Statistics in Question

SHEILA M GORE

ASSESSING METHODS— MANY VARIABLES

Seeing a way through many variables is an exciting challenge. By making the best use of latent information a whole range of problems can be solved—screening patients for referral to a hypothyroid clinic¹; predicting the outcome for a patient with severe head injury so that limited resources for intensive care are used constructively²; identifying before operation patients at high risk of developing postoperative deep vein thrombosis³ (for whom prophylactic heparin is justified); explaining variation in cardiovascular mortality⁴ and making diagnoses⁵; ensuring that clinical trial results are not distorted by chance imbalance between randomised treatment groups⁷; or setting up a prognostic staging system that accords with survival.⁶ Doctors who have identified such a problem in their own specialty are advised to consult a statistician about the details of analysis, which may be fairly complicated. The simple underlying principle is to combine some or all of the many explanatory variables as a descriptive, predictive, diagnostic, or prognostic index for the appropriate outcome or dependent variable.

A second class of problems—for example, understanding how a disease presents—leads to the analysis of multi-way contingency tables, and statistical consultation is again advised.

(21) *Identify a common structure for the five problems shown in table I and suggest why the methods of analysis differ in detail but not in principle.*

—in each of the five problems there is a set of predictor or explanatory variables, some or all of which will be combined to explain, predict, or describe the dependent or outcome variable

—methods of analysis differ in detail because of (i) restrictions on scale for the outcome variable (probabilities are constrained to be in the interval [0,1], the risk of dying is non-negative); (ii) choice between an additive or multiplicative model (subject matter and the data usually determine which is more appropriate); (iii) requiring the random error terms in the statistical model to have constant variance to avoid the need for weighted regression

—the principle behind multivariate methods should be illustrated in a table or graph that shows how successful the predictions are in relation to the observed outcome

—ideally the success of any statistical model should be tested on new data

COMMENT

Inferences from any method of analysis are reliable only when the data have been collected in a disciplined way, the series of study patients being in some defensible sense representative. Data from clinical trials generally satisfy this requirement because the entrance criteria define the study population; a series of consecutive patients is acceptable provided that there is no important seasonal variation or other time trend either in the presentation of the disease or its outcome that would preclude making general inferences. Authors should recognise and comment upon the limitations of their sample. They must try to assess how these limitations might promote or eliminate particular variables in an explanatory model.

Many variables seem at first to be a matrix of insufferable complexity, but simple summaries and illustrations can be used to good effect⁹ to give an initial impression of predictor variables and to group together those that are highly interdependent. This is the foundation on which any multivariate statistical model is built. The next step is deciding the form of the regression function and whether the outcome variable needs to be transformed to induce a particular structure of error. Statistical advice is usually necessary. In practice, predictors are most often combined linearly, as in the five examples that follow: when the outcome variable is transformed to control variance linearity often comes as a by-product. The examples are in order of increasing technical difficulty as regards estimating the coefficients in the prediction equation—a statistical problem. A reporting problem is that the transformed scales on which it is convenient and reasonable to perform statistical analyses

are less familiar in examples (3) to (5) than in examples (1) and (2). Authors should communicate their findings by mentioning the methodology briefly but concentrate on the clinical interpretation and applicability of the results in language that is familiar to readers. A good example is the description by Clayton *et al*³ of a predictive index for postoperative deep vein thrombosis. The common thread in the five examples is the combining of several explanatory variables to describe or predict an outcome. Keep that idea firmly in mind and the barrier of transforming the problem for analytic reasons becomes trivial. Also good reporting should emphasise the clinical interpretation of results, and how they should be used—to help decision making or for test reduction.⁶ It is important for doctors to recognise when methods, as described here, might be applied in their specialty and to seek statistical collaboration on what are often exciting and challenging problems—statistically as well as clinically.

(1) Mean heart rate during operation¹⁰ depends on the surgeon's age and seniority, resting heart rate, the type and length of the operation, scrub-up time, medication (such as beta-blockade), and so on. The problem is to combine some or all of these explanatory variables in a prediction equation for mean heart rate during the operation to give a better understanding of their relative importance. The outcomes for individual surgeons will, of course, deviate randomly from the predicted means, and this accounts for the error terms that bedevil statistics. Suppose that mean heart rates from 80 to 140 beats per minute have been observed. Possibly the residual random variation increases with the increase in predicted mean heart rate, but in the first instance constant variance could be assumed. Common sense suggests that a reasonable form for the prediction equation is a baseline mean heart rate β_0 , which is modified up or down by adding weighted terms (the coefficients $\beta_1, \beta_2, \dots, \beta_8$ are the weights) to account for age, seniority, length of operation, and so on. If one or more of the weights is not significantly different from zero then the corresponding explanatory variable might be eliminated from the prediction equation without serious loss of predictive value. Selecting a small number of explanatory variables is a non-trivial statistical problem. (The availability of statistical packages on computers does not lessen the need for statistical thinking. Blunderbuss approaches to statistical modelling are wasteful.) A multiple regression model for predicted mean heart rate during operation combining explanatory variables linearly has the form

$$y_{x_1, x_2, \dots, x_8} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_8 x_8$$

where y denotes predicted mean heart rate.

(2) Pocock *et al*⁴ noted that water hardness and other factors have a proportional effect on cardiovascular mortality. That is to say, if towns A and B are similar in all respects except water hardness (0.2 mmol/l in town A and 1.3 mmol/l in town B, say) then the SMR (standardised mortality ratio for cardiovascular

diseases) for town B is expected to be 8% less than in town A, so that if town A's SMR is 100, town B has SMR = 92, whereas if town A has SMR of 150, town B's SMR will be 138. This is what we mean by a constant *proportional* effect of water hardness on SMR. A multiplicative (proportional) model for SMR is equivalent to an additive model for log SMR and so Pocock *et al* used the logarithm of SMR as the dependent variable in a multiple regression model. The dependent and explanatory variables are different from those described in the multiple regression model above but transforming the outcome variable has induced the same structure, and the method of analysis is the same thereafter. Figure 1 shows actual SMR against predicted SMR, prediction being based on the set of five explanatory variables in table I. The authors noted that some towns, especially in Scotland, had a higher cardiovascular mortality than the model predicted, and this geographical clustering in mortality is being investigated. Departures from a predicted model, as in this case, often suggest new lines of research. The performance of a regression model should always

Actual SMR for cardiovascular diseases plotted against SMR predicted from five-variable model for 234 towns.

be reported—on a familiar scale—so that readers as well as authors appreciate the relevance of the modelling.

(3) A simple prognostic index for predicting before operation which patients will develop postoperative deep vein thrombosis was established by Clayton *et al*³ using clinical and coagulation data obtained before operation on 124 patients, of whom 20 had thrombosis. The structure of this problem is similar to the previous examples but the emphasis—identification of high-risk patients—is different and the technical detail more difficult. In the original 124 cases a binary outcome was observed

TABLE I—Problems with a common structure

	Outcome variable				
	(1)	(2)	(3)	(4)	(5)
	Surgeon's mean heart rate during operation	(Logarithm of) standardised mortality ratio for cardiovascular disease	Postoperative deep vein thrombosis	Tardive dyskinesia: absent, mild, moderate, severe	Risk of dying after myocardial infarction
Explanatory or predictor variables, some or all of which are selected:	Age Seniority Specialty Smoking habit Length of operation Scrub-up time Resting heart rate Beta-blockade	Total water hardness Rainfall Maximum temperature % manual workers Car ownership	Age % overweight for height Preoperative stay Cigarette smoking Varicose veins Malignant disease Fibrinogen Factor VIII as % of normal Euglobulin lysis time Serum FR antigen	Age group Inpatient/outpatient status Dose of antipsychotic drugs (chlorpromazine equivalents) Duration of treatment Anticholinergic medication Parkinsonism Sex	Timolol/placebo Sex Age group Previous infarction Angina Treated hypertension Diuretic treatment before admission Beta-blocker treatment before admission Heart failure Lowest systolic blood pressure < 100 mm Hg Arrhythmias in acute stage Site of infarct Smoking habit

(presence or absence of postoperative deep vein thrombosis) and we want to predict, for future patients, a probability. Directly predicting the probability of postoperative deep vein thrombosis for a given type of patient is troublesome for two reasons: firstly, because variance depends on the estimated probability, and secondly, a predicted probability outside the interval [0,1] would be nonsensical. Both these difficulties are avoided by translating the problem into predicting the log odds which has an unrestricted range—the conversion between the log odds (logit) and probability scales is shown in figure 2. (The

log odds scale is one of several on which proportions can be analysed.) Clayton *et al* called the log odds scale the *predictive index*. The five out of 10 variables that they identified as having the best predictive power gave the following index

$$I = -11.3 + 0.009x_1 + 0.22x_2 + 0.085x_3 + 0.043x_4 + 2.19x_5$$

where x_1 = euglobulin lysis time (minutes), x_2 = concentration of fibrin-related antigen (mg/l), x_3 = age (years), x_4 = percentage overweight for height, and x_5 = presence or absence of varicose veins (scored as 1 or 0 respectively).

High positive values of the predictive index are associated with a high risk of developing postoperative deep vein thrombosis. In a prospective study of 62 new cases,¹¹ using a cut-off point of -2 correctly identified nine out of 10 patients and incorrectly identified seven out of 52 patients as being at risk of developing postoperative deep vein thrombosis (see figure 3). This discrimination was as good as that obtained on the original data. Validating prediction models on the original data, which they were designed to reproduce, gives an overoptimistic view of their performance. Only by testing models on new data is their worth firmly established.

(4) Kidger *et al*¹² described a scoring system for orofacial

Distribution of preoperative predictive indices for the 62 patients studied

- Patients who did not develop deep vein thrombosis
- ▲ Patients who did develop deep vein thrombosis

and neck movements in patients with psychiatric illness. By grading the total score to give close agreement with doctors' rating of tardive dyskinesia each patient was allocated to one of four ordered categories: tardive dyskinesia—absent, mild, moderate, or severe. The next question is whether factors such as age, duration of treatment, dose of antipsychotic drugs (measured in chlorpromazine equivalents), Parkinsonism, anticholinergic medication, inpatient/outpatient status etc, in some combination partially explain severity of tardive dyskinesia. An extension of the logistic model in example (3)—appropriate when the outcome variable is an ordered category—might be the method of choice. Intuitively a graded or ordered outcome is more informative than collapsing scores to give a crude classification: tardive dyskinesia—absent or present.

(5) Patients were randomised after myocardial infarction to timolol or matched placebo.⁷ By chance—assuming no randomisation leak—the placebo group includes a higher proportion of patients aged 65 years or older, patients with arrhythmias in the acute stage, patients with a clinical history of treated hypertension, or patients who had taken diuretics. Of course, these variables are interrelated, older patients being more likely to have arrhythmias in the acute stage and to have been on diuretic treatment before admission, but it will be important to check that a definite effect of treatment persists after allowing retrospectively for moderate imbalance between the groups as randomised. Attention focuses on the time-specific risk of dying or hazard, a convenient measure of the changing force of mortality in time. If we assume a proportional hazards model—a reference point that is not always valid¹³—then we claim that there is a basic form of hazard which is proportionately increased or decreased according to the set of explanatory variables which describes a particular patient.

The constant of proportionality is most often taken to be the *exponential* of a linear combination of the explanatory variables—exponentiation being to ensure that the final estimates are non-negative. Whether the patient was randomised to timolol would be included as one of the explanatory variables and if the treatment effect were still significant after adjustment for risk factors as described above, then the investigators would be reassured that the advantage was probably genuine and not an artefact because of moderate imbalance between the randomised groups.⁷ It is certainly not the case, as Mitchell suggests,¹⁴ that an imbalance in risk factors, significant at the 1% level, necessarily translates into a survival difference, significant also at the 1% level. What the implications for survival are may be sorted out by an analysis of the type described above. Perhaps Professor Mitchell's whimsical account was intended as a rebuke against the failure to report fully a sophisticated analysis?

Regression models for survival can be adapted easily to give a prognostic classification of patients as in chronic lymphocytic leukaemia¹⁵ or breast cancer.⁸

Fairly complex statistical methods have been discussed in relation to the problems in this commentary. Complexity is not always necessary. Comparing statistical techniques in the context of diagnosing hypothyroidism, Gardner and Barker¹ concluded that a simple method—counting the total number of symptoms present—was as effective in determining a rule for

Complexity is not always necessary

Common sense should not be divorced from statistical thinking

referral as more complicated methods. Thomson *et al*¹⁸ modified a predictive index for which patients should be asked to give consent for mastectomy because they wanted to avoid mammography in younger women.

Common sense should not be divorced from statistical thinking.

(22) *Calcium as a measure of water quality was not included in the five-variable regression model for log SMR reported by Pocock et al.⁴ Does this mean that water calcium is irrelevant as an explanatory variable for cardiovascular mortality?*

—no

—excluded variables are superfluous, not irrelevant

COMMENT

The results of Pocock *et al*⁴ do not imply that the variables excluded from the prediction are useless indicators of cardiovascular mortality. The implication is rather that, after taking account of the information already in the prediction equation, the remaining variables are redundant in the sense of not adding anything further to the explanation. A different regression model—in which water calcium replaced total water hardness as a measure of water quality—could have been proposed without serious loss. Regression models give an explanation not the explanation, because selection of predictor variables requires judgment as well as technical skill.

Regression models give an explanation not the explanation: selection of predictor variables requires judgment as well as technical expertise

(23) *There is recent evidence from another clinical trial that treatment A is more successful in male patients. How would you expect the statistician to allow for this in a regression model that takes account of several risk factors?*

—by including an interaction term

COMMENT

The effect of treatment A would be measured by two indicator variables, one pointing to male patients, the other to female patients. If the corresponding estimated coefficients were significantly different the effect of treatment A probably differs between the sexes. Interaction terms are not convincing unless there is a sensible interpretation of them, or prior justification for their inclusion. Doctors should advise the statistician if they have good reason to suspect that a treatment will be more effective in one group of patients than another.

(24) *Tumour size, fixation/ulceration, clinical disease of the homolateral axillary nodes, and presence/absence of other signs (associated with poor prognosis) were recorded for 3695 of the*

3922 patients in the Western General Hospital breast cancer series. Table II—a cross-classification of patients by these clinical features—shows part of the data. How many patients with no other signs were referred with tumour size 5 cm or more, no fixation to the overlying skin, and no nodal disease?

—74

—analysis of multi-way contingency tables leads to general inferences about how disease presents

COMMENT

Analysing the completed four-way contingency table (table II) answers questions such as: Is tumour size larger when there

TABLE II—Cross tabulation by clinical features

	Other signs absent: fixation/ulceration			Other signs present: fixation/ulceration		
	1	2	3	1	2	3
<i>Nodes 1</i>						
Size 1	249	160	4	7	12	1
2	206	531	9	12	103	10
3	74	242	13	16	131	50
<i>Nodes 2</i>						
Size 1	83	69	0			
2	118	305	10			
3	45	174	11			
<i>Nodes 3</i>						
Size 1						
2						
3						
<i>Factor</i>				<i>Coding</i>		
Tumour size	Size			≤2 cm	1	
				3–4 cm	2	
				≥5 cm	3	
Fixation/ulceration				No fixation, no ulceration	1	
				Fixation, no ulceration	2	
				Ulceration	3	
Homolateral axillary nodes				Not affected	1	
				Mobile	2	
				Matted	3	

are other signs present? Does increased tumour size mean more disease in the homolateral axillary nodes? Does the apparent association of tumour size and fixation resolve if we take into account the state of the axillary nodes? Is there an important second-order interaction of fixation/ulceration, nodal disease, and presence/absence of other signs, for example?

These questions are not answered satisfactorily by considering only two factors at a time. A statistician would usually be consulted about the analysis of multi-way contingency tables. The usefulness of this approach is that it leads to general inferences about how a disease such as breast cancer presents.

I am grateful to Dr S J Pocock and colleagues for permission to reproduce figure 1 and to Dr A J Crandon and others for figure 3 under question 21.

References

- Gardner MJ, Barker DJP. Diagnosis of hypothyroidism: a comparison of statistical techniques. *Br Med J* 1975;ii:260-2.
- Teasdale G, Parker L, Murray G, Knill-Jones R, Jennett B. Predicting the outcome of individual patients in the first week after severe head injury. *Acta Neurochir* 1979;suppl 28:161-4.
- Clayton JK, Anderson JA, McNicol GP. Preoperative prediction of postoperative deep vein thrombosis. *Br Med J* 1976;ii:910-2.
- Pocock SJ, Shaper AG, Cook DG, *et al.*, British Regional Heart Study: geographic variations in cardiovascular mortality, and the role of water quality. *Br Med J* 1980;280:1243-9.
- Bouckaert A. Computer diagnosis of goitres. III Optimal subsymptomatology. *J Chron Dis* 1971;24:321-7.
- Card WI, Emerson PA. Test reduction. I Introduction and review of published work. *Br Med J* 1980;281:543-5.
- The Norwegian Multicenter Study Group. Timolol-induced reduction in

- mortality and reinfarction in patients surviving acute myocardial infarction. *N Engl J Med* 1981;**304**:801-7.
- ⁸ Myers MH, Axtell LM, Zelen M. The use of prognostic factors in predicting survival for breast cancer patients. *J Chron Dis* 1966;**19**: 923-33.
- ⁹ Gore SM. Assessing methods—a feel for other things. *Br Med J* 1981; **283**:775-7.
- ¹⁰ Foster GE, Evans DF, Hardcastle JD. Heart-rates of surgeons during operations and other clinical activities and their modification by oxprenolol. *Lancet* 1978;ii:1323-5.
- ¹¹ Crandon AJ, Peel KR, Anderson JA, Thompson V, McNicol GP. Postoperative deep vein thrombosis: identifying high-risk patients. *Br Med J* 1980;**281**:343-4.
- ¹² Kidger T, Barnes TRE, Trauer T, Taylor PJ. Sub-syndromes of tardive dyskinesia. *Psychol Med* 1980;**10**:513-20.

- ¹³ Gore SM. Assessing methods—survival. *Br Med J* 1981;**283**:840-3.
- ¹⁴ Mitchell JRA. Timolol after myocardial infarction: an answer or a new set of questions? *Br Med J* 1981;**282**:1565-70.
- ¹⁵ Montserrat E, Rozman C. Subclassification of stage II chronic lymphocytic leukaemia with prognostic and therapeutic implications. *Lancet* 1979; ii:854.
- ¹⁶ Thomson HJ, Miller SS, Gore SM, Bayliss A. Consent for mastectomy. *Br Med J* 1980;**281**:1097-8.

Sheila M Gore, MA, is a statistician in the MRC Biostatistics Unit, Medical Research Council Centre, Hills Road, Cambridge CB2 2QH.

No reprints will be available from the author.

Lesson of the Week

Opiate toxicity in elderly patients

T H CARADOC-DAVIES

Opiate drugs are used to relieve pain and dyspnoea in patients with myocardial infarction and left ventricular failure. The usual doses of diamorphine are 5 mg intravenously or intramuscularly,¹ and morphine 10 mg intravenously or intramuscularly, both repeated if necessary. This gives relief to most patients. An elderly woman was recently admitted with myocardial infarction and left ventricular failure. She was unconscious and cyanosed with pinpoint pupils. She had received a "standard" dose of opiate, and her striking response to treatment with naloxone prompted me to collect similar cases.

Case reports

Over three months there were about 450 acute admissions to the geriatric unit of this hospital. All of the patients were aged over 75 years. Seven cases of opiate toxicity were detected: three admitted by general practitioners, three from a casualty department, and one in a ward bed. The criteria were the triad of pinpoint pupils, low level of consciousness, and depressed respiration (all were deeply cyanosed); response to naloxone; and confirmation that an opiate had been given. The following features were recorded after a clinical examination: level of consciousness, simple neurological examination, pulse rate, blood pressure, colour and respiration, and signs of left ventricular failure. Arterial blood gases were not measured owing to the urgency of giving naloxone to the patients. Electrocardiograms were taken and the concentration of cardiac enzymes measured to establish the presence or absence of myocardial infarction. A chest x-ray examination confirmed the clinical impression of left ventricular failure. Information about the dose and the route and time of administration of the narcotic and other drugs was established.

The data from the seven patients were analysed. (a) Age range

Elderly patients may suffer from opiate toxicity unless they are given less than the normal adult dose intravenously

75 to 90 years. (b) Indications for treatment with an opiate: pain only in three patients, dyspnoea in two, and both in two. (c) Blood pressure was low in three, compared with previous or subsequent readings. In one patient (case 5) it rose after treatment with naloxone. (d) Other drugs: patients with left ventricular failure had all received frusemide, one had received aminophyllin. Two patients (cases 1 and 4) had been given perphenazine prophylactically with heroin. Those who had received Cyclimorph also received cyclizine, which it contains. (e) The time from when the opiate was given to reversal ranged from 15 minutes to seven hours. (f) Outcome: five patients had sustained myocardial infarction, and three of these died (cases 2, 4, 6 of persistent hypotension, sudden death, and in left ventricular failure, respectively). (h) Chest pain and severe dyspnoea did not recur after reversal with naloxone in any patient.

Case 6, who had received 20 mg of heroin, required 0.8 mg of naloxone before responding and needed a further 0.4 mg after six hours, suggesting that diamorphine has a longer half-life than naloxone in elderly patients.

Discussion

The mode of action of the opiates in patients acutely ill with myocardial infarction or left ventricular failure is not clearly understood. The results from different studies are conflicting.²⁻⁶ Opiates certainly act centrally in reducing dyspnoea, pain, and sympathetic drive, which may be vitally important in reducing cardiac output and lessening the chance of life-threatening arrhythmias occurring. There may be peripheral actions on resistance and capacitance vessels.

The common side effects may be grouped as central: nausea, vomiting, and respiratory depression; and peripheral: hypo-

St Helen's Hospital, Hastings

T H CARADOC-DAVIES, MB, senior registrar in geriatric medicine (present address: Wakari Hospital, Dunedin, New Zealand)