Improving the selection of patients for upper gastrointestinal endoscopy

S A Naji, P W Brunt, S Hagen, N A G Mowat, I T Russell, T S Sinclair, T M H Tang

Abstract

A retrospective study was undertaken to investigate how endoscopies yielding positive findings differ a priori from those yielding negative findings: and how those judged 'helpful' (in the sense of influencing management) differ a priori from those judged 'unhelpful'. A total of 483 patients undergoing endoscopy was sampled and a wide range of data abstracted, including 48 patient characteristics available to the gastroenterologist at the time of the decision to perform endoscopy. Sixty nine per cent of endoscopies were positive. Multivariate statistical analysis identified four variables which taken together were strongly predictive of a positive endoscopy. The resulting mathematical formula correctly predicted the outcome of 76% of endoscopies. Eighty two per cent of the endoscopies were retrospectively classified by the gastroenterologists as helpful. Six variables were strongly predictive of a helpful endoscopy. The corresponding formula correctly predicted the findings of 84% of endoscopies. Comparison of the two analyses shows that the two sets of predictions differ substantially. Thus it is important that decision tools should be based not on the crude distinction between positive and negative, but on the more useful distinction between helpful and unhelpful in influencing management.

(Gut 1993; 34: 187-191)

Upper gastrointestinal endoscopy has been shown to yield higher diagnostic accuracy, sensitivity, and specificity than radiological techniques.1-4 Gastroenterology units have therefore been established throughout the United Kingdom, most offering an endoscopy service limited to patients screened initially at hospital, although some offer an open access service to general practitioners.5-10 The results has been a large increase in the number of endoscopies performed11 and, in some studies, a greater proportion of patients has been found to be normal or to have minor or insignificant disease.11 Most studies have reported that as many as 30-40% of patients undergoing endoscopy are found to be normal.12-14 It has also been suggested that the increased use of endoscopy does not influence the diagnosis or prognosis of gastric cancer15 or the major complication rate of peptic ulcer disease.16 Some authors have taken these studies to indicate that despite the diagnostic superiority of endoscopy over radiological techniques, there are no concomitant short term benefits in patient management.1 Others have concluded that there is a need to improve the use of endoscopy through better selection and targeting of patients.11 12 16

We undertook a retrospective study of elective upper gastrointestinal endoscopies performed in a single unit. Our primary aims were to investigate (1) whether patients undergoing endoscopies which show significant pathology differ a priori from patients in whom significant pathology is not found, and (2) whether endoscopies adjudged to have influenced patient management differ a priori from those adjudged not to have influenced patient management. Our secondary aim was to assess the potential of a particular multivariate statistical method (stepwise logistic discriminant analysis) to improve the selection of patients for endoscopy.

Methods

PATIENTS

All 485 patients who underwent elective upper gastrointestinal endoscopy (excluding endoscopic retrograde cholangiopancreatoigraphy) at the Grampian Health Board Gastrointestinal Unit (then based at Woodend General Hospital, Aberdeen) during February, June and October 1987 (to avoid any seasonal variations) were considered for the study. The patients were retrospectively identified from the theatre books and further data were collected by one of the authors (TMHT) from routine sources, including the endoscopy reports filed in the Gastrointestinal Unit, the files of the radiology and pathology departments and, where necessary, the patients' medical records. The data extracted from these sources included patient characteristics such as age, sex, source of referral, recorded symptoms, previous investigations, endoscopy findings, and management plans. Adequate data were available on all but two patients, yielding a final sample of 483 patients, comprising 235 (48.6%) men and 248 (51.4%) women. Ages ranged from 14 to 91 years, with a mean (SD) of 55.2 years (18.5). All patients had been referred for endoscopy from hospital wards or outpatient clinics in Aberdeen, with the gastrointestinal outpatient clinic and inpatient ward accounting for 64% of the total.

ENDOSCOPY

Most endoscopic examinations were performed by one of three consultant gastroenterologists, with a smaller proportion performed by the gastroenterology senior registrar or registrar. All patients were fasted overnight, had Xylocaine local anaesthetic spray to the oropharynx, and were sedated with intravenous
diazepam. Examinations were undertaken with standard forward viewing fibreoptic endoscopes.

CRITERIA FOR SUCCESS OF ENDOSCOPY
Two separate classifications of the 483 endoscopies were undertaken:

Positive or negative
On the basis of the endoscopy reports, the endoscopies were divided into two groups:
(a) Those that showed any pathology other than hiatus hernia alone were classified as 'positive';
(b) Those that showed no pathology other than hiatus hernia alone were classified as 'negative'.

Helpful or unhelpful
The medical records of all 483 patients were reviewed by each consultant (PWB, NAGM, TSS) to validate the positive or negative classification of each endoscopy and to decide whether it had been a helpful or unhelpful examination. In cases where opinion differed, the classification was agreed after discussion of all the factors between the three reviewers. Accordingly, the endoscopies were divided into two groups: (a) those which were adjudged to have influenced patient management were classified as 'helpful'; (b) those which were adjudged not to have influenced patient management were classified as 'unhelpful'.

STATISTICAL ANALYSIS
Stepwise logistic discriminant analysis was used to identify which of the 48 abstracted variables available to the gastroenterologist at the time of the decision to perform an endoscopy were predictive of the findings of that endoscopy in terms of both criteria - positive or negative and helpful or unhelpful. This technique selects in a stepwise fashion that subset of variables that best discriminates between patients in the two groups of interest and results in a mathematical equation (known as a discriminant function) that can be used to predict the group to which any future patient will belong.

The stepwise selection of variables for this discriminant function was terminated when none of the remaining variables was able to improve the 'maximum log likelihood' (a measure of the predictive power of the discriminant function) by an amount that would have been statistically significant, nominally at the 1% level recommended by Russell and Gregson. This strict significance level was chosen to allow for the multiplicity of statistical tests implied by the adoption of a stepwise procedure.

Results
Table I shows that the proportion of endoscopies classified as negative (31%) was similar to that in previous studies. However, most (67%) of the negative endoscopies were considered to have been helpful, and it was only 79 (16%) of all endoscopies that did not influence patient management. A significant proportion (37%) of these unhelpful endoscopies were classified as positive examinations.

Positive or negative
Stepwise logistic discriminant analysis identified four of the 48 variables known at the time of the endoscopy which, taken together, were strongly predictive of the findings (positive or negative) of the endoscopy. The first variable to enter the discriminant function was whether the examination was a review of a previous endoscopy; if so, a positive result was significantly more likely. From the remaining 47 variables, the patient's age (whether more or less than the sample mean of 55) made the next greatest improvement in predictive power; older patients were significantly more likely to generate a positive endoscopy. The third predictive variable selected was whether the indication for endoscopy was other than the most common ones of anaemia, anorexia, bleeding, 'dyspepsia', dysphagia, heartburn, pain, vomiting, and weight loss. Patients with a less common indication were significantly less likely to generate a positive examination. The fourth and final predictive variable to enter the discriminant function was whether diffuse epigastric pain was recorded as an indication for endoscopy: the absence of recorded pain was significantly less likely to lead to a positive endoscopy. None of the remaining 44 variables improved the predictive power of the function enough to justify inclusion. The discriminant function derived from the four selected variables correctly predicted the findings (positive or negative) of 76% of the endoscopies studied. Table II shows this discriminant function converted into a simplified scoring system suitable for routine clinical use. Figure I shows the group distribution of scores for the 483 endoscopies
Improving the selection of patients for upper gastrointestinal endoscopy

**Figure 1:** Prediction of positive endoscopy: grouped distribution of scores from logistic discriminant analysis. Note: the figures in parentheses indicate the predicted probability of a helpful endoscopy in this range of scores.

Using this scoring system; this shows a very close relationship between predictive scores and actual findings.

HELPFUL OR UNHELPFUL

The same 48 variables were subjected to a second analysis to discriminate between helpful and unhelpful endoscopies. Six variables were identified which taken together were strongly predictive of the findings (helpful or unhelpful) of an endoscopy. The first variable to enter the discriminant function was whether diffuse epigastric pain was recorded as an indication for endoscopy; recorded pain predicted an unhelpful endoscopy. The next predictive variable selected was source of referral for endoscopy; referrals from the gastroenterology ward were significantly more likely to result in helpful endoscopies. The third variable predicting helpful findings was a recent abnormal barium investigation. The presence of heartburn as an indication for endoscopy entered as the fourth variable; this was predictive of an unhelpful endoscopy. Whether the examination was a review of a previous endoscopy and whether the endoscopy was performed by a particular gastroenterologist became the last two variables in the discriminant function, both predictive of a helpful examination. None of the remaining 42 variables could significantly improve the predictive power of the discriminant function. The discriminant function based on the six selected variables correctly predicted the findings (helpful or unhelpful) of 84% of the endoscopies studied. Table III shows this discriminant function converted into a simplified scoring system suitable for routine clinical use. Figure 2 shows the grouped distribution of scores for the 483 endoscopies using this scoring system; again there is a very close relationship between predictive scores and actual findings.

Thus there were two variables common to both discriminant functions; whether the endoscopy was a review and whether diffuse pain was recorded as an indication. Since stepwise logistic discriminant analysis is designed to yield robust prediction rather than a robust selection of predictive variables, however, it was still possible that these two discriminant functions were essentially equivalent. We therefore tested this hypothesis using the method of Gregson. A highly significant test statistic was obtained, indicating that the two discriminant functions were very different; in other words, the factors that discriminate between positive and negative endoscopies differ fundamentally from those that discriminate between helpful and unhelpful endoscopies.

**Table III**

<table>
<thead>
<tr>
<th>To predict the findings (helpful or unhelpful) of endoscopy in a known patient.</th>
<th>Possible</th>
<th>Scored</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 Begins with a score of:</td>
<td>14</td>
<td>14</td>
</tr>
<tr>
<td>1 If patient was referred from gastroenterology ward</td>
<td>Add 16</td>
<td></td>
</tr>
<tr>
<td>2 If recent barium meal was abnormal</td>
<td>Add 9</td>
<td></td>
</tr>
<tr>
<td>3 If endoscopy is a review</td>
<td>Add 14</td>
<td></td>
</tr>
<tr>
<td>4 If endoscopy to be performed by Dr A</td>
<td>Add 12</td>
<td></td>
</tr>
<tr>
<td>5 If patient has diffuse epigastric pain</td>
<td>Subtract 11</td>
<td></td>
</tr>
<tr>
<td>6 If patient has heartburn</td>
<td>Subtract 17</td>
<td></td>
</tr>
</tbody>
</table>

Total z score =

Key:
The patient’s z score will be between -14 and +65. A score of zero indicates a 50% chance of a helpful endoscopy. A negative score indicates less than 50% chance of a helpful endoscopy. A positive score indicates greater than 50% chance of a helpful endoscopy.

More specifically, the chance of helpful endoscopy (p) is given by: p = e^(z/2) / (1 + e^z) where e = 2.718, the base of natural logarithms.

**Discussion**

This retrospective study of 483 essentially random upper gastrointestinal endoscopies has shown that almost one third did not show significant pathology, a proportion similar to those found in previous studies. It has been suggested that, although "the value of negative endoscopic findings cannot be assessed", "there is little objective evidence of benefit". Our study calls both of these propositions into question. The value of a negative endoscopy (or any other diagnostic procedure) can and should be assessed in the same terms as a positive endoscopy, i.e by
determining its effect upon the subsequent management of the patient.

In this study, 67% of the negative endoscopies were adjudged to have influenced patient management. Consequently, the 16% of endoscopies (both positive and negative) that were not thought to have contributed to patient management represent a much lower proportion than suggested in previous studies. The value of a negative endoscopy must not be underestimated; not only may it rule out serious disease and thus reassure both patient and doctor, it may also result in less radical treatment with concomitant benefits for patients and for NHS costs. Furthermore, the comparison of our two discriminant functions (between positive and negative endoscopies and between helpful and unhelpful endoscopies) showed that the two sets of predictions differed substantially and were not derived from the same factors. Accordingly, decisions about the most appropriate use of diagnostic endoscopy (whether routine clinical decisions about individual patients, or decisions about the use of resources) should be based not upon the crude and potentially misleading distinction between positive and negative (or abnormal and normal), but upon the clinically more relevant distinction between endoscopies which influence management and those which do not.

Although stepwise discriminant analysis is a statistical technique that yields predictions that are both accurate and simple to calculate, it suffers from one major limitation—the resulting decision rules (for example, Tables II and III) cannot be interpreted as explaining why some endoscopies are helpful while others are not. The basic problem is that among the many potential predictive variables, there are almost certain to be many that are highly correlated with each other; it is then a matter of chance which of two or more such variables is selected for the decision rule. For example, the rules displayed in Tables II and III both imply that an endoscopy was more likely to be helpful if diffuse epigastric pain was not recorded as an indication. Nevertheless, it would be wrong to conclude that a genuine absence of pain predicts a helpful endoscopy. An alternative, and perhaps more likely, explanation is that pain is usually recorded as an indication only in the absence of more objective indications. If so, diffuse epigastric pain appears in the decision rule, not as a variable in its own right, but as a proxy for those objective indications with which it is highly correlated. Similarly, the finding that whether or not the endoscopy was performed by Dr A is predictive is likely to be mediated by highly correlated objective factors, such as referral patterns and case mix. Fortunately, this limitation does not compromise the accuracy of the predictions arising from the decision rule—the justification for using a multivariate statistical technique. What must be borne in mind, however, is that although this method of deriving predictive formulae is generalisable to any context, the individual variables which are shown to have predictive power in one context will not necessarily be predictive in another. Robust predictive models should be context sensitive.

This leads to three main conclusions about the selection of patients for upper gastrointestinal endoscopy. First, this study has shown that decision rules for selecting patients can be derived from multivariate statistical techniques; they yield accurate predictions and can be made simple to use. Secondly, such rules should be based, not on the traditional distinction between positive and negative tests, but on the more pragmatic distinction between helpful tests that contribute to patient management and unhelpful tests that do not. Thirdly, although retrospectively collected data may be useful in showing the feasibility of such a decision rule (as in this study), the construction of rules for routine clinical use should be based on locally relevant data that are prospectively defined and collected; this recommendation will improve the effectiveness of the resulting decision rules, and may also make them easier to interpret.

In the current climate of limited resources for health care, and ever increasing demands upon gastrointestinal services, it is essential to analyse the true value of carrying out diagnostic procedures. The fact that upper gastrointestinal endoscopy is, with the exception of certain specific indications like dysphagia, replacing barium meal as the first investigation in many centres highlights the importance of systematic evaluation.

Our findings suggest how the selection of patients for endoscopy can be improved. Furthermore, they provide the basis for prospective studies which can lead to better use of resources in future.

We wish to thank Mr Foud Ziade for his valuable technical assistance and Kathleen McIntosh for the preparation of this manuscript.

This research was funded by the Chief Scientist Office of the Scottish Home and Health Department. However, the content of this report is the responsibility of the authors, not the Scottish Home and Health Department.

SAN, SH, ITR and TMHT were supported by the Scottish Home and Health Department. PWB, NAGM and TSS were supported by Grampian Health Board.

Improving the selection of patients for upper gastrointestinal endoscopy