluster design with 32 practices would require the total sample size to be trebled to maintain the same level of power.

We shall discuss the use of the intracluster correlation coefficient in these calculations in a future statistics note.

1 Florey C du V. Sample size for beginners. *BMJ* 1993;306:1181-4.
2 Machin D, Campbell MJ. *Statistical tables for the design of clinical trials.* Oxford: Blackwell, 1987.
3 Chosidow O, Chastang C, Brue C, Bouvet E, Izri M, Monteny N, et al. Controlled study of malathion and d-phenothrin lotions for *Pediculus humanus* var *capitis*-infested schoolchildren. *Lancet* 1994;344:1724-7.
4 Bland JM, Kerry SM. Trials randomised in clusters. *BMJ* 1997;315:600.
5 Kerry SM, Bland JM. Analysis of a trial randomised in clusters. *BMJ* 1998;316:54.
6 Meade TW, Roderick PJ, Brennan PJ, Wilkes HC, Kelleher CC. Extracranial bleeding and other symptoms due to low dose aspirin and low intensity oral anticoagulation. *Thromb Haemostasis* 1992;68:1-6.

Division of General Practice and Primary Care, St George's Hospital Medical School, London SW17 0RE

Sally M Kerry, *statistician*
Department of Public Health Sciences

J Martin Bland, *professor of medical statistics*

Correspondence to: Mrs Kerry.

BMJ 1998;316:549