Interpreting treatment effects in randomised trials

Gordon H Guyatt, Elizabeth F Juniper, Stephen D Walter, Lauren E Griffith, Roger S Goldstein

Department of Clinical Epidemiology and Biostatistics, McMaster University, Hamilton, Ontario, Canada L8N 3Z5
Gordon H Guyatt, *professor*
Elizabeth F Juniper, *professor*
Stephen D Walter, *professor*
Lauren E Griffith, *research biostatician*

Division of Respiratory Medicine, University of Toronto, Toronto, Ontario, Canada
Roger S Goldstein, *professor*

Correspondence to:
Dr Guyatt
guyatt@fhs.csu.mcmaster.ca

BMJ 1998;316:690–3

The need to measure the impact of treatments on health related quality of life has led to a rapid increase in the variety of instruments available and in their use as measures of outcome in clinical trials. One limitation of instruments that purport to measure health related quality of life is difficulty interpreting their results. In the past decade, investigators have progressed in making these questionnaire results interpretable. For example, we have shown that when questionnaires present response options in the form of seven point scales with verbal descriptions for each option (see box), the smallest difference that patients consider important is often approximately 0.5 per question. A moderate difference corresponds to a change of approximately 1.0 per question, and changes of greater than 1.5 can be considered large. Thus, for example, in a domain with four items, patients will consider a 1 point change in two or more items as important. This finding applies across different areas of function, including dyspnoea, fatigue, and emotional function in patients with chronic airflow limitation[1]; and symptoms, emotional function, and activity limitations in adults[2] and children[3] with asthma, parents of children with asthma,[4] and adults with rhinoconjunctivitis.[5] Initially, we used comparisons in the same patient to establish this difference, but more recently we have replicated this finding using differences between patients.[6]

Summary points

- Several questionnaires on quality of life related to health are available, but interpreting their results may be difficult
- For some questionnaires, we now know that the smallest change in score that patients consider important is 0.5
- Even if the mean difference between a treatment and a control is appreciably less than the smallest change that is important, treatment may have an important impact on many patients
- A method for estimating the proportion of patients who benefit from a treatment when the outcome is a continuous variable has been developed
- The method is outlined using two examples, one a crossover trial and the other a parallel group design
- This approach emphasises the need to establish ranges of health related changes that represent trivial, small but important, moderate, and large changes in addition to mean differences

Assumptions

Clinicians and investigators tend to assume that if the mean difference between a treatment and a control is appreciably less than the smallest change that is important, then the treatment has a trivial effect. This may not be so. Let us assume that a randomised clinical trial shows a mean difference of 0.25 in a questionnaire in which the minimal important difference is 0.5. It might be concluded that the difference is unimportant and that the result does not support giving the treatment. This interpretation assumes that every patient treated scored 0.25 better than they would have done had they received the control and ignores the possibility that treatment might have a heterogeneous effect. Depending on the true distribution of results, the appropriate interpretation might be different.

Consider a situation in which 25% of the treated patients improved by a magnitude of 1.0, while the other 75% did not improve at all (mean change of 0). This would mean that the 25% of those treated obtained a moderate benefit from the intervention. Using the method that has recently been developed for interpreting the size of treatment effects—the number needed to treat—investigators have found that doctors often treat 25 to 50 patients, even as many as 100, in order to prevent a single adverse event.[7 8] Thus, the hypothetical treatment with a mean difference of 0.25 and a number needed to treat value of 4 proves to have a powerful effect.

We have developed a method for estimating the proportion of patients who benefit from a treatment when the outcome is a continuous variable. We outline this method using two examples, one a crossover trial and the other a parallel group design.

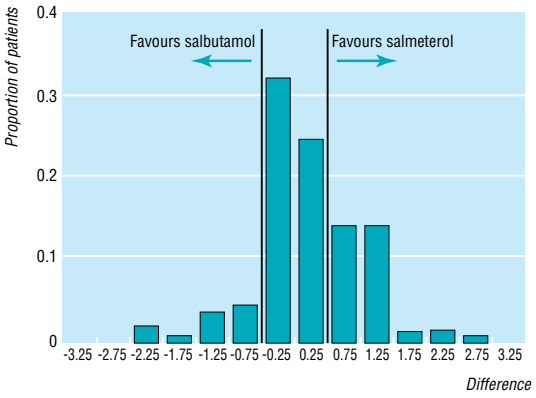
Crossover trial

To complete the asthma quality of life questionnaire, patients rate the impairments they have experienced during the previous 14 days and respond to 32 questions on seven point scales similar to that in the

Seven point scale with verbal descriptors

The following options were given for response to the question "How short of breath have you felt during the last two weeks while climbing stairs?"
1—extremely short of breath
2—very short of breath
3—quite a bit short of breath
4—moderate shortness of breath
5—some shortness of breath
6—a little shortness of breath
7—not at all short of breath

In the seven point scales used in this study, 7 represents the best possible function, and 1 the worst possible function.



Difference in the activity domain of the asthma quality of life questionnaire between periods of treatment with salmeterol and salbutamol

box.[9] In a multicentre double blind crossover randomised trial lasting 12 weeks, 140 patients received salmeterol (50 µg, twice daily), salbutamol (200 µg, four times daily) or placebo plus salbutamol (to be opened as needed). Each patient received all three regimens and used the questionnaire to rate their quality of life in relation to their asthma at the end of each study period.[10]

The mean differences between salmeterol and salbutamol, and between salmeterol and placebo, met conventional criteria for significance. In the current analysis, we examined and compared the distribution of different scores in the salmeterol, salbutamol, and placebo periods. We reasoned that the number of patients who had obtained important benefit from treatment would be the number with a difference of 0.5 or more favouring the treatment period, minus the number with a difference of 0.5 or more favouring the control period. This measure is analogous to the conventional risk difference, with 1 divided by the difference in risk being the number needed to treat.

The figure shows the distribution of differences between the salmeterol and salbutamol treatment periods in the activity domain of the asthma quality of life questionnaire and the difference in the proportion of the distribution in the important benefit compared with the important deterioration ranges. The distribution is approximately normal.

Table 1 shows that for both comparisons, differences between treatments failed to reach the threshold of the minimal important difference for the activity limitation section of the asthma quality of life questionnaire. In the symptom section of the questionnaire, the difference between salmeterol and salbutamol bordered on the minimal important difference. The only comparison in which the minimal important difference was clearly exceeded was that between salmeterol and placebo in the symptom section of the questionnaire.

In contrast to these mean differences, many patients had scores that were more than 0.5 better for salmeterol compared with salbutamol treatment for both symptoms and activity limitations. Fewer had scores that were 0.5 or more better for salbutamol compared with salmeterol. The difference in the proportions is even greater for the comparison between salmeterol and placebo (table 1).

Comparing salmeterol and salbutamol, clinicians would need to treat 4.5 patients for one patient to gain important benefit in the activity domain (or 45 for 10 to benefit). However, the number needed to treat for salmeterol compared with placebo in the activity domain is 2.9.

Parallel group trial

The chronic respiratory questionnaire, which includes 20 items measuring dyspnoea, fatigue, emotional function, and mastery (the extent to which patients feel in control), was developed for use in patients with moderate or severe chronic airflow limitation, and uses seven point scale response options.[11] Seventy eight patients with chronic airflow limitation were randomly allocated to a six month programme of respiratory rehabilitation or to conventional community care. We used differences between the patients' chronic respiratory questionnaire scores at baseline and after 24 weeks reported in the primary analysis of the trial results in the current analysis.[12] Mean differences between treatment and control for three domains reached significance.

The analysis of the parallel group trial provides additional challenges beyond those of the crossover trial. In theory, to calculate the proportion who improved on treatment we would have needed to know how rehabilitation patients would have fared had they received standard care, and how the standard care patients would have fared had they received rehabilitation. However, we could not observe these data directly because patients received only one treatment or the other. We do, however, know the proportion who improved, remained the same, and deteriorated relative to their baseline status in both treatment and control groups (table 2).

Table 1 Differences between groups given different treatments for asthma

Asthma quality of life questionnaire domains	Difference between treatments		Proportion better on salmeterol	Proportion better on salbutamol or placebo	Proportion who benefited	No needed to treat for a single patient to benefit
	Mean	P value				
Salmeterol *v* salbutamol						
Symptoms	0.50	<0.0001	0.42	0.12	0.30	3.3
Limitations on activity	0.30	<0.0001	0.32	0.10	0.22	4.5
Salmeterol *v* placebo						
Symptoms	0.68	<0.0001	0.50	0.09	0.41	2.4
Limitations on activity	0.43	<0.0001	0.42	0.08	0.34	2.9

Table 2 Calculating the proportion of patients who benefited from receiving rehabilitation in a parallel group trial*

	Treatment		
Control	Improved (0.48) (x)	Unchanged (0.42) (y)	Deteriorated (0.10) (z)
Improved (0.28) (a)	0.13 (ax)	0.12 (ay)	0.03 (az)
Unchanged (0.49) (b)	0.24 (bx)	0.21 (by)	0.05 (bz)
Deteriorated (0.23) (c)	0.11 (cx)	0.10 (cy)	0.02 (cz)

*Data from the dyspnoea domain of the chronic respiratory questionnaire. The number needed to treat for one patient to benefit from rehabilitation is calculated by adding up cells of those who improved (bx+cx+cy), subtracting the cells of those who deteriorated (ay+az+bz), and dividing 1 by the result.

Table 3 shows the proportion of patients in the rehabilitation and control groups whose dyspnoea scores increased by more than 0.5 (improved), changed between −0.5 and 0.5 (unchanged), and fell by more than 0.5 (deteriorated). We can refer to the proportions improved, unchanged, and deteriorated in the two groups as the "marginals." Given these marginals, there is, in general, no single way of filling in the individual cells in table 2—indeed, there are many possibilities. We have assumed that treatment and control responses are independent. Making this assumption, we obtain estimates of the individual cell values by multiplying the corresponding marginals (for instance, in table 2 we obtain the value for cell ax by multiplying the proportion improved in the rehabilitation group by the proportion improved in the standard care group). In table 2, cells ax, by, and cz represent patients whose outcome is the same irrespective of treatment. Patients in cells ay, az, and bz fared better receiving standard care than rehabilitation, and patients in cells bx, cx and cy fared better receiving rehabilitation than standard care. Thus, the proportion who received benefit from treatment is $(bx + cx + cy) - (ay + az + bz)$, which in this case is $(0.24 + 0.11 + 0.10) - (0.12 + 0.03 + 0.05) = 0.25$ (0.24 without rounding error). The number needed to treat value is therefore 1/0.24, or 4.2.

Table 3 gives the full results and shows that the mean difference between treatment and control groups exceeded the minimal important difference in two of the four domains. However, for all four domains, the difference in the proportion improved compared with deteriorated in the two treatment groups was similar, leading to consistent number needed to treat values of between 2.5 and 4.4.

Interpretation of treatment effects

The notion of taking a continuous variable, specifying a threshold that defines an important difference, and examining the proportions of patients who reach that threshold is not new. In considering the treatment of hypertension, Rose emphasised the difference between mean differences in populations and the impact these differences might have on individuals. In one specific example, Duffy argues persuasively that knowledge of mean changes in alcohol consumption in a population does not allow one to estimate change in the proportion of heavy drinkers. Rather, ascertaining the proportion of heavy drinkers requires direct measurement.[13] Another good example of this approach comes from a recent controlled trial of tissue plasminogen activator treatment in patients with acute stroke.[14] In reporting the results of this study, the authors presented both mean values of functional measures and differences in the proportions of patients who reached a threshold level of function.

What we have done that is new is to anchor the threshold difference using the smallest difference that patients consider important—the minimal important difference. We have shown how the method can be applied in both crossover and parallel group trials, how to generate the number needed to treat for one patient to benefit from therapy, and how superficial examination of mean differences can produce very misleading conclusions.

When mean differences fall below the minimal important difference, clinicians may intuitively conclude that the treatment has a small, and possibly unimportant, effect. Similarly, doctors who observe a mean difference that is appreciably greater than the minimal important difference may be ready to assume that each patient benefits. This is not necessarily the case. For example, we found a mean difference of 0.7 in the mastery domain of the chronic respiratory questionnaire between those who received and did not receive rehabilitation. Despite this substantial difference, the number needed to treat was 2.5. This means that for every five patients who complete a rehabilitation programme, only two will be better off—a result that may have major implications for the cost effectiveness of the intervention.

Our approach is not restricted to health related quality of life or functional status measures, but applies to any clinical variable. For instance, the interpretation of changes in pulmonary function, exercise capacity, or renal or cardiac function could all be analysed in this way. For these variables, however, the concept of the minimal important difference may be questioned. If renal failure requires dialysis or if cardiac function deteriorates to the point that a heart transplant is nec-

Table 3 Differences between patients with chronic airflow limitation who were receiving rehabilitation and patients given conventional care

Chronic respiratory questionnaire domain	Difference between groups		Estimated proportion better on rehabilitation	Estimated proportion better on conventional care	Proportion benefiting from rehabilitation	No needed to treat for a single patient to benefit
	Mean	P value				
Dyspnoea	0.60	0.0003	0.44	0.20	0.24	4.1
Fatigue	0.45	0.06	0.45	0.23	0.23	4.4
Emotional function	0.40	0.001	0.47	0.17	0.30	3.3
Mastery	0.71	0.0001	0.54	0.15	0.39	2.5

SIMON SHEPHEARD/IMPACT

Ascertaining the proportion of heavy drinkers requires direct measurement

essary, the importance for the patient is clear. Smaller changes in physiological function are important not in themselves, but rather through their effects on patient function and her or his health related quality of life. When considering differences that are important to patients, it may be more appropriate to measure function and health related quality of life directly rather than physiological variables.

Our approach is a way of making data more interpretable—we do not advocate its use as the only analysis. Power may be lost when converting continuous variables to dichotomous or categorical variables. We believe the initial analysis should examine whether differences in continuous variables meet criteria for significance. Once investigators have excluded chance as an explanation for differences between groups they can examine the proportions of patients who have deteriorated, remained the same, or improved as an aid in interpreting the importance of the results.

This approach emphasises the need to establish ranges of health related quality of life, symptoms, and functional status questionnaire changes that represent trivial, small but important, moderate, and large changes. When they understand these ranges, investigators reporting clinical trials should present not only mean differences but also the difference in the proportion of patients who experience important improvement, and the associated number needed to treat. Presenting the results in both ways will reduce the risk of important misinterpretation of randomised trials that directly measure aspects of living that are important to patients.

Funding: Supported in part by a grant from the Medical Research Council of Canada.

Conflict of interest: None.

1 Jaeschke R, Guyatt G, Keller J, Singer J. Measurement of health status: ascertaining the meaning of a change in quality-of-life questionnaire score. *Controlled Clin Trials* 1989;10:407-15.
2 Juniper EF, Guyatt GH, Willan A, Griffith LE. Determining a minimal important change in a disease-specific quality of life questionnaire. *J Clin Epidemiol* 1994;47:81-7.
3 Juniper EF, Guyatt GH, Feeny DH, Ferrie PJ, Griffith LE, Townsend M. Measuring quality of life in children with asthma. *Quality Life Res* 1996;5:36-46.
4 Juniper EF, Guyatt GH, Feeny DH, Ferry PJ, Griffith LE, Townsend M. Measuring quality of life in parents of children with asthma. *Quality Life Res* 1996;5:27-34.
5 Juniper EF, Guyatt GH, Griffith LE, Perrie PJ. Interpretation of rhinoconjunctivitis quality of life questionnaire data. *J Allergy Clin Immunol* 1996;98:843-5.
6 Redelmeier DA, Goldstein RS, Guyatt GH. Assessing the minimal important difference in symptoms: a comparison of two techniques. *J Clin Epidemiol* 1996;49:1215-9.
7 Laupacis A, Sackett DL, Roberts RS, An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med* 1988;318: 1728-33.
8 Jaeschke R, Guyatt G, Shannon H, Walter SD, Cook DJ, Heddle N. Basic statistics for clinicians. III. Assessing the effects of treatment: measures of association. *Can Med Assoc J* 1995;152:351-7.
9 Juniper EF, Guyatt GH, Epstein RS, Ferry PJ, Jaeschke R, Hillers TK. Evaluation of health-related quality of life in asthma: development of a questionnaire for use in clinical trials. *Thorax* 1992;47:76-83.
10 Juniper EF, Johnston PR, Borkhoff CM, Guyatt GH, Boulet LP, Haukioja A. Quality of life in asthma clinical trials: comparison of salmeterol and salbutamol. *Am J Respir Care Med* 1995;151:66-70.
11 Guyatt GH, Berman LB, Townsend M, Pugsley SO, Chambers LW. A measure of quality of life for clinical trials in chronic lung disease. *Thorax* 1987;42:773-8.
12 Goldstein RS, Gort EH, Guyatt GH, Stubbing D, Avendano MA. Prospective randomised controlled trial of respiratory rehabilitation. *Lancet* 1994;344:1394-7.
13 Duffy JC. Alcohol consumption and control policy. *J R Stat Soc A* 1991; 156:225-30.
14 The National Institute of Neurological Disorders and Stroke rt-PA Stroke Study Group. Tissue plasminogen activator for acute ishemic stroke. *N Engl Med J* 1995;333:1581-7.

(Accepted 5 October 1997)

The new genetics

Psychological responses to genetic testing

Theresa M Marteau, Robert T Croyle

Predicting disease on the basis of biological markers, such as serum cholesterol concentration or blood pressure, is not new; the ability to predict disease using DNA is. As the scope for genetic testing extends beyond testing for single gene disorders to testing large sections of the population for genes associated with common disorders it is important to consider what effect this will have on individuals and on society as a whole. Research into the psychological impact of genetic testing in Huntington's disease, cystic fibrosis, breast cancer, and ovarian cancer has shown that an individual's decision to undergo testing and his or her response on receiving the results are influenced by many factors. This article discusses the nature of these factors and the implication they have for the introduction of widespread genetic screening.

The psychological impact of screening for biological markers associated with increased risk of disease has been well researched.[1] Extrapolating from these findings to predict the impact of population based genetic screening of asymptomatic individuals is difficult, though—partly because the predictive value of genetic tests for some disorders is high, and partly

Summary points

As genetic screening becomes widespread, its psychological impact on individuals, their families, and society as a whole needs to be assessed

The psychological impact of predictive genetic testing for Huntington's disease, breast cancer, and ovarian cancer, when offered with expert counselling before and after testing, depends more on pretest expectations, mood, and social support than the results of the test itself

Distress associated with screening may be reduced by careful assessment before and after testing, counselling, and support

Research is needed to determine the most effective and practical counselling strategies for the increasingly large number of people who will be offered genetic tests for treatable conditions

This is the second of four articles discussing the broader implications of advances in genetics

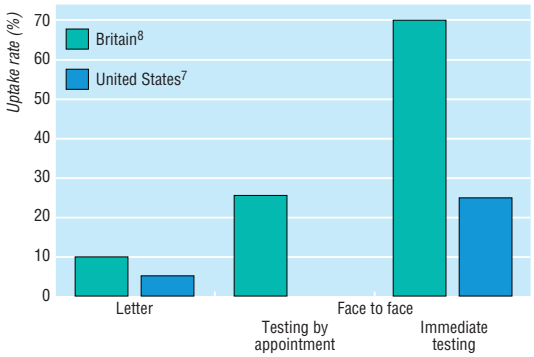
Psychology and Genetics Research Group, United Medical and Dental Schools of Guy's and St Thomas's Hospitals, London SE1 9RT

Theresa M Marteau, *professor of health psychology*

continued over

BMJ 1998;316:693–6

Department of Psychology, University of Utah, Salt Lake City, UT 84112, USA
Robert T Croyle, *associate professor of psychology*
Correspondence to: Professor Marteau t.marteau@umds.ac.uk



Uptake of carrier testing for cystic fibrosis in non-pregnant populations, according to method of offering the test

because the results of genetic testing impact on families as well as individuals.

At present, most of the genetic tests that are carried out are reproductive tests which provide information about the chances of genetic disorders in future children—for example, carrier testing for cystic fibrosis. But predictive tests, which give people information about their own chances of developing a disease, are being carried out with increasing frequency. These include presymptomatic tests for genetic mutations associated with dominantly inherited conditions with complete penetrance (having the mutation is invariably associated with disease, as for Huntington's disease) and predispositional tests, which test for gene mutations that are risk factors for a disease (having the mutation means an increased risk, but not a certainty, of developing a disease—for example, genetic testing for hereditary breast and ovarian cancer). Predispositional testing is set to become the main type of genetic test offered in the near future, as genes predisposing to common diseases such as cancer, Alzheimer's disease, heart disease, and diabetes continue to be discovered.[2]

Factors influencing the decision to undergo genetic testing

Uptake rates for genetic tests are higher when there are effective ways of treating or preventing the condition. If little can be offered, most people do not want information about their risk status. Thus the uptake for DNA predictive testing is about 10% for Huntington's disease, for which there is no treatment[3 4]; for breast cancer, for which there is some possibility of prevention and treatment, it is about 50%[5]; and it is around 80% for familial adenomatous polyposis, for which there is effective treatment.[6]

Uptake of genetic tests also depends on how a test is offered. This has been shown most clearly in population screening for cystic fibrosis. When people get invited by letter, fewer than 10% come forward—but uptake is greatly increased when testing is offered in person and if, in addition, it can be carried out immediately (fig).[7 8]

Which method of invitation is preferable is debatable. For conditions for which there is an effective treatment, high uptake may be the most important goal. If the only intervention being offered is termination of pregnancy, the quality of the decision to undergo a test, or not, may be a more appropriate goal than high uptake. There is some evidence that methods of offering tests that result in high uptake are associated with decisions based on less information and hence of poorer quality.[7]

Interest in undergoing testing is more strongly related to perceived risk than objective risk.[9 10] The extent to which individuals feel uncertain about their risk for a particular condition, their need for certainty, and the extent to which tests will provide that certainty are each important in determining whether they undergo a particular genetic test. Reducing uncertainty is one of the most common reasons for undergoing a predictive DNA test (A Binchy et al, unpublished data).[3] Women are more likely than men to undergo carrier tests,[8] presymptomatic tests,[11] and predispositional tests.[6] This may be because of differences in their knowledge about health threats and a difference in the way they cope with adverse information about their health, with men being more likely than women to engage in minimisation.[11 12] Societal and cultural factors are also important. Thus while uptake rates in the United States and Britain are similar for Huntington's disease and breast cancer, rates for cystic fibrosis are lower in the United States.[3–5 7 8] This may reflect a more negative attitude to termination of pregnancy as well as greater concern about insurance.

How individuals respond to genetic testing

People who find out that they carry a mutation that predisposes them or a possible child to a disease tend to be more distressed than those whose test results are negative, although the distress is usually within a normal range. The prediction of catastrophic reactions, including suicide, among those at risk for Huntington's disease who received unfavourable test results have not been fulfilled. Indeed, people receiving positive test results have, overall, experienced some decrease in psychological distress as the uncertainty over their genetic status has decreased.[13] Unexpectedly, some people receiving negative test results experience difficulties in adjusting to their revised risk status.[13] Among women undergoing predictive DNA testing for breast cancer, the test result seems to have relatively little impact on general levels of anxiety or depression.[5 14] Long term follow up data on psychological morbidity are available only on patients who have undergone predictive testing for Huntington's disease. They suggest that carriers do not become more distressed over time and people with negative results do not experience further decreases in psychological distress.[15] This illustrates that factors other than test results are important in predicting and understanding responses to genetic testing.

The importance of individual characteristics

Most people participating in population based screening programmes expect negative test results. Most of those who are known to be, or believe themselves to be, at high risk expect positive results.[16] This may explain why people who are aware that they are at risk experience less distress after getting the results of their test than do those who are not aware of being at

risk.[17 18] There are gender differences too, with women more likely than men to report negative feelings after genetic testing.[11 12 19] The amount of social support and the psychological resources that people have also affect their ability to cope.[20]

In Huntington's disease, a person's mood before testing is a better indication of how they react to their test result than is the result of the test itself.[11 21] Those who are distressed before they undergo genetic testing are particularly at risk of an adverse psychological outcome after testing, and they need to be identified and given additional support early on.

For many people the term genetic, in relation to an illness, carries negative connotations. They wrongly assume that an illness with a genetic cause is not preventable and not treatable. In a recent study people were asked to imagine that they had been tested by their general practitioner and found to have an increased risk of heart disease.[22] Half the participants were told that this increased susceptibility had been determined by a genetic test. For the other half, the type of test was unspecified. When risk was determined by a genetic test, heart disease was seen as less preventable.

The extent to which individuals consider a condition to be preventable is an important predictor of whether they follow advice on how to reduce the risk of developing the condition or ameliorate the condition once it has developed.[23] This may mean that if people do not consider genetic tests in the same light as tests for biological risk factors of disease they will not be motivated to change their behaviour. This view is supported by a study in which smokers randomly allocated to be tested for a genetic susceptibility to lung cancer were found to be no more likely to quit smoking than those who were not tested.[24] Those who were tested, however, perceived their risks of lung cancer as greater and were more fearful of this than those not given such information.

How testing is conducted

Initial studies of testing for Huntington's disease and breast cancer have included standardised protocols involving one or more consultations to help people decide whether to proceed with testing. The offer of the test is separated in time from the taking of a biological sample to conduct the DNA test. Such programmes also offer counselling after test results are revealed. The reason such programmes have not had catastrophic effects is likely to be due to such counselling. We do not yet know whether such elaborate counselling is necessary for people undergoing predictive testing for other conditions for which treatment is available, such as breast cancer and ovarian cancer. Research is needed to determine the most effective components of pretest and post-test counselling and the most efficient ways of providing it in anticipation of the increasingly large group of people who will be offered genetic tests.

The way risk information is presented affects how it is perceived and responded to. For example, doctors are more willing to prescribe drugs if evidence of their effectiveness is presented in terms of relative risks rather than absolute risks.[25 26] The way genetic risk information is given varies widely—as relative and absolute risks, probabilities and percentages, and numerous verbal descriptors.[27–29] One experiment with university students showed that judgments about personal vulnerability to a range of negative outcomes were sensitive to relative, but not absolute, information on risk.[30] The impact of presenting the results of genetic tests in different ways has yet to be determined.

JOANNE O'BRIEN/FORMAT

Amount of social support affects people's ability to cope with results of genetic testing

Impact of genetic testing on families

Relationships among siblings, parents, and offspring can be complicated by the different test results that individuals receive. For example, some of those found not to carry the gene for Huntington's disease were rejected by their families when they were found no longer to have one of the key bonds that had previously tied them together: being at risk for Huntington's disease.[31] Partners may be affected more than those who undergo testing—again in Huntington's disease, a study showed that the partners of those who tested positive experienced more post-test distress and poorer quality of life than did the carrier.[16] Partners of non-carriers, however, experienced less hopelessness than their tested partners. These findings suggest the importance of providing support for relatives as well as for the individual undergoing testing.

Impact on society

The conduct of research into the genetic basis for disease and, more recently, of complex behaviours, as well as the clinical provision of genetic testing, all serve to emphasise the inherited component of the human condition. Beliefs about the causes of any problem, in particular how controllable it is seen to be, influence how others respond to people who have the problem.[32] For example, health professionals have a more positive attitude towards helping patients with heart disease who do not smoke.[33] How might a genetic emphasis on the causes of illness or behaviour affect general attitudes and in particular those of health professionals?

If the effect is to make the outcome seem less controllable, a genetic emphasis may have a positive effect. So, for example, the claim of homosexuality as genetic in origin was greeted by some gay rights activists as

Best practice in genetic testing

In the light of current evidence, best practice for the conduct of genetic testing (presymptomatic, predispositional, and prenatal) includes the following points:

- The written protocol for the conduct of the testing programme should include how the laboratory tests are to be conducted and how communication with patients is to be managed
- Before they decide whether to undergo a test, clear and simple information should be presented to those eligible for testing. Such information should include the advantages and disadvantages of testing, as well as the meaning of any possible test result
- The initial offer of a test should be separated in time (a day or more) from a biological sample being taken
- Test results should be explained and support offered to all those tested and their relatives
- The effectiveness of a testing programme in achieving good understanding as well as facilitating behaviours that reduce risk, without high levels of emotional distress or false reassurance, needs to be assessed, not assumed

heralding the end of blame for such a sexual orientation. If, on the other hand, the effect of a genetic screening programme is to emphasise the importance of taking preventive actions to avoid that risk, people may regard an adverse outcome as controllable. This may result in people with such outcomes receiving less help and more blame.

In line with this, concerns have been expressed that genetic testing in pregnancy together with the offer of termination of affected pregnancies will make us less tolerant as a society towards disability and difference—and might lead to blaming of parents who do not use genetic tests and subsequently give birth to a child with a disability. Though there is indirect evidence to suggest that parents may be blamed for not undergoing tests,[34] this remains an important and little researched question.

Conclusion

Research concerning the psychological impact of genetic testing is limited, but it does suggest that adverse psychological reactions are uncommon when testing is provided within a testing programme that separates the offer of testing from the taking of a biological sample for the test, and that provides clear information and emotional support both before and after testing (box). Research is now needed to determine how much and what type of information and support are required for the increasing numbers of people being offered genetic testing, and how these are most efficiently provided, to achieve good understanding of a test and its results, as well as facilitating behaviours to reduce risk, without high levels of emotional distress or false reassurance.

Funding: TMM is supported by the Wellcome Trust; RTC is supported by grants from the National Cancer Institute.

Conflict of interest: None.

1 Croyle RT, ed. *Psychosocial effects of screening for disease prevention and detection*. New York: Oxford University Press, 1995.
2 Bell J. Genetic advances: implications for clinical medicine. *BMJ* 1998;316:618-20.
3 Craufurd D, Dodge A, Kerzin-Storrar L, Harris R. Uptake of presymptomatic predictive testing for Huntington's disease. *Lancet* 1989;i:603-5.
4 Quaid KA, Morris M. Reluctance to undergo predictive testing: the case of Huntington disease. *Am J Med Genet* 1993;45:42-5.
5 Lerman C, Narod S, Schulman K, Hughes C, Gomez-Caminero A, Bonney G. BRCA1 testing in families with hereditary breast-ovarian cancer. A prospective study of patient decision making and outcomes. *JAMA* 1996;275:1885-92.
6 Evans DGR, Maher ER, Maclead R, Davies DR, Craufurd D. Uptake of genetic testing for cancer predisposition—ethical issues. *J Med Genet* 1997;34:746-9.
7 Tambor ES, Bernhardt BA, Chase GA, Faden RR, Geller G, Hofman KJ, et al. Offering cystic fibrosis carrier screening to an HMO population: factors associated with utilization. *Am J Hum Genet* 1994;55:626-37.
8 Bekker H, Modell M, Dennis G, Silver A, Mathew C, Bobrow M, et al. Uptake of cystic fibrosis carrier testing in primary care: supply push or demand pull? *BMJ* 1993;306:1584-6.
9 Struewing JP, Lerman C, Kase RG, Giambarresi TR, Tucker MA. Anticipated uptake and impact of genetic testing in hereditary breast and ovarian cancer families. *Cancer Epidemiol Biomarkers Prev* 1995;4:169-73.
10 Marteau TM, Kidd J, Cook R, Michie S, Johnston M, Slack J, et al. Perceived risk not actual risk predicts uptake of amniocentesis. *Br J Obstet Gynaecol* 1991;98:282-6.
11 Tibben A, Frets PG, van den Kamp JJP, Niermeijer MF, Vegter-van der Vlis M, Roos RAC, et al. On attitudes and appreciation 6 months after predictive DNA test for Huntington disease in the Dutch program. *Am J Med Genet* 1993;48:103-11.
12 Marteau TM, Dundas R,, Axworthy D. Long term cognitive and emotional impact of genetic testing for carriers of cystic fibrosis: the effects of gender and test result. *Health Psychol* 1997;16:51-62.
13 Wiggins S, Whyte P, Huggins M, Adam S, Theilmann J, Bloch M, et al. The psychological consequences of predictive testing for Huntington's disease. *N Engl J Med* 1992;1401-5.
14 Croyle RT, Smith KR, Botkin JR, Baty B, Nash J. Psychological responses to BRCA1 mutation testing: preliminary findings. *Health Psychol* 1997;16:63-72.
15 Tibben A, Timman R, Bannink EC, Duivenvoorden HJ. Three-year follow-up after presymptomatic testing for Huntington's disease in tested individuals and partners. *Health Psychol* 1997;16:48-63.
16 Lynch HT, Watson P, Conway TA, Lynch JF, Slominski-Caster SM, Narod SA, et al. DNA screening for breast/ovarian cancer susceptibility based on linked markers. *Arch Intern Med* 1993;153:1979-87.
17 Marteau T, Anionwu E. Evaluating carrier testing: objectives and outcomes. In: Marteau TM, Richards M, eds. *The troubled helix: social and psychological implications of the new human genetics*. Cambridge: Cambridge University Press, 1996:123-39.
18 Marteau TM, Kidd J, Cook R, Johnston M, Michie S, Shaw RW, et al. Screening for Down's syndrome [letter]. *BMJ* 1988;297:1469.
19 Watson EK, Mayall E, Chapple J, Dalziel M, Harrington K, Williams C, et al. Screening for carriers of cystic fibrosis through primary health care services. *BMJ* 1991;303:504-7.
20 Lazarus RS, Folkman S. *Stress, appraisal and coping*. New York: Springer, 1984.
21 Decruyenaere M, Evers-Kiebooms G, Boogaerts A, Cassiman JJ, Cloostermans T, Demyttenaere K. Prediction of psychological functioning one year after the predictive test for Huntington's disease and impact of the test result on reproductive decision making. *J Human Genet* 1996;33:737-43.
22 Marteau TM, Senior V. Illness representations after the human genome project: the perceived role of genes in causing illness. In: Petrie K, Weinman J, eds. *Perceptions of illness and treatment: current psychological research and implications*. Amsterdam: Harwood Academic Press, 1997:241-66.
23 Skinner EA. Personality processes and individual differences: A guide to constructs of control. *J Pers Soc Psychol* 1996;71:549-70.
24 Lerman C, Gold K, Audrain J, Lin TH, Boyd NR, Orleans CT, et al. Incorporating biomarkers of exposure and genetic susceptibility into smoking cessation treatment: effects on smoking-related cognitions, emotions, and behavior change. *Health Psychol* 1997;16:87-99.
25 Bobbio M, Demichelis B, Giustetto G. Completeness of reporting trial results: effects on physicians willingness to prescribe. *Lancet* 1994;343:1209-11.
26 Bucher HC, Weinbacher M, Gyr K. Influence of method of reporting study results on decision of physicians to prescribe drugs to lower cholesterol concentration. *BMJ* 1994;309:761-4.
27 Allanson A, Michie S, Marteau TM. Presentation of screen negative results on serum screening for Down syndrome: variations across Britain. *J Med Screening* 1997;4:21-2.
28 Marteau TM, Plenicar M, Kidd J. Obstetricians presenting amniocentesis to pregnant women: practice observed. *J Reprod Infant Psychol* 1993;11:3.
29 Hallowell N, Statham H, Murton F, Green J, Richards MPM. "Talking about chance": the presentation of risk information during genetic counseling for breast and ovarian cancer. *J Genet Counseling* 1997;6:269-86.
30 Klein WM. Objective standards are not enough: affective, self-evaluative behavioral responses to social comparison information. *J Pers Soc Psychol* 1997;72:763-74.
31 Tibben A, Vegter-van der Vlis M, Niermeijer MF, van der Kamp JJP, Roos RAC, Rooijmans HGM, et al. Testing for Huntington's disease with support for all parties [letter]. *Lancet* 1990;335:553.
32 Billings PR, Kohn MA, de Cuevas M, Beckwith J, Alper JS, Natowicz MR. Eugenics: past, present and the future. *Am J Hum Genet* 1991;49:1109-18.
33 Holtzman NA, Shapiro D. Genetic testing and public policy. *BMJ* (in press).
34 Marteau TM, Drake H. Attributions for disability: the influence of genetic screening. *Soc Sci Med* 1995;40:1127-32.

Continuing medical education

Maintaining standards in British and Canadian medicine: the developing role of the regulatory body

Lesley Southgate, Dale Dauphinee

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

Centre for Health Informatics and Multiprofessional Education (CHIME), University College London Medical School, Whittington Hospital Campus, London N19 5NF
Lesley Southgate, *professor of primary care and medical education*
Medical Council of Canada, Ottawa, Ontario, Canada
Dale Dauphinee, *executive director*
Correspondence to: Professor Southgate l.southgate@chime.ucl.ac.uk
Series editors: Hans Asbjørn Holm and Tessa Richards

BMJ 1998;316:697–700

Summary points

In the United Kingdom and Canada there is increased interest in the accountability of medical practitioners

Regulating bodies are adopting proactive methods, and emphasis is placed on maintaining standards for registered practitioners

Identification of seriously deficient performance rests initially on peer review of actual practice

Programmes to maintain standards must include opportunities for remedial education; successful outcomes will follow only if there is identified professional, educational, and financial support
The safety of patients is the primary consideration

While health care is being reformed throughout the Western world, another change has emerged without as much public attention: the appearance of strategies to increase the degree of accountability of medical practitioners. As part of this interest in accountability, the scope of standards for practice has widened to include activities beyond the traditional actions of regulating bodies, such as dealing with doctors' misconduct and impairment and relying on patients' complaints to detect these. The approaches to and pacing of these changes differ, but the underlying trends are the same. Strategies to anticipate and prevent a decline in doctors' performance are now a central concern for regulating bodies, which are increasingly adopting proactive or interventional methods.[1 2]

We describe the recent changes in the ways in which the medical profession is regulated in the United Kingdom and some developments in Canada to enhance doctors' performance. We will discuss the implications flowing from the introduction of a minimum standard, the methods by which it is defined and assessed, and the emerging approaches to monitor and enhance doctors' performance.

Forces for change

In both North America and the United Kingdom the pressure for greater accountability of doctors is being felt at all levels of governance and regulation of practitioners: in hospital standards committees, in utilisation review by public agencies and third party payers, and in various professional regulatory and licensing bodies.[3] The movement is away from the traditional approach of primary reliance on setting standards for entry into practice and towards placing new emphasis on maintaining standards in practice.

One prominent phenomenon is that the increasing role of lay members on the governing boards of regulatory bodies has opened the self governance of the profession to wider scrutiny. For example, in Britain the General Medical Council recently increased its lay representation to 25%, at the same time reducing representation from universities and medical royal colleges. But perhaps the most powerful influence has come from the need to control healthcare costs while maintaining quality, shown in government cutbacks to healthcare budgets in Canada and the wave of managed care sweeping the United States.[4] As part of their need to be more cost effective, funding bodies are documenting practitioners' performance on a regular basis.[5] For example, the publication of report cards on the performances of individual doctors is not unheard of in the United States.[6 7] Health reforms in the United Kingdom which emphasise cost effectiveness, clinical effectiveness, and evidence based medicine have also contributed to a focus on the performance of doctors in clinical practice,[8] matched by the professional support for the introduction of the professional performance procedures by the GMC.[9]

Professional regulation and maintenance of practice standards

Britain

In the United Kingdom the GMC maintains the medical register and regulates the entry of practitioners to it. The education committee of the council (by statute an independent body) inspects and accredits medical schools awarding the basic medical degree, and over recent years the GMC has had a profound effect on the undergraduate curriculum through its publication *Tomorrow's Doctors*.[10] A national standard for medical graduates is maintained by this mechanism, underpinned by the system of external examiners, which operates between the universities. In 1997 the GMC has issued guidance about the attributes expected of new doctors by the end of the preregistration house year and in so doing has signalled its intention to require improvements in the learning experience provided for new medical graduates before they achieve full registration.[11] The guidance also includes an indication of the ways in which the progress of medical graduates should be assessed so that problems of poor performance, which may become intractable during the later stages of the doctor's career, can be identified and remedied early.

Once a practitioner is registered he or she must maintain good standing and practice in accordance with the guidance set out in *Duties of a Doctor: Good Medical Practice*.[12] This important text, which includes

DAVID HITCH

sections on relationships with patients and colleagues as well as on clinical method, describes the boundaries and standards for modern medical practice and by implication maps out the territory wherein the council's legitimate regulatory activities will be exercised in relation to individual clinical practice.

The GMC has powers to impose conditions on practice and to suspend practitioners or to remove them from the register on grounds of ill health or conduct. Since 1 September 1997 it also has direct powers within the Medical (Professional Performance) Act 1995 to assess poorly performing practitioners and to limit or suspend their practice if their performance is found to be seriously deficient. The introduction of a national standard to identify seriously deficient performance across all medical disciplines and at any stage of a doctor's career will have a profound effect on local governance and self regulation throughout the profession and should reinforce public confidence in standards of medical practice in the United Kingdom.[13]

North America

The approach to standards for licensure (registration) and the approach to maintenance of standards of practice are similar in Canada and the United States. After receiving the medical degree, all graduates must pass licensing examinations such as the qualifying examinations of the Medical Council of Canada, or the medical licensing examination in the United States.

Canadian Model for the Monitoring and Enhancement of Physician Performance

Monitoring	*Enhancement*
Initial screen of all practitioners	Feedback to all physicians
Assessment of physicians defined at risk by initial screening	Continuing medical education programmes for groups and individuals
Detailed assessment of high risk physicians	Remedial programmes

The existence of a national standard by examination for entry to the register is in direct contrast with Britain, where the function is delegated to the universities. The examination process is conducted in two (Canada) or three (United States) parts over time and serves as the basis of general licensure by the licensing authorities in the individual provinces or states. In both countries, licensure and maintenance of licensure are responsibilities of the province or state.

In common with the British GMC, Canadian licensing bodies are responsible for ongoing maintenance of standards for practitioners. Several provincial licensing authorities have developed physician review and enhancement of performance programmes (PREPP) which are designed to assess the knowledge and skills of practitioners who are referred after being identified by the complaint process, through self referral, or from referral by colleagues.[14] More recently the Federation of Medical Licensing Authorities of Canada has developed a two component model by which it intends to monitor and enhance the performance of all practising physicians in Canada, referred to as the Canadian Model for the Monitoring and Enhancement of Physician Performance (MEPP). Three national workshops of all stakeholders in the Canadian medical community established a consensus about the mechanisms for monitoring and the approaches to enhancement and remediation.[15–18] Pilot projects were carried out to assess various approaches to the monitoring aspects of the new programme. The MEPP model has three monitoring steps and three dimensions for enhancement that run in parallel and association with them (box).

The Canadian approach will in effect sample the entire register of practitioners. A different procedure has been adopted in the United Kingdom. Identification of serious deficiency of performance within the performance procedures of the GMC is triggered by a complaint to the council. Medical and lay screeners then decide whether there is a case that the doctor is dysfunctional and whether the dysfunction is best addressed within the performance procedures rather than through the health or conduct routes. With the GMC dealing with the issue of very poor performance in practice, the need for regular testing of all career doctors is reduced, particularly if individual, local, and national professional standards are maintained by self assessment, local peer review, and external review by the national professional bodies.[3]

Identifying doctors who are performing badly

Similar methods for assessing clinical performance for practitioners at risk have been adopted by the licensing bodies in Britain and Canada. Performance is initially assessed by peer review in the setting of actual practice. Tests of competence follow for those few practitioners whose standard of performance remains in doubt.

Canada

In Canada, in those provinces with existing physician review and enhancement of performance programmes, practitioners have been identified by self referral, referral by colleagues, random office audits,

Assessment by GMC during peer review visit

- Assessment of medical record keeping
- Discussion of the management of the doctor's own cases and clinical work
- Observation of aspects of actual practice
- Audit of clinical outcomes
- Interviews with third parties
- Structured interview with the doctor
- Site tour to determine the circumstances of practice

and investigation of patients' complaints. Under the first step of the monitoring component of the MEPP model, the performance of all physicians will be screened through fee for service billing patterns, peer assessment questionnaires, and patient satisfaction questionnaires.[16] These components of the monitoring mechanisms are being piloted in Alberta and Quebec by the local licensing authorities.[19 20] It is expected that the cost will be low, under £20 per physician. The vast majority of practitioners will "pass" this first cut without a problem and enhancement steps will be feedback, primarily for reassurance.

At the second step of monitoring, about 10% of practitioners could be identified from the first screen as at risk or in need and will move to a second level of assessment such as hospital audit, office audit, and structured interview of the physician. Enhancement will focus on continuing education programmes for individuals or groups.

Individually oriented monitoring and enhancement will apply to very few doctors (1-2%) and requires an objective assessment of need. The majority of the components for the third step are embodied in the existing physician review and enhancement programmes and consist of written tests of practice related knowledge, assessment of basic clinical skills by objective structured examinations, oral examinations that may include role play, and detailed interviews in order to set out a specific remedial programme for that individual.[18]

Britain

In Britain, once a doctor on the register enters the assessment stage of the professional performance procedures, the approach taken by the GMC assessors has much in common with the second and third steps in the Canadian model. The assessment will be in two phases: a peer review visit to the practice, rapidly followed by tests of competence if serious deficiency of performance cannot be ruled out by the assessors during phase one. The assessment methods, which derive their overall validity from the content of practice set out in *Good Medical Practice,* have been endorsed by the council and were implemented in September 1997.[21]

In phase one, the assessment consists of a peer review of performance conducted by a team of two medical and one lay assessor from the GMC. They will review the performance of the doctor in the setting of actual practice using equivalent methods, standards, and documentation for all disciplines (box).

In phase two the assessments take the form of tests of competence designed to assess the knowledge, skills and attitudes necessary for the practice in which the doctor is engaged. By this stage the assessors have not been able to rule out serious deficiency of performance and they will be seeking further evidence on which to come to a conclusion. The relation between competence (can do) and performance (does do) is complex, in that the first does not always predict the second. But here, performance is in doubt. By testing competence the assessors can discover, when serious deficiency is found, whether poor performance is because the doctor cannot or will not practise at an acceptable standard. This has major implications for the recommendations for remedial education and training.

Peer support and peer managed learning

The current approach in Canada is to design a learning package for the needs of the poorly performing doctor, based on the assessment at either step 2 or step 3. It is key that members of the profession have ownership in this process by functioning as evaluators and assisting in developing the remedial learning programmes. The extent of this feeling was shown at the third MEPP workshop in 1996, where the profession identified key characteristics needed for the programme to succeed:

- A peer to peer approach
- An educational contract with one to one traineeship or mentorship
- Willingness of doctors to participate in the enhancement process
- Support from the various medical organisations.

There was also strong support for feedback that was as immediate as possible and non-judgmental.[17]

In Britain the primary purpose of the assessments within the performance procedures is to describe the performance of the doctor in practice in a degree of detail which gives a sound basis for the decisions that must be taken within the GMC about the doctor's fitness to practise. The intention is to restore the doctor to effective clinical practice, providing patients or the public are not placed in jeopardy.[13] The report from the assessors will be available to the doctor, who will be able to use it to plan a programme of remedial education and training with the help of the regional postgraduate dean or the regional director of postgraduate general practice education.

Nature of remedial education

In Canada it is proposed that remedial education be individually oriented, based on the assessment and

Assessment after peer review of practice

Britain	*Canada*
Tests of competence in phase 2 of performance procedures	Tests of competence in step 3 of MEPP
Test of practice related knowledge and clinical thinking in written or oral form	Written test of practice related knowledge
Consultation skills (clinical thinking and communications skills; standardised patients are often used[23])	Basic clinical skills tested within a simulated office oral or structured clinical examination (often with standardised patients)
Practical clinical skills using structured clinical examination	

degree of risk, and that it be one to one and non-judgmental. The real question, which is not yet solved, is who will pay and whether the universities and their affiliated teaching institutions will be able to accept the challenge on these terms.

Similar approaches will be adopted in Britain, where the situation is still fluid, with wide recognition of the problems associated with remedial education and training at this standard. Not least, the ethical and legal framework within which these doctors can see and treat patients under supervision needs further clarification and the educational support, NHS management, and financial support for the clinical teams who undertake to receive them must be identified. The cost of the assessment will be borne by the GMC, and financial support for the doctor to undertake a remedial programme may be available from NHS trusts and health authorities. What is clear in both Britain and Canada is that a successful outcome is possible only if the individual doctor accepts the responsibility both for undertaking the remedial programme and for achieving an improvement in clinical performance.

Continued poor performance

In both the Canada and Britain the licensing authorities take the responsibility for doctors who remain below the acceptable level of performance after a remedial programme. In the current Canadian enhancement programmes in certain provinces, this paradigm is well established and tested. Results in some programmes indicate that a matter of fact but non-judgmental approach will work for most practitioners with deficits, but a very small number will represent too high a risk to their patients. For that group the judgment of the licensing authorities becomes operative, as it does for doctors who are unable to raise their level of performance after the remedial programme. The experience in Ontario has shown that some individuals are too far down the scale to recover because of poor cognition or deteriorated knowledge and skills.[22] For them the solution is negotiated retirement, knowing full well that the licensing authorities have the responsibility and mandate to act.

Experience in Britain will build up and be made public in the years after the introduction of the performance procedures. At this stage it is impossible to predict the outcome of remediation and reassessment, although the Canadian experience seems relevant.

1 Norman GR, Davis DA, Lamb S, Hanna E, Caulford P, Kaigas T. Competency assessment of primary care physicians as part of peer review program. *JAMA* 1993;270:1046-51.
2 Cunnington JPW, Hanna E, Turnbull J, Kaigas T, Norman GR. Defensible assessment of the competency of the practising physician. *Acad Med* 1997;72:9-12.
3 Irvine D. The performance of doctors. 1. Professionalism and self regulation in a changing world. *BMJ* 1997;314:1540-2.
4 Dauphinee WD. Assessing clinical performance. Where do we stand and what might we expect? *JAMA* 1995;274:741-3.
5 Brook RH, Kamberg CJ, McGlynn EA. Managing the quality of care. *JAMA* 1996;276:476-80.
6 Topol EJ, Califf RM. Scorecard cardiovascular medicine: its impact and future directions. *Ann Intern Med* 1994;120:65-70.
7 Green J, Wintfield N. Report cards on cardiac surgeons: assessing New York State's approach. *N Engl J Med* 1995;332:1229-32.
8 Secretary of State for Health. *Primary care: the future*. London: HMSO, 1996.
9 General Medical Council. *The new performance procedures: consultative document*. London: GMC, 1997.
10 General Medical Council. *Tomorrow's doctors: recommendations on undergraduate medical education*. London: GMC, 1993.
11 General Medical Council. *The new doctor: recommendations on general clinical training*. London: GMC, 1997.
12 General Medical Council. *Duties of a doctor: good medical practice*. London: GMC, 1995
13 Irvine D. The performance of doctors. 2. Maintaining good practice, protecting patients from poor performance. *BMJ* 1997;314:1613-5.
14 Page GG, Bates J, Dyer SM, Vincent DR, Bordage G, Jacques A, Sindon A, et al. Physician-assessment and physician-enhancement programs in Canada. *Can Med Assoc J* 1995;153:1723-8.
15 Lescop J, Kaigas T, Waymouth V, Smith S. *Proceedings of the workshop on maintenance of competence - monitoring of performance*. Ottawa: Federation of Medical Licensing Authorities of Canada, 1994.
16 Lescop J, Kaigas T, Waymouth V, Smith S. *Proceedings of the Aylmer II workshop on monitoring and enhancement of physician performance*. Ottawa: Federation of Medical Licensing Authorities of Canada, 1995.
17 Lescop J, Kaigas T, Waymouth V, Smith S. *Proceedings of the Aylmer III workshop on monitoring and enhancement of physician performance. Building the bridges between enhancement and monitoring in the Canadian model: the feedback continuum*. Ottawa: Federation of Medical Licensing Authorities of Canada, 1996.
18 Federation of Medical Licensing Authorities of Canada. *Position paper on the Canadian model for monitoring and enhancement of physician performance*. Ottawa: FMLAC, 1996.
19 Ward B. Update on MEPP project: implementation of step I across Canada. Report from Alberta. Annual Meeting of the Federation of Medical Licensing Authorities of Canada, Ste Adele, Quebec, 29 May 1997.
20 Jacques A. Update on MEPP project: implementation of step I across Canada. Report from Quebec. Annual Meeting of the Federation of Medical Licensing Authorities of Canada, Ste Adele, Quebec, 29 May 1997.
21 Southgate L. *Performance procedures: a report on the development and piloting of the assessment methods*. General Medical Council, 21 May 1997.
22 Caulford PG, Lamb SB, Kaigas T, Hanna E, Norman GR, Davis DA. Physician incompetence: specific problems and predictors. *Acad Med* 1994;69(suppl):S16-8.

A strange experience

Taking precautions

I once tried a door to see if it was locked and it wasn't, so I opened it. This was nearly 56 years ago, when I was one of several wounded prisoners of war in a hospital near Naples. We were having a brief supervised walk in the hospital grounds and I contrived not to be noticed for a minute or two's exploration. What I saw made quite an impression—this was before I went to medical school—for I had stepped into a mortuary. In the centre of the roof was a large bell. And hanging down from it were several chains, each—except for a few spare ones—attached to the wrists of one of the corpses lying on slabs around the walls. I am reminded of this whenever tabloid headlines announce that someone supposed to be dead has turned out not to be.

Thurstan Brewin, *retired clinical oncologist, Oxford*

We welcome articles up to 600 words on topics such as *A memorable patient, A paper that changed my practice, My most unfortunate mistake,* or any other piece conveying instruction, pathos, or humour. If possible the article should be supplied on a disk. Permission is needed from the patient or a relative if an identifiable patient is referred to. We also welcome contributions for "Endpieces," consisting of quotations of up to 80 words (but most are considerably shorter) from any source, ancient or modern, which have appealed to the reader.