For Debate . . .

Standards for the use of ordinal scales in clinical trials

C RONALD MacKENZIE, MARY E CHARLSON

Abstract

Ordinal scales are frequently used in clinical trials to quantify outcomes which are non-dimensional. They may be regarded as either single state or transition measures based on whether they assess the outcome at a single point in time or directly examine change which has occurred between two points in time. Each has unique structural and operating characteristics, so that different methodological standards for their construction and utilisation are required. All trials employing ordinal scales published in three leading journals between 1980 and 1984 were examined.

For both types of scales the individual ranks must be clearly defined, mutually exclusive, and ordered in a hierarchical progression. Further, both types must be able to detect equally both improvement and deterioration. For this to be ensured with the single state scales the population under study must not be clustered at one extreme of the scale at entry into the trial. With the transition scales internal symmetry must be achieved. Strategies for determining the comparisons to be performed should include emphasis on within patient analysis for crossover trials. Concordance between scale scores and the other measures of outcome employed in the trial must be evaluated.

Frequent violations were uncovered in the studies reviewed, and it is hoped that the simple rules outlined will prove useful in the planning and evaluation of future trials.

Introduction

In clinical medicine ordinal scales are frequently used to quantify phenomena which are not truly dimensional. Well known examples include the 1+ to 4+ system for assessing peripheral oedema, the New York Heart Association classification for patients with cardiac failure,[1] and the TNM cancer staging system.[2] More complex scales have been developed to measure phenomena such as functional state[3 4] and disease severity in acutely ill patients.[5] These scales are single state measures because they are designed to assess patients at a single point in time. In a clinical trial such scales are often employed to establish a patient's state at entry and again at the trial's conclusion. Differences between assessments are then attributed to the treatment. In contrast, transition scales measure this change directly. Regardless of their baseline state, patients report whether they are better, the same, or worse at the conclusion of the trial.

Guidelines for the construction and utilisation of ordinal scales as measures of outcome in clinical trials have not been formalised. In this paper a set of standards are developed and applied to recent trials which have employed such scales as measures of outcome. Since single state and transition scales pose different methodological problems, the two types of scales and the respective examples from publications are analysed separately.

Methods

The January 1980 to December 1984 issues of three leading journals were reviewed. These were the *New England Journal of Medicine*, *Annals of Internal Medicine*, and the *British Medical Journal*. Studies were included if they met the following criteria: (*a*) the study was a clinical trial, and (*b*) at least one ordinal scale was used to measure outcome. Twenty six studies using 44 scales met these criteria.[6-31]

For each study compliance with the pertinent methodological standard was rated according to whether (*a*) the standard was met, (*b*) the standard was violated, or (*c*) compliance with the standard could not be evaluated or the standard did not apply to the scale in question (tables I and II). Each study was reviewed separately and independently by each of us but no explicit attempt was made to determine interobserver agreement. Table III lists the scales specifically discussed below.

Single state scales

SCALE RANKS

The first standard is that scales must be composed of individual elements which are clearly defined and the ranks must be discrete and non-overlapping (mutually exclusive). Failure to ensure this may create serious ambiguities, as seen in a trial which examined the benefits of adrenergic blockade in patients with subarachnoid haemorrhage. Patients were evaluated using an ordinal, single state measure of neurological outcome.[6] The scale ranks were (1) alert, no neurological signs; (2) alert, minor neurological signs; (3) alert, major signs; (4) drowsy, minor signs; (5) drowsy, major signs; and (6) comatose. Outcome was deemed to be good if a patient was grade 1 or 2 and poor if grade 3 or more. The signs that made up the individual ranks (minor and major neurological signs, drowsiness, coma) were never defined. As a result we cannot be sure what the ranks actually mean—Is a patient who is drowsy with minor neurological signs really worse off than an alert patient with major signs? Is hemiparesis a major or a minor sign? When does severe drowsiness become coma? This lack of definition makes it impossible to assess whether the ranks are mutually exclusive or whether the order is reasonable. This is crucial in interpreting the report because the reported benefits of treatment depended entirely on where the cut off point was placed. If, for instance, grade 4 (drowsy, minor signs) were considered a good outcome, the overall conclusions would be different.

Thus the first guideline requires that all the elements of the scale must be clearly defined and the ranks must be discrete. This standard was met in 24 instances, violated in three, and was not evaluable in the remaining 12 examples (table I).

Not only must the individual scale ranks be clearly defined but the order of the ranks must be reasonable. Ambiguously ordered ranks or the inclusion of qualitatively different phenomena within the scale may create difficulties, as seen in a recent study of patients receiving intensive immunosuppression for progressive multiple sclerosis.[7] In that study two ordinal scales were used as outcome measures. The disability status scale is based on disability and accompanying neurological signs, while the ambulation index quantifies a patient's ambulatory state. With the disability status scale patients are assigned to grade 6 if they require a mechanical aid (canes, crutches, or

Department of Medicine, Cornell University Medical Center, New York NY 10021

C RONALD MacKENZIE, MD, assistant professor of medicine
MARY E CHARLSON, MD, associate professor of medicine and Henry J Kaiser Family Foundation faculty scholar in general internal medicine

Correspondence and requests for reprints to: Dr MacKenzie.

braces) for walking. In contrast, all other ranks depend on the disability experienced by the patient. Taken to extreme, a patient who walks with a cane but works and is self sufficient might be rated as worse off than a patient who is not able to care for himself. In a scale designed to measure disability, including a category based solely on the use of mechanical aids for walking is not sensible and raises questions about whether someone with grade 6 disability is actually worse off than someone with grade 4. As grade 6 was the most frequent rank among patients, regardless of the treatment group, the results in the study using the disability status scale are difficult to interpret.

The study also used an ambulation index which measures physical ability based on both the time required and distance walked as well as the need for mechanical aids. In this scale there is a reasonable hierarchical progression of scale ranks; it seems clear that a patient with grade 4 mobility is worse off than the patient with grade 3.

Thus the second guideline for the single state scales is that the scale elements must assess the same phenomena and be ordered in a reasonable hierarchical manner. This guideline depends directly on the presence of well defined mutually exclusive ranks, without which the hierarchical progression cannot be assessed. Of the 39 examples, 22 satisfied this standard and three did not. In the remaining 14 compliance was not evaluable (table I).

CLINICAL CONTEXT

As scales frequently measure a constellation of clinical phenomena, it may be difficult to know the true clinical importance of the overall scale scores or score differences. This is particularly true with the more complex scales. For example, with more simple ordinal scales such as the New York Heart Association classification a clinician can perceive whether a change from class III to class II is important. With the more elaborate scales, however, the meaning of a given change in score becomes more uncertain. A statistically significant difference between treatment groups may not be of much clinical relevance.

One method of addressing this problem is for the investigator to provide a context for interpreting the scores or differences in scores. An example of this approach is illustrated in a study on the impact of free medical care on adult health.[8] Data on the general health of a large population were obtained by a questionnaire which is scored on a standardised scale of 1-100. A context

TABLE I—*Compliance with guidelines for single state scales**

Guideline	Standard met	Standard violated	Standard not evaluable
(1) Are the scale elements clearly defined and discrete?	24	3	12
(2) Are the scale ranks arranged in a straightforward hierarchical manner?	22	3	14
(3) Is there reason to believe that a given change in scale score reflects a clinically meaningful difference?	18	0	21
(4) Does the scale range allow for the equal measurement of improvement and deterioration in the population?	23	4	12
(5) Was the strategy used in the comparison appropriate?	22	14	3
(6) Were the outcomes obtained with the scale similar to other measured outcomes? If not, do the authors offer a plausible reason?	29	6	4

* Table refers to 37 scales used in 39 instances in studies under review.

TABLE II—*Compliance with guidelines for transition scales*

Guideline	Standard met	Standard violated
(1) Are the scale elements clearly defined and discrete?	7	1
(2) Are the scale ranks arranged in a straightforward hierarchical manner?	7	1
(3) Is there reason to believe that a given change in scale score reflects a clinically meaningful difference?	7	1
(4) Is the scale constructed to allow for equal measurement of improvement and deterioration (symmetrical)?	4	4
(5) In a crossover trial was the strategy used in the comparison appropriate?	6	2
(6) Were the outcomes obtained with the scale similar to other measured outcomes? If not, do the authors offer a plausible reason?	5	3

TABLE III—*Examples of single state and transition scales*

Single state scales	Transition scales
Heart failure score[9]; point value 1-4 based on severity of: Dyspnoea Crackles Abnormal heart rate Right heart failure Chest film abnormality	Patient evaluation[13]: 1 = Excellent 2 = Good 3 = No change 4 = Worse
Skin surface damage score[10]: 0-3 Redness 0-3 Surface damage 0-3 Lichenification (Total × 20 zones = skin surface score)	Symptom score[14]: −2 = Worsening −1 = New symptom 0 = Improvement +2 = Resolution
Neurological grade[6]: 1 = Alert, no neurological signs 2 = Alert, minor neurological signs 3 = Alert, major signs 4 = Drowsy, minor signs 5 = Drowsy, major signs 6 = Comatose	Patient evaluation[16]: −3 = Much, much worse −2 = A lot worse −1 = A little worse 0 = The same +1 = A little better +2 = A lot better +3 = Much, much better

Disability status scale[7]:
  0 = Normal neurological findings
  1 = No disability; minimal neurological signs
  2 = Minimal disability
  3 = Moderate disability but ambulatory
  4 = Relatively severe disability; ambulatory, self sufficient 12 hours/day
  5 = Disability precludes full day's work without provisions; walks unaided only for several blocks
  6 = Device required for walking (cane, crutches, walker)
  7 = Wheelchair bound, able to wheel, enters and leaves chair unaided
  8 = Bedbound, able to use arms
  9 = Totally helpless and bedridden

Ambulation index[7]:
  0 = Asymptomatic; fully active
  1 = Walks normally; fatigue interferes with demanding activities
  2 = Abnormal gait or episodic imbalance noted by family; walks 25 ft in 10 s
  3 = Walks independently, 25 ft in 20 s
  4 = Requires unilateral support (cane, crutch); walks 25 ft in 20 s
  5 = Requires bilateral support (canes, walker) and walks 25 ft in 20 s; unilateral support but requires >20 s to walk 25 ft
  6 = Requires bilateral support and ≥20 s to walk 25 ft; uses wheelchair occasionally
  7 = Walks only few steps with bilateral support; unable to walk 25 ft; mostly in wheelchair
  8 = Restricted to wheelchair, transfers independently
  9 = Restricted to wheelchair, unable to transfer

Ritchie index[11 32]:
  0 = Not tender
  1 = Tender
  2 = Tender and winces
  3 = Tender, winces, and withdraws

for understanding the meaning of the scale scores is provided by clear definitions for both "good" and "ill" health based on the entry distribution of scale scores. For instance, being "ill" was defined as being in the lowest fifth of the score distribution at enrolment. Even more important, *clinical* examples were provided to describe the meaning of the various subscale score differences. As an example, a 10 point difference on the physical subscale would be equivalent to the effect of mild, chronic osteoarthritis. Such descriptions are useful for providing the necessary information to make the results interpretable so that their clinical usefulness can be assessed.

Thus the third guideline is that the scale scores must be placed in a clinical context. This is of particular importance as the scales increase in complexity and the meaning of a score becomes more obscure. In the review 18 examples satisfied this requirement, while in 21 it could not be evaluated (table I).

SCALE RANGE

The most fundamental requirement for scales used as measures of outcome is that they must detect equally both improvement and deterioration in the population under study. To do this a single state scale must have a sufficient range to encompass the spectrum of the phenomenon in the population. For example, if the population under study lies at one extreme or the other on the scale at the beginning of the study change in one direction may be impossible to detect.

In a recent crossover trial oral digoxin was evaluated in patients with chronic congestive heart failure.[9] Heart failure was assessed in 25 patients using a scale composed of five clinical and radiological observations, each assigned a score 0-4 based on severity. A total score of zero indicated the best clinical state and 20 the worst. Although the patients had had a mean score of 8 at some point before the trial, their mean heart failure score was 1·2 (SD

1·4) at the beginning of the study, close to the "best" possible score. As a result the study population lay at one extreme (the "best" end) of the scale; thus immediately before the trial the measurement of change in one direction—in this case improvement—was constrained. As measured by the scale, digoxin was efficacious overall. Among the 11 patients who did not respond to digitalis, only one had a third heart sound (S3) or gallop rhythm, typical of patients with decompensated cardiac failure. The authors concluded that digitalis did not work in patients without a third heart sound. Of the 11 non-responders, however, seven had heart failure scores of zero (the best possible score) when receiving placebo; therefore, if improvement had occurred with digoxin it could not have been detected by the scale (since zero was the best possible score and the effect of treatment was measured as placebo (that is, zero) minus digoxin). Though the study showed an overall beneficial effect of digoxin, the authors' claim that only those patients who had an S3 gallop benefited is not warranted; in fact, benefit may have occurred that was not measured because the heart failure score cannot detect improvement in patients without signs of decompensated cardiac failure.

Whether the clustering of patients at one extreme on a scale may pose an important problem in utilising the scale may be assessed by examining where the patients are on the scale at entry to the study. In a recent investigation of children with severe atopic eczema treated with combined oral and nasal beclomethasone a surface damage score was calculated.[10] The skin surface was divided into 20 separate zones of roughly equal area. A score (0-3) was recorded for each zone for (a) redness, (b) surface damage (vesiculation, crusting, excoriation, or any combination), and (c) lichenification. The scores were summed to give a total of 0-60 for each feature. At entry into the study the mean scores ranged from 19 to 26. As the patients can move from anywhere between zero and 60 on the scale measurement of both improvement and deterioration was possible.

Thus the fourth requirement for the use of single state scales in the measurement of outcome is that the scale must be able to measure improvement and deterioration equally in the population under study. Though there may be circumstances in which change will occur in only one direction (TNM cancer staging), these are the exception rather than the rule. Whether this standard is met may be assessed by examining the patients' scores at entry to the trial. If the patients are clustered at one extreme on the scale, then the scale may be unable to detect a change occurring in one direction. When such circumstances were present we deemed the standard to be violated. Of the 39 instances in which single state scales were employed, this requirement was met in 23 and violated in four instances. In the remaining 12 little or no information was provided about the patients' scores at entry into the trial, so the standard could not be evaluated (table I).

USE OF SINGLE STATE SCALES

Single state ordinal scales may be used in two ways to assess outcomes. A between patient analysis compares the outcomes in two or more groups. For example, if a trial of digoxin in congestive heart failure were being analysed in this manner the heart failure scores in patients treated with digoxin would be compared with those in patients receiving placebo; differences between groups would be attributed to the treatment. The problem with this analysis is that the patients may have entered the trial with very different initial scores. The between group analysis assesses only which group is best overall at the end of the trial, not the extent of improvement in each group. If, instead, the differences between the scores at entry and those at completion of the trial are calculated for each patient and the differences examined, a within patient analysis would be performed. In trials using ordinal scales as outcome measures it is the within patient changes that are most important. Such within patient analyses must be performed in crossover trials, where, in this example, the score during digoxin treatment would be compared with the same patient's score while receiving placebo.

In a recent report on the use of lymphoplasmapheresis in rheumatoid arthritis 14 patients were randomised to receive either intensive plasmapheresis or sham treatment.[11] Evaluations comprised a range of laboratory and clinical measurements which included subjective assessments of wellbeing and a system for evaluating joint tenderness and inflammation known as the Ritchie index.[32] With the scale of wellbeing patients were asked to score themselves on a scale of 1 to 10 (10 representing best).

For both the scale of wellbeing and the Ritchie index an examination of the within patient differences would be the preferred comparison in a trial of this design. Instead, the authors analysed the scores in the two groups at the end of the treatment period using a between patient comparison. For example, with the scale of wellbeing the score in the placebo group (6·14; SD 1·67) was compared with that in the treatment group (6·78; SD 0·756); the authors reported the difference to be significant (p<0·05). While it is difficult to attach much clinical importance to differences of this magnitude, the key problem with this approach is that the between patient comparison obscures what happens to the individual. As the study's conclusions hinged on the apparent improvements in these two outcomes, and all other clinical variables (walking time, grip strength, and morning stiffness) failed to improve, the authors' claims regarding efficacy are difficult to accept.

In contrast, a study of the effectiveness of a geriatric evaluation unit used a within patient analysis.[12] Patients ready for discharge from an acute care hospital were randomly assigned to either a geriatric evaluation unit or to receive routine follow up care and then followed up for a year. Physical and mental functioning were assessed using various standardised ordinal scales. In patients who survived at one year between patient analysis showed no significant differences between the groups. When the surviving patients in each group were examined using a within patient analysis, however, three out of the four scales utilised in the study showed significant differences favouring the geriatric unit.

Thus the fifth guideline assesses whether the analysis of the data centred on the within patient differences. In 22 instances the correct analysis was performed, while in 14 it was not. In three examples the information was insufficient to assess the strategy employed (table I).

Transition scales

Feinstein and Wells first introduced the concept of a "transition taxonomy" for scales used for directly evaluating change and suggested a useful design for their construction.[33] By this method change is detected using arbitrary ordinal ratings that represent both the magnitude and the direction of the change in question. As with the single state scales the individual ranks must be clearly defined, mutually exclusive, and arranged in a hierarchical manner. One additional characteristic influences the performance of these scales—namely, symmetry.

SCALE SYMMETRY

The concept of scale symmetry relates to the basic requirement that to measure outcomes a scale must detect equally improvement and deterioration. A double blind crossover trial of the treatment of Raynaud's phenomenon compared the serotonin antagonist ketanserin with placebo.[13] An ordinal transition scale by which patients evaluated the effect of treatment on symptoms was employed, constructed as (1) excellent, (2) good, (3) no change, and (4) worse. This asymmetric design, which provides two opportunities for improvement (excellent, good) against one for deterioration (worse), increases the likelihood of a positive response (improvement) and could potentially bias the results.

In another study thyroxine replacement therapy in patients with subclinical hypothyroidism was examined.[14] A different approach was adopted in the construction of the symptom score. A transition scale was utilised composed of five ordinal ranks: (-2) worsening, (-1) development of a new symptom, (0) no change, ($+1$) improvement, ($+2$) resolution. Despite the apparent symmetry of the ranks (-2 to $+2$), the inclusion of "new symptom" (-1) destroys its internal symmetry and confounds the interpretation of the results.

Thus transition scales must be symmetrical in their structure. This guideline was satisfied by four of the eight transition scales found in the literature review (table II).

USE OF TRANSITION SCALES

In contrast with the single state scales, transition measures assess directly within patient differences. In the usual trial patients are assigned to one treatment or another and the scale scores at entry are later compared with those at completion of the study; transition scales assess the difference directly, thereby achieving a within patient analysis. In a crossover trial the transition score during active treatment must be compared with the score during placebo treatment. Failure to make this comparison nullifies the advantage of the crossover design. In an investigation concerning the treatment of Raynaud's phenomenon nifedipine was compared with placebo in a crossover fashion.[15] Although each patient received the two treatments and the transition scores were reported for both, the response of a given patient (to treatment) was never analysed in relation to the placebo scores. Instead, the authors analysed the total scores for the treatment and placebo periods using a between patient analysis; benefit (moderate to pronounced improvement) was reported in nine of the 15 patients while receiving nifedipine versus two while receiving placebo. When, however, differences between nifedipine and placebo in individual patients are examined (within patient comparison) only four of the 15 patients had experienced a moderate to pronounced improvement with nifedipine (as compared with placebo); of the remaining patients, 11 experienced, at best, minimal improvement in their symptoms.

In contrast, in a study of plasmapheresis among patients with rheumatoid arthritis a crossover design with sham apheresis as the placebo was employed.[16] Various assessments of disease activity were employed and included the patients' subjective evaluation of joint symptoms using an ordinal transition scale: (−3) much, much worse, (−2) a lot worse, (−1) a little worse, (0) the same, (+1) a little better, (+2) a lot better, and (+3) much, much better. All assessments were performed before the trial (time 1) and then after one, two, three (time 4), eight, and 12 weeks of treatment. The analysis proceeded as appropriate for a crossover trial with differences occurring in the treatment intervals calculated for each patient. For the subjective evaluations (transition scale) there was by definition no baseline (time 1) assessment. Thus the comparison performed was apheresis time 4 minus sham treatment time 4. At the conclusion of the study no significant differences were found between the groups.

When the studies were evaluated for the use of within patient differences six were found in which the preferred analysis was performed (table II).

Concordance

The concept of concordance applies to both the single state and transition scales and reflects whether the scales yield results that are consistent with the other measured outcomes. Differences, if present, should be reasonably explained. For example, we should expect that patients experiencing an exacerbation of rheumatoid arthritis would show increases in sedimentation rate along with a deterioration in their American Rheumatism Association classification. Thus when other outcome measures are employed it is a useful and easily applied validation procedure to examine whether all the outcomes of interest are concordant. Though the correlations will not be perfect, we should be concerned when similar trends do not occur, unless the authors offer a plausible explanation for the dichotomy. When the concept was applied to all 47 examples (single state and transition scales combined) 34 satisfied the criteria and nine did not. In four instances scale concordance could not be evaluated or was not applicable because only a single measure of outcome was utilised.

Conclusion

In this review we have attempted by example to delineate some of the problems with the application of ordinal scales to measure outcomes in clinical research. Guidelines easily applied by the clinician or the investigator have been developed to serve as methodological criteria for the evaluation of research using scoring systems to assess clinical outcomes. We hope that these simple rules will provide a useful framework by which future investigation may be planned and evaluated.

References

1 Criteria Committee of the New York Heart Association. *Diseases of the heart and blood vessels. Nomenclature and criteria for diagnosis.* 6th ed. Boston: Little, Brown and Company, 1964:112-3.
2 UICC: International Union Against Cancer. *TNM Classification of malignant tumours.* Geneva: International Union Against Cancer, 1974:51-5.
3 Bergner M, Bobbit RA, Carter WB, Gilson BS. The sickness impact profile: development and final revision of a health status measure. *Med Care* 1981;19:787-805.
4 Nelson E, Conger B, Douglass R, et al. Functional health status levels of primary care patients. *JAMA* 1983;249:3331-8.
5 Wagner DP, Knaus WA, Draper EA. Statistical validation of a severity of illness measure. *Am J Public Health* 1983;73:878-84.
6 Walter P, Neil-Dwyer G, Cruickshank JM. Beneficial effects of adrenergic blockade in patients with subarachnoid haemorrhage. *Br Med J* 1982;284:1661-4.
7 Hauser SL, Dawson DM, Lehrich JR, et al. Intensive immunosuppression in progressive multiple sclerosis. A randomized, three-arm study of high dose intravenous cyclophosphamide, plasma-exchange, and ACTH. *N Engl J Med* 1983;308:173-80.
8 Brook RH, Ware JE, Rogers WH, et al. Does free care improve adults' health? Results from a randomized controlled trial. *N Engl J Med* 1983;309:1426-33.
9 Lee DCS, Johnson RA, Bingham JB, et al. Heart failure in outpatients. A randomized trial of digoxin versus placebo. *N Engl J Med* 1982;306:699-705.
10 Heddle RJ, Soothill JF, Bulpitt CJ, Atherton DJ. Combined oral and nasal beclomethasone dipropionate in children with atopic eczema: a randomised controlled trial. *Br Med J* 1984;289:651-3.
11 Wallace D, Goldfinger D, Lowe C, et al. A double-blind controlled study of lymphoplasmapheresis versus sham apheresis in rheumatoid arthritis. *N Engl J Med* 1982;306:1406-10.
12 Rubinstein LZ, Josephson KR, Wieland D, et al. Effectiveness of a geriatric evaluation unit. A randomized controlled trial. *N Engl J Med* 1984;311:1664-70.
13 Roald OK, Seem E. Treatment of Raynaud's phenomenon with ketanserin in patients with connective tissue disorders. *Br Med J* 1984;289:577-8.
14 Cooper DS, Halpern R, Wood LC, et al. L-Thyroxine in subclinical hypothyroidism. a double-blind, placebo-controlled trial. *Ann Intern Med* 1984;101:18-24.
15 Rodeheffer RJ, Rommer JA, Wigley F, et al. Controlled double-blind trial of nifedipine in the treatment of Raynaud's phenomenon. *N Engl J Med* 1983;308:880-3.
16 Dwosh IL, Giles AR, Ford PM, et al. Plasmapheresis therapy in rheumatoid arthritis. A controlled, double-blind, crossover trial. *N Engl J Med* 1983;308:1124-9.
17 Carpenter WT, Sadler JH, Light PD, et al. The therapeutic efficacy of hemodialysis in schizophrenia. *N Engl J Med* 1983;308:669-75.
18 Ericsson CD, Dupont HL, Sullivan P, et al. Bicozamycin, a poorly absorbable antibiotic, effectively treats traveller's diarrhea. *Ann Intern Med* 1983;98:20-5.
19 Fischer BH, Marks M, Reich T. Hyperbaric-oxygen treatment of multiple sclerosis. A randomized, placebo-controlled, double-blind study. *N Engl J Med* 1983;308:181-6.
20 Nicotra MB, Rivera M, Awe RJ. Antibiotic therapy of acute exacerbations of chronic bronchitis. A controlled study using tetracycline. *Ann Intern Med* 1982;97:18-21.
21 Bryson YJ, Dillon M, Lovett M, et al. Treatment of first episodes of genital herpes simplex virus infection with oral acyclovir. *N Engl J Med* 1983;308:916-20.
22 Kroc Collaborative Study Group. Blood glucose control and the evolution of diabetic retinopathy and albuminuria. A preliminary multi-center trial. *N Engl J Med* 1984;311:365-71.
23 Allen GS, Ahn HS, Preziosi TJ, et al. Cerebral arterial spasm—a controlled trial of nimodipine in patients with subarachnoid hemorrhage. *N Engl J Med* 1983;308:619-24.
24 Veith RC, Raskind MA, Caldwell JH, et al. Cardiovascular effects of tricyclic antidepressants in depressed patients with chronic heart disease. *N Engl J Med* 1982;306:954-9.
25 Anderson JL, Marshall HW, Bray BE, et al. A randomized trial of intracoronary streptokinase in the treatment of acute myocardial infarction. *N Engl J Med* 1983;308:1312-8.
26 Finlay IG, Scott R, McArdle CS. Prospective double-blind comparison of buprenorphine and pethidine in ureteric colic. *Br Med J* 1982;284:1830-1.
27 Kirk P, Lennard-Jones JE. Controlled trial of azathioprine in chronic ulcerative colitis. *Br Med J* 1982;284:1291-2.
28 Robson MH, France R, Bland M. Clinical psychologist in primary care: controlled clinical and economic evaluation. *Br Med J* 1984;288:1805-8.
29 Morrow GR, Morrell C. Behavioral treatment for the anticipatory nausea and vomiting induced by cancer chemotherapy. *N Engl J Med* 1982;307:1476-80.
30 Intermittent Positive Pressure Breathing Trial Group. Intermittent positive pressure breathing therapy of chronic obstructive pulmonary disease. A clinical trial. *Ann Intern Med* 1983;99:612-20.
31 Nocturnal Oxygen Therapy Trial Group. Continuous or nocturnal oxygen therapy in hypoxemic chronic obstructive lung disease. A clinical trial. *Ann Intern Med* 1980;93:391-8.
32 Ritchie DM, Boyle JA, McInnes JM, et al. Clinical studies with an articular index for the assessment of joint tenderness in patients with rheumatoid arthritis. *Q J Med* 1968;37:393-406.
33 Feinstein AR, Wells CK. A new taxonomy for rating change in functional activities of patients with angina pectoris. *Am Heart J* 1977;93:172-82.

What advice should be given to relatives about the likely duration of a victim's life after injury to the spinal cord producing paraplegia and quadriplegia?

Patients who sustain a spinal injury have a reduced life expectancy. Until spinal centres were developed, as the result of the second world war, all these patients died within a few months of injury from severe pressure sores and ascending urinary tract infection. As a result of the development of specialist centres, patients have a much improved life expectancy and, provided that they receive adequate treatment, a paraplegic patient's life expectancy is reduced by about 10 years, mainly due to the risk of developing pressure sores. Patients who have higher lesions—that is, tetraplegia—are always at risk of developing pressure sores and urinary tract infections but the main risk is overwhelming chest infection, which can develop within 24 hours. It is vital that relatives learn how to look after the patients while they are in hospital and that when the treatment is completed they have short trial periods at home. After discharge it is especially important that they remain well hydrated and in continuous contact with a spinal unit. Should the patient develop either pressure sores or urinary tract infections a spinal specialist should be contacted immediately.—J R SILVER, consultant in spinal injuries, Aylesbury.

A young man's diet is milk, pulses, bread, vegetables, fresh fruit, and a few ounces of cheese a week. He is in good health and able to do hard work, though subject to depressive episodes. Is the diet likely to be lacking in calcium, and if so could the lack produce psychiatric symptoms before physical ones?

Although the diet is rather restricted in composition, it is clearly adequate to support physical activity and overall health. The details of the diet are not sufficient to permit the calculation of intakes completely but the milk alone would contribute over 600 mg a day of calcium so that the diet contains adequate calcium.[1] The foods making up the diet are rich in dietary fibre and the associated phytates reduce the bioavailability of the calcium (and iron and zinc) in the diet, but it would seem unlikely that the calcium supply is inadequate.[2] This could, however, be tested by providing a mineral supplement. The diet does look monotonous and dietetic advice on constructing a more varied and attractive diet would be worth exploring.—D A T SOUTHGATE, head, nutrition and food quality division, Food Research Institute, Norwich.

1 Paul AA, Southgate DAT. *McCance and Widdowson's the composition of foods.* London: HMSO, 1978.
2 Department of Health and Social Security. *Recommended daily amounts of food energy and nutrients for groups of people in the United Kingdom.* (Report on health and social subjects 15.) London: HMSO, 1979.