J Clin Epidemiol Vol. 44, No. 11, pp. 1263–1270, 1991
Printed in Great Britain.

0895-4356/91 $3.00 + 0.00

USING ORDINAL LOGISTIC REGRESSION TO ESTIMATE THE LIKELIHOOD OF COLORECTAL NEOPLASIA

SCOTT R. BRAZER,* FRANK S. PANCOTTO, THOMAS T. LONG III, FRANK E. HARRELL JR, KERRY L. LEE, MALCOLM P. TYOR and DAVID B. PRYOR

Department of Medicine, Cabarrus Memorial Hospital, Concord, North Carolina and Department of Medicine and Community of Family Medicine, Duke University Medical Center, Durham, NC 27710, U.S.A.

(Received in revised form 26 February 1991)

Abstract—The utility of ordinal logistic regression in the prediction of colorectal neoplasia was demonstrated in a group of 461 consecutive patients undergoing colonoscopy in a community practice. One hundred twenty-nine patients had adenomatous polyps and 34 had colorectal adenocarcinoma. An ordinal logistic regression model developed in a random subset (292 patients) identified five predictors of colorectal neoplasia. Colorectal neoplasia risk could be predicted using the patient's age, sex, hematocrit, fecal occult blood test result and indication for colonoscopy. The risk of colorectal neoplasia in the remaining subset of patients (169) could be reliably estimated from the model. Ordinal logistic regression analysis in this select group of patients can accurately estimate the likelihood of colorectal neoplasia. Because the generalizability of our findings are unknown, the model should not be applied to other patients. However, application of this technique to an unselected group of patients not already referred for colonoscopy could provide unbiased estimates of colorectal neoplasia risk in individual patients.

Predictive models Ordinal logistic regression

INTRODUCTION

As the cost of health care soars, clinicians are under increasing pressure to improve the accuracy of their predictions of patient risk [1]. Diagnostic predictions can be improved by applying regression models to already available information using regression models. Ordinal logistic regression is a valuable technique which: (1) uses all the information available in an ordinal dependent variable [2], (2) makes few assumptions [3, 4], and (3) can use both continuous and nominal predictor variables. The predictive accuracy of the model can be verified [4]. Regression models can yield predictions which are superior to those of expert clinicians in academia and private practice [5, 6].

It is estimated that 155,000 persons developed colorectal cancer in the U.S. during 1990 [7]. The overall mortality, estimated at over 60,000 for 1990, has not decreased in 25 years [8] in part because methods have not been developed that will accurately identify patients at a time when they have localized disease. At the present time, patients' risk for colorectal neoplasia is deemed "average" or "standard" if they are over 50 and have no other risk factors or "high" if they have one or more risk factors (a history of colorectal neoplasia, positive family history of colorectal neoplasia, inflammatory bowel disease (IBD), etc.) [9, 10]. This scheme assigns equal weight to

*All correspondence should be addressed to: Scott R. Brazer, MD, MHS, P.O. Box 3662, Duke University Medical Center, Durham, NC 27710, U.S.A.

all risk factors and disregards the gradation of risk that may occur with individual risk factors such as increasing age [11, 12] and in patients with more than one risk factor. A more precise estimation of the risk for cancer of the colon may facilitate diagnosis [13] and identify patients likely to benefit from more exhaustive investigation. The purpose of this investigation was to examine the utility of ordinal logistic regression in the prediction of colorectal neoplasia in a selected subset of consecutive patients undergoing colonoscopy.

MATERIALS AND METHODS

Patient population and data collection

Data were prospectively collected on 461 consecutive patients undergoing initial colonoscopy in a community hospital during the 13-month study. The procedures were performed and data collected by one of two board certified gastroenterologists (FSP or TTL).

Historical items recorded included: gastrointestinal bleeding (overt or occult); bowel habits; abdominal pain; family history of colorectal neoplasia and a prior history of either colorectal neoplasia, IBD, breast or genital carcinoma, cholecystectomy, or abdominal radiation therapy. In addition to the above, a specific "reasons for colonoscopy" was identified by the consulting gastroenterologist. "Reasons" included: gastrointestinal bleeding (overt or occult); abnormal barium enema; prior history of colorectal neoplasia; abnormal proctoscopic examination suggesting colorectal neoplasia or IBD; unexplained diarrhea or abdominal pain; iron deficiency; positive family history of colorectal neoplasia; and IBD. Items recorded from the physical examination were limited to the presence of significant anal disease (active fissures, significant hemorrhoidal disease, or prominent, inflamed skin tags). Laboratory data included the hematocrit and, if available, the MCV, serum iron, TIBC, and ferritin. The extent of the colon examined was taken from the endoscopy report. Tissue pathology was taken from the pathologists' report. Colorectal neoplasia was defined as colorectal carcinoma or adenoma.

Data were stored on an IBM PC/XT using dBase III (Ashton-Tate, Inc., Torrance, CA) and transmitted to Triangle Universities Computation Center for statistical analysis using SAS (SAS Institute, Inc., Cary, NC).

Methods of analysis

The patient sample was divided by random sampling without replacement into two groups while stratifying on age, sex, and presence of colorectal neoplasia. The training sample comprised 63% (292 patients) of the total sample and was used to identify characteristics associated with colorectal neoplasia. The test sample, used to validate the model, included the remaining 37% (169) of patients. Clinical characteristics independently associated with colorectal neoplasia were identified by ordinal logistic regression [14, 15] in the training sample.

The ordinal logistic model assumes proportional odds; the odds ratio for a given predictor variable does not vary when going from one level to the next for different levels of the outcome [3]. The predictor variables are assumed to be linearly and additively related to the logit [4]. The parameters of the logistic model are estimated by maximum likelihood.

The ordinal scale for colorectal neoplasia was graded from 0 to 3; where 0 was no neoplasia, 1 was an adenomatous polyp of less than 5 mm, 2 was an adenomatous polyp of greater than or equal to 5 mm, and 3 was cancer. Ten candidate variables were examined. These included: age, sex, hematocrit, iron deficiency as a reason for colonoscopy, abdominal pain, constipation, the fecal occult blood test (FOBT), hematochezia, diarrhea or pain as sole indication for the colonoscopy and a presence of a known risk factor. The risk factor was scored 0 (no risk factors) or 1 (one or more risk factors). Risk factors included: a personal history of an adenomatous polyp or colorectal cancer, female genital tract or breast cancer, radiation therapy, IBD, cholecystectomy, or positive family history of colorectal cancer. The number of candidate variables were limited to reduce the risk of "overfitting" caused by spurious associations [16]. "Overfitting" can be demonstrated when the ratio of patients with the outcome of interest to candidate variables falls below 10. Restricted cubic spline functions of age and hematocrit were fitted to examine linearity.

A stepwise variable selection was used. Candidate variables were included if the χ_1^2 measuring independent association with the outcome was greater than 3.84 ($p < 0.05$). All two-way interactions were examined with a global test using a score statistic.

The proportional odds assumption of the logistic regression model was examined in three

ways. First, the cumulative logit for each outcome category was plotted vs quintiles of age. Second, plots of log odds ratios with 95% confidence limits for each variables in the model were plotted vs the point of dichotomization (i.e. 0 vs 1, 2, 3; 0, 1 vs 2, 3; and 0, 1, 2 vs 3). Finally, a partial proportional odds model was fitted to test the proportional odds assumption [3].

Two different validation methods were used to assess the predictive accuracy of the model developed in the training sample when applied to the test sample. In the first method, patients were divided into two age and two sex groups. The predicted probability of outcome 1 or worse (any colorectal neoplasia), outcome 2 or worse (adenomatous polyp $\geqslant 5$ mm or colorectal cancer), and outcome 3 (colorectal cancer) was calculated for each patient and compared with the observed outcome status: patient in outcome group 1 or worse, 2 or worse, and 3 [4].

In the second method, the reliability relating predicted probabilities to observed outcomes is estimated by logistic-linear calibration. In this method, a reliability curve is estimated by fitting a second logistic model, with the independent variable being the predicted log odds of the outcome (formulated from the training sample) and the dependent variable being a dichotomous indicator of whether or not the outcome actually occurred (from the test sample). If the predictions are reliable, the reliability curve generated should be a 45° line. Tests for overall unreliability are reported.

Somer's rank correlation [17] between the linear combination of factors in the logistic regression model and the outcome level was computed in the training and test sample to quantify predictive ability. A nomogram was generated from the training sample to allow predictions to be made in individual patients in the test sample [18].

RESULTS

Endoscopy of the entire colon was accomplished in 93% of the 461 patients; 63% were outpatients. Nearly two-thirds of the 461 patients were female and over half were over 60 years of age (median = 61; interquartile range 48–71). Common presenting symptoms included: abdominal pain (46%), hematochezia (37%), constipation (23%), and diarrhea (17%). A history of IBD was given by 20 patients. A history of colorectal neoplasia was given by 42 patients (9%); 5% had a previous diagnosis of genital or breast cancer and 5 patients had radiation therapy in the past. A remote history of peptic ulcer disease was given by 12% of patients, a history of cholecystectomy in 10% and a family history of colon cancer in 10%. Stools were guaiac positive in 34% of patients; 114 patients did not have a FOBT. The median hematocrit was 40% (interquartile range 36–44%). Eight patients had missing hematocrit values. These 8 patients were not included in the logistic regression analysis.

Most patients underwent colonoscopy to evaluate occult or overt bleeding (38%). Other recorded reasons included: abnormal barium enema (32%), abnormal proctoscopy (14%), abdominal pain (9%), iron deficiency (6%), history of colorectal neoplasia (5%), and IBD (3%) (more than one reason possible per patient).

There were 4 complications including 3 post-polypectomy bleeding episodes requiring transfusion only and 1 cecal perforation requiring surgery. There were no deaths.

Neoplastic (adenoma or carcinoma) lesions were confirmed by histopathology in 163 patients. One hundred twenty-nine had adenomatous polyps (87% $\geqslant 0.5$ cm) and 34 had cancer. An additional 4 patients with colon tumors were identified. Two patients had metastatic lesions to the colon and 2 had colonic lymphoma. These patients were not considered to have colorectal neoplasia for purposes of analysis. Of the 34 colorectal cancers, 27 (79%) were localized to the bowel wall. Six had spread to regional lymph nodes (Dukes' Stage C) and only one patient had distant metastases at presentation.

Plots of the cumulative logit of lesion vs age demonstrated near linearity. A score test of linearity (vs a restricted cubic spline) was insignificant (χ^2_3 4.79, $p = 0.2$). Conversely, the restricted cubic spline of hematocrit demonstrated significant deviation from linearity (χ^2_3 8.35, $p = 0.04$). A plot of the restricted cubic spline function suggested a quadratic relationship. The model using the quadratic term was superior to the model using the restricted cubic spline (χ^2_2 6.77, $p = 0.03$ vs χ^2_4 9.78, $p = 0.04$).

There were no significant interactions between the dummy variable coding for a missing FOBT and the 5 predictors ultimately chosen for the model (χ^2_5 3.89, $p = 0.57$). A plot of the cumulative logit of lesion vs FOBT negative, missing and positive showed the

Table 1. Univariate analysis of clinical characteristics associated with colorectal neoplasia

	Chi-square	p	Association with neoplasia
Advancing age	31.88	<0.0001	+
Diarrhea or pain only	29.60	<0.0001	−
Positive FOBT	12.14	<0.001	+
Male sex	11.86	<0.001	+
Abdominal pain (not sole indication for colonoscopy)	5.17	0.023	−
Hematocrit	2.88	0.089	+
Positive risk factor	0.84	0.36	
Iron deficiency	0.41	0.52	
Hematochezia	0.05	0.83	
Constipation	0.02	0.88	

missing values between negative and positive. Therefore, missing FOBT were set to the mean.

Clinical characteristics associated with colorectal neoplasia as determined by univariate analysis are shown in Table 1.

Independent, statistically important characteristics which correlated with colorectal neoplasia are shown in Table 2. These characteristics included not having diarrhea or pain as the sole indication for evaluation (χ^2_1 42.23, $p < 0.0001$), advancing age χ^2_1 33.49, $p < 0.0001$), the quadratic hematocrit term (χ^2_1 15.95, $p < 0.0001$), male sex (χ^2_1 6.65, $p < 0.01$), a positive FOBT (χ^2_1 5.12, $p = 0.02$), and hematocrit (χ^2_1 3.27, $p = 0.07$). Iron deficiency, hematochezia, a positive history risk factor score, constipation, and abdominal pain did not correlate with colorectal neoplasia after adjustment for the variables included in the model ($p > 0.05$). A global test of all two way interactions was not signifiant (χ^2_{39} 31.87, $p = 0.8$).

The model does not violate the proportional odds assumption. The cumulative logit plot of lesion vs age reveals three nearly parallel lines. Plots of the log odds ratios for the FOBT, sex, age, hematocrit and hematocrit × hematocrit show no clear deviation from a line with a slope of zero. Finally, a global test of proportional odds suggested no deviation from proportionality (χ^2_{10} 9.57, $p = 0.48$).

The predicted probability of disease is very similar to the actual prevalence when patients were divided into four age and sex groups [Figs 1(A–D)]. Figures 2(A) and (B) demonstrate the reliability of predictions for any colorectal neoplasia and for large polyps ($\geqslant 5$ mm) or cancer respectively in the test sample. An overall test for unreliability was insignificant both for any colorectal neoplasia (χ^2_2 2.35, $p = 0.3$) and for large polyps or cancer (χ^2_2 0.45, $p = 0.8$).

The Somer's correlation was 0.475 in the training sample and 0.463 in the test sample, indicating that slight overfitting to the training sample caused slight inability of the model to validate in test sample.

DISCUSSION

This study demonstrates the ability of ordinal logistic regression analysis to provide accurate predictions of risk for colorectal neoplasia in selected patients undergoing colonoscopy. The logistic regression model makes few assumptions which should be verified and tested. The five variables included in the model meet the linearity, additivity and proportional odds assumptions. These five characteristics, available from a standard clinical exam, include the patient's age, sex, FOBT results, hematocrit, and indication for colonoscopy. The study

Table 2. Multivariable analysis of clinical characteristics associated with colorectal neoplasia

	Beta	Chi-square	p
Diarrhea or pain only	−9.56	42.23	<0.0001
Advancing age	0.0474	33.49	<0.0001
Hematocrit × hematocrit	0.0035	15.95	<0.0001
Male sex	0.441	6.65	<0.01
Positive FOBT	0.5798	5.12	0.02
Hematocrit	−0.199	3.27	0.07

Not associated: history risk factor score, iron deficiency, hematochezia, constipation, or abdominal pain ($p > 0.05$).

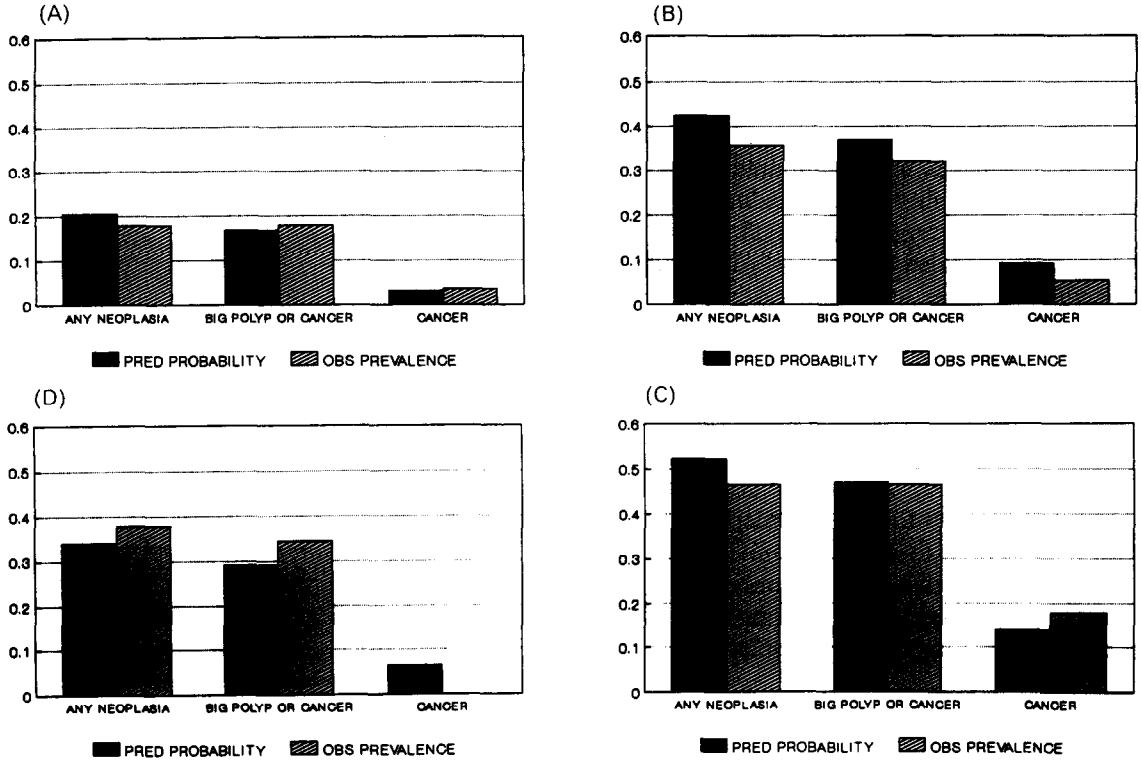

Fig. 1. Predictions in test population. Clockwise from upper left: (A) represents women 16–62 years, (B) women 63–93 years, (C) men aged 24–58 years, and (D) men 59–89 years. Average predicted probabilities of colorectal neoplasia, colorectal adenomatous polyps ⩾5 mm or cancer, and colorectal cancer compared with the observed prevalence. Sex-specific median age was used to derive age groupings.

demonstrates that the information contained in each characteristic contributes independent information to the other characteristics. Risk estimates may be easily generated from a nomogram (Fig. 3). [N.B. The nomogram is provided to demonstrate the feasibility of such estimates and should not be used to develop risk estimates in individual patients.] Estimates based on the model developed in the training sample are reliable when applied to patients in the test sample.

The results of our study demonstrate that risk estimates of colorectal neoplasia in a randomly selected test sample closely correspond to that predicted. Because estimates were only developed and validated in patients who had been referred to a gastroenterologist for evaluation and ultimately underwent colonoscopy, it is possible that risk estimates in patients not referred for colonoscopy might be substantially different. Several types of bias could be introduced through the referral process for

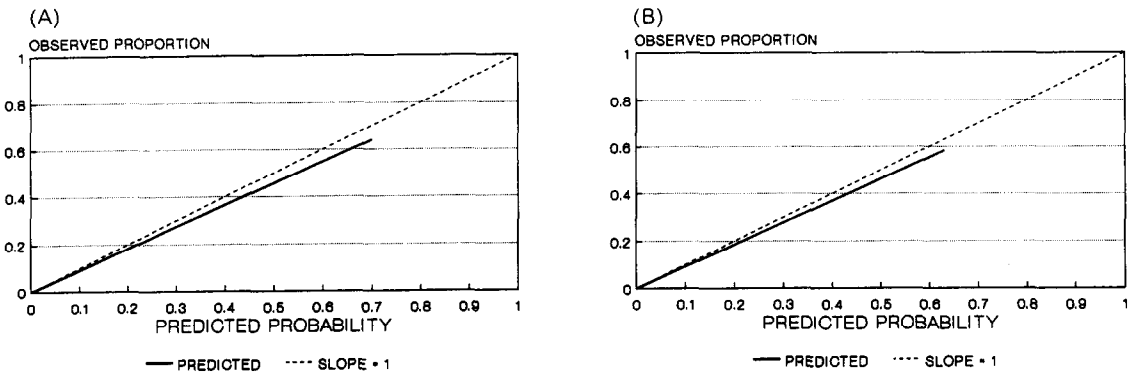

Fig. 2. Reliability of the predictions. Predictions of any colorectal neoplasia (A) and adenomatous polyps ⩾5 mm or cancer (B) based on the training sample were applied to all patients in the test sample. Shown are the probabilities of colorectal neoplasia estimated by logistic-linear calibration (solid line). If predictions are reliable, the solid line should fall on the line with a slope of 1 (dashed line).

evaluation and the decision process to perform colonoscopy after that evaluation [19, 20]. Consequently, our model should not be used to generate risk estimates in clinical practice.

Future studies will be required to determine the generalizability of our findings. However, clinical characteristics identified in our study should be included as potentially important risk factors in these studies involving unselected cohorts. Approaches such as ours are likely to be useful in developing unbiased risk estimate models when applied to these populations.

The study also demonstrates the value of ordinal logistic regression to identify important characteristics of ordinal outcomes. Colorectal neoplasia usually follows an orderly progression from adenoma to carcinoma [21]. The ordinal scale weights this progression appropriately: small polyps are worse than no polyps, large polyps are worse than small polyps; and cancer is worse than large polyps.

The model codes the risk of advancing age on a continuous scale. Previous studies [11, 12] have shown that the risk of colorectal cancer nearly doubles with each decade of life after the age of 50. The model's estimates for age-related risk is a 1.6 fold rise (95% CI: 1.36–1.89) in neoplasia risk with each decade of life.

The model also recognizes that risks vary given the presence of more than one risk factor. The FOBT has been subject to much recent criticism as a screening test for colorectal neoplasia due to its relatively low sensitivity and specificity [22]. However, the predictive value of this test is dependent on the prevalence of disease in the population studied which varies with age. Prior studies have shown that colorectal cancer is found in approximately 12% and adenomatous polyps in approximately 35% of patients over 63 with a positive FOBT who volunteer for screening. Conversely, colorectal cancer is found in 5% and adenomatous polyps in 26% for similar subjects under 63 [10, 23]. The logistic model provides more precise risk estimates of a positive FOBT by using age on

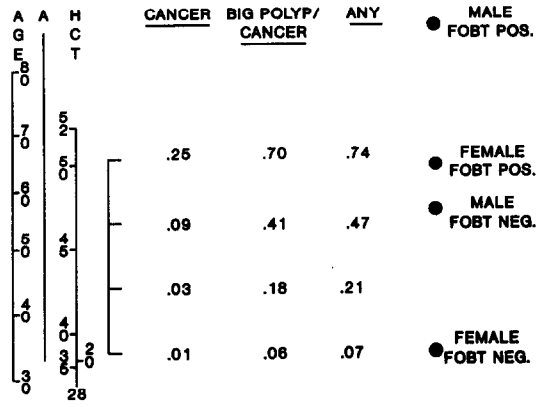

Fig. 3. A nomogram for estimating the likelihood of any colorectal neoplasia, an adenomatous polyp ⩾5 mm or cancer. *Directions for use*: (1) If diarrhea or pain is the sole indication for colonoscopy, probability of colorectal neoplasia <0.01. If not, locate the patient's age on the scale at the left and their hematocrit on the scale to its right. (2) Place a ruler between these points and mark this point on line A. (3) Locate the point appropriate to the patient's sex and FOBT result. (4) Place a ruler between this point and the mark on line A. (5) Read the probability of colorectal neoplasia on the center scale.

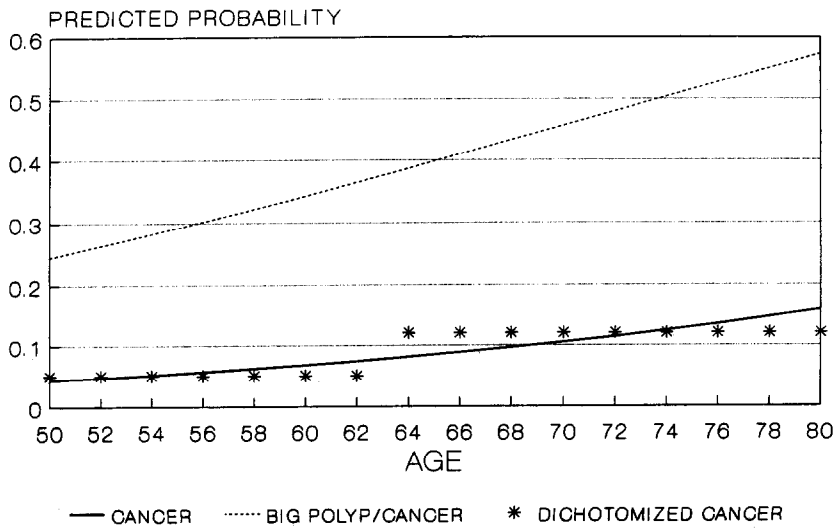

Fig. 4. Improved predictions of risk using age as a continuous variable. The predicted probability of colorectal cancer (solid line) and any colorectal neoplasia (dashed line) in women with a positive FOBT vs published risk [11, 27] estimates when dichotomizing age at 63 years (*).

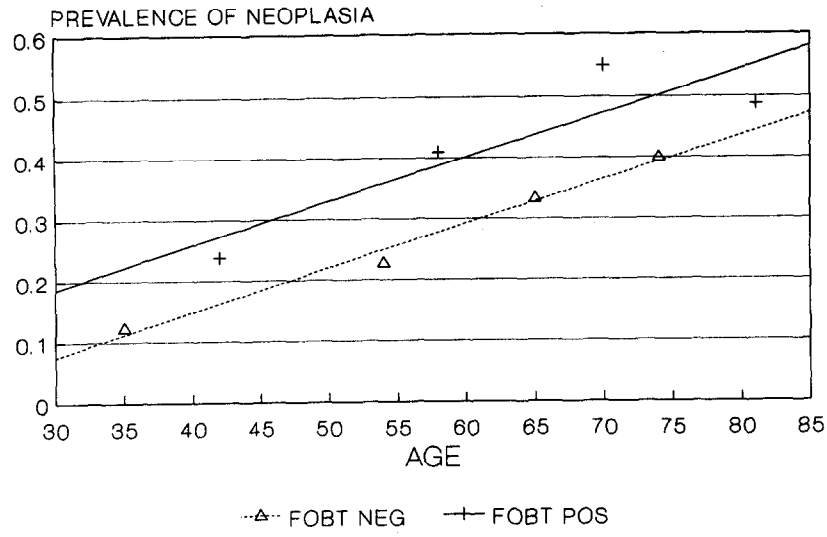

Fig. 5. Lack of interaction between age and FOBT. The patients have been grouped by their FOBT result (positive = solid line, negative = dashed line) and thereafter divided into four quantiles by age. The mean prevalence of any colorectal neoplasia is plotted. The lack of interaction is shown by the two parallel lines.

a continuous scale (Fig. 4). Furthermore, the study demonstrates that the information content in the FOBT is similar at different ages, as demonstrated by the lack of an important interaction between the two characteristics (Fig. 5).

While the presence or absence of given predictors in the model may make biologic or epidemiologic "sense", their presence (or absence) may only represent biases present in the patient population studied. For example, while some investigators have found an increased prevalence of colorectal neoplasia in men undergoing screening colonoscopy [24], the association of male sex with colorectal neoplasia seen in this study may have resulted from the increased investigation of women with the irritable bowel syndrome. Known risk factors for the development of colon cancer: IBD [25–27]; family [28, 29] or personal [30, 31] history of colorectal neoplasia; cholecystectomy [31]; and history of breast or female genital tract cancer [11, 32]; were not included in the model. Again, this may reflect peculiarities of the population studied or the low prevalence of these risk factors in the subjects studied.

Symptoms and signs commonly used to detect colon cancer (abdominal pain, constipation, hematochezia, and iron deficiency) were not helpful in the identification of patients at high risk for colorectal neoplasia. These systems may have prompted the referral of these patients to the gastroenterologist; their value may have been to discriminate a moderate risk subset from those not referred. However, an overwhelming percentage (96%) of patients with colorectal neoplasia in the current study had localized (Duke's Stage A or B) colon cancer or colorectal polyps. These lesions rarely cause pain, altered bowel habits, hematochezia, or iron deficiency [33].

Endoscopic examination of the entire colon is the gold standard for the diagonsis of colon neoplasia. Several reasons, including unacceptable cost and risk proscribe routine application of colonoscopy unless a high risk group is the target. Until more sophisticated diagnostic tests (e.g. accurate genetic and/or biochemical tests) become available, patients at high risk can only be identified by analysis of clinical characteristics obtained from patients with localized colorectal neoplasia. Although a selection bias is implicit in the present referral pattern, the model developed from analysis of clinical variables reliably predicted colon neoplasia in a test sample of this population. Particularly important, is the application of this model to a relatively large number of patients harboring localized disease. Application of similar techniques to independent patient populations could provide rationale for this approach to early diagnosis of colon neoplasia.

One may conclude that accurate estimates of colorectal neoplasia in selected patients are possible using ordinal logistic regression analysis. It is hoped that this technique, not the specific model developed, is applied to an unselected population in order to obtain a

model providing unbiased predictions. Future studies should include the clinical characteristics identified in our study as candidate variables as well as a detailed dietary [34–36] and bowel habit history [37].

Acknowledgements—This work was supported, in part, by research grants from the National Institutes of Health AM07166-10, the National Center for Health Services Research HS04873, HS05635, the National Heart, Lung, and Blood Institute HL36587, the National Library of Medicine 1G08 LM04613, and a grant from the Andrew W. Mellon Foundation, New York.

REFERENCES

1. Hlatky MA, Califf RM, Harrell FE Jr, Lee KL, Mark DB, Muhlbaier LH, Pryor DB. Clinical judgement and therapeutic decision making. **J Am Coll Cardiol** 1990; 15: 1–14.
2. Armstrong BG, Sloan M. Ordinal regression models for epidemiologic data. **Am J Epidemiol** 1989; 129: 191–204.
3. Peterson BL, Harrell FE Jr. Partial proportional odds models for ordinal response variables. **Appl Stat** 1990; 39: 205–217.
4. Harrell FE, Lee KL. The practical value of logistic regression. **Proc 10th Annual SAS User's Group International Conference**, Cary, NC. SAS Institute; 1985: 1031–1036.
5. Lee KL, Pryor DB, Harrell FE Jr *et al.* Predicting outcome in coronary disease: statistical models versus expert clinicians. **Am J Med** 1986; 80: 553–560.
6. Kong DF, Lee KL, Harrell FE Jr *et al.* Clinical experience and predicting survival in coronary disease. **Arch Intern Med** 1989; 149: 1177–1181.
7. Silverberg E, Boring CC, Squires TS. Cancer Statistics, 1990. **CA** 1990; 40: 9–26.
8. Silverberg E, Holleb AI. Major trends in cancer: 25 year survey. **CA** 1975; 25: 2–7.
9. Winawer SJ. Screening for colorectal cancer: an overview. **Cancer** 1980; 45: 1093–1098.
10. Winawer SJ, Sherlock P. Surveillance for colorectal cancer in average-risk patients, familial high-risk groups, and patients with adenomas. **Cancer** 1982; 50: 2609–2614.
11. Schottenfield D, Winawer SJ. Large intestine. In: Schottenfeld D, Fraumen J Jr, Eds. **Cancer Epidemiology and Prevention**. Philadelphia: W.B. Saunders; 1982: 703–727.
12. Winawer SJ, Sherlock, P, Schottenfeld D *et al.* Screening for colon cancer. **Gastroenterology** 1976; 70: 783–789.
13. Sherlock P, Winawer SJ. The role of early diagnosis in controlling large bowel cancer. **Cancer** 1977; 40: 2609–2615.
14. Walker SH, Duncan DB. Estimation of the probability of an event as a function of several independent variables. **Biometrika** 1967; 34: 167–179.
15. Harrell FE. The LOGIST procedure. In: Hastings RP, Ed. **SUGI Supplemental Library User's Guide**, version 5 edn. Cary, NC: SAS Institute Inc.; 1986: 269–293.
16. Harrell FE, Lee KL, Califf RM, Pryor DB, Rossati RA. Regression modelling strategies for improved prognostic prediction. **Stat Med** 1984; 3: 143–153.
17. Goodman LA, Kruskal WH. **Measures of Association for Cross-Classifications**. New York: Springer-Verlag; 1979.
18. Banks J. Nomograms. In: Kotz S, Johnson NL, Eds. **Encyclopedia of Statistical Sciences**. Vol. 6: **Multivariate Analysis—Plackett and Burman Designs**. New York: Wiley; 1985: 265–271.
19. Philbrick JT, Horwitz RI, Feinstein AR. Methodologic problems of exercise testing for coronary artery disease: groups, analysis and bias. **Am J Cardiol** 1980; 46: 807–812.
20. Sackett DL. Bias in analytical research. **J Chron Dis** 1979; 32: 51–63.
21. Boland CB, Itzbowitz SH, Kim YS. Colonic polyps and the gastrointestinal polyposis syndromes. In: Sleisenger MH, Fordtran JS, Eds. **Gastrointestinal Disease: Pathophysiology, Diagnosis, Management**. Philadelphia: W.B. Saunders; 1989: 1487–1489.
22. Simon JB. Occult blood screening for colorectal carcinoma: a critical review. **Gastroenterology** 1985; 88: 820–837.
23. Gilbertsen VA, McHugh R, Schuman L, Williams SE. The earlier detection of colorectal cancers. **Cancer** 1980; 45: 2899–2901.
24. Rex D, Lehman G, Hawes R *et al.* Screening colonoscopy in asymptomatic average-risk volunteers with negative fecal occult blood tests; age and sex differences. **Am J Gastroenterol** 1990; 85: 274A.
25. Devroede GJ, Taylor WF, Sauer WG, Jackman RJ, Stickler GB. Cancer risk and life expectancy of children with ulcerative colitis. **N Engl J Med** 1971; 285: 17–21.
26. Ekbom A, Helmick C, Zack M, Adami H. Ulcerative colitis and colorectal cancer. **N Engl J Med** 1990; 323: 1228–1233.
27. Katzka I, Brody RS, Morris E, Katz S. Assessment of colorectal cancer risk in patients with ulcerative colitis: experience from a private practice. **Gastroenterology** 1983; 85: 22–29.
28. Kussin SZ, Lipkin M, Winawer SJ. Inherited colon cancer: clinical implications. **Am J Gastroenterol** 1979; 72: 448–457.
29. Schottenfeld D. Patient risk factors and the detection of early cancer. **Prev Med** 1972; 1: 335–351.
30. Lotfi AM, Spencer RJ, Ilsturp DM, Melton J. Colorectal polyps and the risk of subsequent carcinoma. **Mayo Clin Proc** 1986; 61: 337–343.
31. Vernick LJ, Kuller LH. A case–control study of cholecystectomy and right-sided colon cancer: the influence of alternative data sources and differential interview participation proportion on odds ratio estimates. **Am J Epidemiol** 1982; 116: 86–101.
32. Schottenfeld D, Berg JW, Vitsky B. Incidence of multiple primary cancers. II. Index cancers arising in the stomach and lower digestive system. **J Natl Cancer Inst** 1969; 43: 77–86.
33. MacRae FA, St John JB. Relationship between patterns of bleeding and hemoccult sensitivity in patients with colorectal cancers or adenomas. **Gastroenterology** 1982; 82: 891–898.
34. Modan B. Role of diet in cancer etiology. **Cancer** 1977; 40: 1887–1891.
35. Hoff G, Moen IE, Trygg K *et al.* Epidemiology of polyps in the rectum and sigmoid colon. **Scand J Gastroenterol** 1986; 21: 199–204.
36. Hoff G, Larsen S. Epidemiology of polyps in the rectum and sigmoid colon. Discriminant analysis for identification of individuals at risk of developing colorectal neoplasia. **Scand J Gastroenterol** 1986; 7: 848–852.
37. Burkitt DP, Walker ARP, Painter NS. Dietary fiber and disease. **JAMA** 1974; 229: 1068–1074.