18 Malmgren R, Collins A, Nilsson CG. Platelet serotonin uptake and effects of vitamin B6-treatment in premenstrual tension. *Neuropsychobiology* 1987;18:83-6.
19 Jadad A, Moore M, Carrol D, Jenkinson C, Reynolds DJM, Gavaghan DJ, et al. Assessing the quality of reports of randomised clinical trials; is blinding necessary? *Control Clin Trials* 1996;17:1-12.
20 Computer programs for epidemiologic analysis: PEPI version 2. Gahlinger PM and Abramson JH. 2. Stone Mountain, GA: USD, 1995.
21 Hasselblad V, Hedges LV. Meta-analysis of screening and diagnostic tests. *Psychol Bull* 1995;117:167-78.
22 Egger M, Smith GD, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;315:629-34.
23 Stokes J, Mendels J. Pyridoxine and premenstrual tension. *Lancet* 1972;i:1177-8.
24 Colin C. Studies on the treatment of mastalgia. *Rev Med Brux* 1982;3:605-9.
25 Barr W. Pyridoxine supplements in the premenstrual syndrome. *Practitioner* 1984;228:425-7.
26 Williams MJ, Harris RI, Dean BC. Controlled trial of pyridoxine in the premenstrual syndrome. *J Int Med Res* 1985;13:174-9.
27 Hagen I, Nesheim BI, Tuntland T. No effect of vitamin B-6 against premenstrual tension. A controlled clinical study. *Acta Obstet Gynecol Scand* 1985;64:667-70.
28 Chakmakjian ZH, Higgins CE, Abraham GE. The effect of a nutritional supplement, Optivite for women, on premenstrual tension syndromes. *J Appl Nutr* 1985;37:12-7.
29 Smallwood J, Ah-Kye D, Taylor I. Vitamin B6 in the treatment of pre-menstrual mastalgia. *Br J Clin Pract* 1986;40:532-3.
30 Stewart A. Clinical and biochemical effects of nutritional supplementation on the premenstrual syndrome. *J Reprod Med* 1987;32:435-41.
31 Kendall KE, Schnurr PP. The effects of vitamin B6 supplementation on premenstrual symptoms. *Obstet Gynecol* 1987;70:145-9.
32 Doll H, Brown S, Thurston A, Vessey M. Pyridoxine (vitamin B6) and the premenstrual syndrome: a randomized crossover trial. *J R Coll Gen Pract* 1989;39:364-8.
33 London RS, Bradley L, Chiamori NY. Effect of a nutritional supplement on premenstrual symptomatology in women with premenstrual syndrome: a double blind longitudinal study. *J Am Coll Nutr* 1991;10:494-9.
34 Kerr GD. The management of the premenstrual syndrome. *Curr Med Res Opin* 1977;4:29-34.
35 Day JB. Clinical trials in the premenstrual syndrome. *Curr Med Res Opin* 1979;6:40-5.
36 Abraham GE, Hargrove JT. Effect of vitamin B6 on premenstrual symptomology in women with premenstrual tension syndromes: a double blind crossover study. *Infertility* 1980;3:155-65.
37 Mattes JA, Martin D. Pyridoxine in premenstrual depression. *Hum Nutr Appl Nutr* 1982;36:131-3.
38 Goei GS, Abraham GE. Effect of a nutritional supplement, Optivite, on symptoms of premenstrual tension. *J Reprod Med* 1983;28:527-31.
39 Fuchs N, Hakim M, Abraham GE. The effect of a nutritional supplement, Optivite for women, on premenstrual tension syndromes. *J Appl Nutr* 1985;37:1-11.
40 Brush MG, Bennett T, Hansen K. Pyridoxine in the treatment of premenstrual syndrome: a retrospective survey in 630 patients. *Br J Clin Pract* 1988;42:448-52.
41 Brush MG. Vitamin B6 treatment of premenstrual syndrome. In: Leklem JE, Reynolds RD, eds. *Clinical and physiological applications of vitamin B6*. New York: Liss, 1988: 363-79. (*Current topics in nutrition and disease*, vol 19).
42 Van den Berg H, Schrijver J, Bruinse HW, van der Ploeg HM. Vitamin B6 and premenstrual syndrome. *Voeding* 1989;50:58-62.
43 Berman MK, Taylor ML, Freeman E. Vitamin B-6 in premenstrual syndrome. *J Am Diet Assoc* 1990;90:859-61.
44 De Souza MC, Walker AF, Bolland KM, Robinson PA. A synergistic effect of magnesium and vitamin B6 supplementation for the relief of symptoms of the premenstrual syndrome (PMS). *Proc Nutr Soc* 1996;56:75A.
45 Harrison W, Endicott J, Rabkin JG, Nee J. Treatment of dysphoric changes: clinical outcome and methodological implications. *Psychol Bull* 1984;118-22.
46 Blue J, Harman J. *Mastalgia review*: St Marks Breast Centre. *NZ Med J* 1998;111:34-7.
47 Villegas-Salas E, Ponce dL, Juarez-Perez MA, Grubb GS. Effect of vitamin B6 on the side effects of a low-dose combined oral contraceptive. *Contraception* 1997;55:245-8.
48 Bernstein AL. Vitamin B6 in clinical neurology. *Ann NY Acad Sci* 1990;585:250-60.
49 Stewart JW, Harrison W, Quitkin F, Baker H. Low B6 levels in depressed outpatients. *Biol Psychiatry* 1984;19:613-6.
50 Adams PW, Rose DP, Folkard J, Wynn V, Seed M, Strong R. Effects of pyridoxine hydrochloride upon depression associated with oral contraceptives. *Lancet* 1973;:897-904.
51 Winston F. Oral contraceptives, pyridoxine, and depression. *Am J Psychiatry* 1973;130:1217-21.
52 Folkers K, Ellis J. Successful therapy with vitamin B6 and vitamin B2 of the carpal tunnel syndrome and need for determination of RDAs for vitamins B6 and B2 for disease states. *Ann NY Acad Sci* 1989;295-301.

(Accepted 11 March 1999)

Influence of hospital and clinician workload on survival from colorectal cancer: cohort study

F Kee, R H Wilson, C Harper, C C Patterson, K McCallion, R F Houston, R J Moorehead, J M Sloan, B J Rowlands

Abstract

Objective To determine whether clinician or hospital caseload affects mortality from colorectal cancer.
Design Cohort study of cases ascertained between 1990 and 1994 by a region-wide colorectal cancer register.
Outcome measures Mortality within a median follow up period of 54 months after diagnosis.
Results Of the 3217 new patients registered over the period, 1512 (48%) died before 31 December 1996. Strong predictors of survival both in a logistic regression (fixed follow up) and in a Cox's proportional hazards model (variable follow up) were Duke's stage, the degree of tumour differentiation, whether the liver was deemed clear of cancer by the surgeon at operation, and the type of intervention (elective or emergency and curative or palliative intent). In a multilevel model, surgeon's caseload had no significant effect on mortality at 2 years. Hospital workload, however, had a significant impact on survival. The odds ratio for death within 2 years for cases managed in a hospital with a caseload of between 33 and 46 cases per year, 47 and 54 cases per year, and ≥55 cases per year (compared to one with ≤23 cases per year) were respectively 1.48 (95% confidence interval 1.03 to 2.13), 1.52 (1.08 to 2.13), and 1.18 (0.83 to 1.68).
Conclusions There was no detectable caseload effect for surgeons managing colorectal cancer, but survival of patients treated in hospitals with caseloads above 33 cases per year was slightly worse than for those treated in hospitals with fewer caseloads. Imprecise measurement of clinician specific "events rates" and the lack of routinely collected case mix data present major challenges for clinical audit and governance in the years ahead.

Department of Epidemiology and Public Health, Queen's University of Belfast, Belfast BT12 6BJ
F Kee, *professor*
C C Patterson, *senior lecturer in medical statistics*
Northern Ireland Centre for Clinical Oncology, Belvoir Park Hospital, Belfast BT8 8JR
R H Wilson, *specialist registrar*
R F Houston, *consultant*

continued over

BMJ 1999;318:1381–6

Introduction

In 1995, the then chief medical officer, Sir Kenneth Calman, made recommendations for improving

Northern Ireland Cancer Register, Belvoir Park Hospital
C Harper, *specialist registrar in public health medicine*
Department of Surgery, Queen's University of Belfast, Belfast BT12 6BJ
K McCallion, *specialist registrar*
Ulster, North Down and Ards Health and Social Services Trust, Dundonald, Belfast BT16 1RH
R J Moorehead, *consultant surgeon*
Department of Pathology, Queen's University of Belfast, Belfast BT12 6BJ
J M Sloan, *reader*
Department of Surgery, Queens Medical Centre, Nottingham NG7 2UH
B J Rowlands, *professor*
Correspondence to: Professor Kee
f.kee@qub.ac.uk

cancer care in England and Wales.[1] Care was to be focused in a network of cancer units and centres that would have sufficient workload to permit appropriate subspecialisation and the development of expertise.

Those charged with implementing this recommendation have been exercised by the putative association between patient outcome and the number of procedures undertaken, whether by surgeon or by hospital. The Royal College of Surgeons has stated only that all patients requiring surgical management should be treated by surgeons with appropriate training and experience.[2]

The Clinical Outcomes Group recently stated that the quality of colorectal surgery can have a strong influence on patient survival.[3] The Centre for Reviews and Dissemination further advised that concentrating surgery in the hands of those with better results could improve survival.[4] Regardless of the reason, surgeon variability may influence outcome by as much as 20%.[5] Evidence on the subject is inconclusive for several reasons, but "if there is no relationship, then the whole process of training and accreditation is called into question."[6]

Detecting a relation often depends on the range over which observations are made, and commissioners of care may reasonably ask whether there is any relation locally between volume and outcome. Equally importantly, it should be asked whether surgical audit can reliably distinguish the "good achievers" from the rest and thus influence the shape of future services.

Our study attempts to address these issues, in particular whether patient mortality is related to the number of cases for which a surgeon or hospital is responsible.

Subjects and methods

Colorectal cancer register—The Northern Ireland colorectal cancer register was established in 1990 with the support of a local cancer charity. Although it produces an independent annual report, it now works more closely with the Northern Ireland cancer register. Clinical and pathological details of colorectal cancers diagnosed in Northern Ireland residents are collected from all hospitals and pathology laboratories serving the provinces (1.6 million population). Although we have previously reported incidence data,[7 8] this report is based on follow up (to the end of 1996) of the first 5 year cohort diagnosed with colorectal cancer between 1990 and 1994.

Subjects—Data recorded on each patient included the age, sex, and address of the patient, the site and Duke's stage of the tumour, the nature of surgery undertaken (emergency or elective and curative intent or palliative intent), and the consultant surgeon and hospital in charge of the case. The address of the patient was mapped to the electoral ward. The province's 566 wards were divided into fifths on the basis of their Townsend score, allowing each to be categorised according to a measure of material deprivation.

Follow up—Each year a follow up questionnaire is dispatched to the consultant in charge and cause of death is ascertained from the registrar general for all deceased patients. Our report deals with follow up status of patients at 31 December 1996. The median follow up period for our cohort was 53.8 months.

Table 1 Demographic and case mix characteristics of cases 1990-4

Characteristic	No (%) of cases (n=3135)
Age group (years)	
<60	686 (21.9)
60-69	828 (26.4)
70-79	1071 (34.2)
⩾80	550 (17.6)
Sex	
Male	1701 (54.3)
Female	1434 (45.7)
Townsend deprivation fifth	
1	753 (24.0)
2	615 (19.6)
3	526 (16.8)
4	500 (15.9)
5	558 (17.8)
Not known	183 (5.8)
Duke's stage	
A	93 (3.0)
B	1319 (42.1)
C	740 (23.6)
D	643 (20.5)
Not known	340 (10.8)
Consultant workload (cases per year)	
⩽9.7	617 (19.7)
9.8-12.7	685 (21.9)
12.8-16.1	604 (19.3)
16.2-24.9	599 (19.1)
⩾25.0	614 (19.6)
Not known	16 (0.5)
Consultant experience (years)	
⩽13	612 (19.5)
14-17	639 (20.4)
18-22	753 (24.0)
23-30	443 (14.1)
⩾31	568 (18.1)
Not known	120 (3.8)
Hospital workload (per year)	
⩽23	600 (19.1)
24-32	643 (20.5)
33-46	584 (18.6)
47-54	571 (18.2)
⩾55	737 (23.5)
Tumour site	
Right colon	882 (28.1)
Total colon	204 (6.5)
Left colon	895 (28.5)
Rectum	1073 (34.2)
Multiple	68 (2.2)
Not known	13 (0.4)

Statistical methods

We first explored simple univariate predictors of survival at 2 years using a χ^2 test for each variable. We then identified independent predictors by using logistic regression. Surgery over the 5 year period was undertaken in 19 different hospitals and by 71 different surgeons. The surgeons and the hospitals were ranked according to the annual number of cases for which they were responsible, and the distribution was subdivided into fifths. Thirteen surgeons each managed fewer than 10 cases over the 5 year period, and we aggregated them and their combined total cases[59] into one group. Using data from the medical

Table 2 Univariate predictors of survival at 2 years

	Total No	No (%) of deaths at 2 years	P value (χ^2)
Age group (years)			
<60	686	210 (30.6)	<0.001
60-69	828	267 (32.2)	
70-79	1071	381 (35.6)	
≥80	550	238 (43.3)	
Duke's stage			
A	93	10 (10.8)	<0.001
B	1319	253 (19.2)	
C	740	254 (34.3)	
D	643	410 (63.8)	
Not known	340	169 (49.7)	
Differentiation			
Well differentiated	280	70 (25.0)	<0.001
Moderate	2088	646 (30.9)	
Poor	413	206 (49.9)	
Not known	354	174 (49.2)	
Liver deemed clear by surgeon at operation			
Yes	2171	623 (28.7)	<0.001
No	469	324 (69.1)	
Not known	495	149 (30.1)	
Type of procedure			
Elective+curative intent	1798	393 (21.9)	<0.001
Elective+palliative intent	576	399 (69.3)	
Emergency+curative intent	261	81 (31.0)	
Emergency+palliative intent	169	123 (72.8)	
Not known	331	100 (30.2)	

register, we also described each surgeon according to the number of years they had been a medical practitioner.

We used multilevel logistic regression modelling[9–11] of the survival status at 2 years after surgery to allow for possible clustering among cases within surgeons and among surgeons within hospitals. In constructing this three level hierarchy, we assigned patients to the hospitals in which their surgeon was mainly based.

Although we have presented findings using the logistic regression model for fixed follow up at 2 years, the results were broadly comparable to those when Cox's proportional hazards models analysis for the entire follow up period was used.

For illustrative purposes, we split the data into two periods, representing performance for the first 3 years and the last 2 years. The performance of each surgeon (according to their patients' adjusted mortality), was assessed in the first of these periods by comparing it with an (arbitrarily defined) acceptable mortality. We then determined which surgeons achieved or fell below this acceptable standard in the second period.

Results

Of the 3217 new patients registered for the period, 82 (2.5%) either had a biopsy only or there was insufficient information to determine the procedure, and these patients were excluded from the survival analysis.

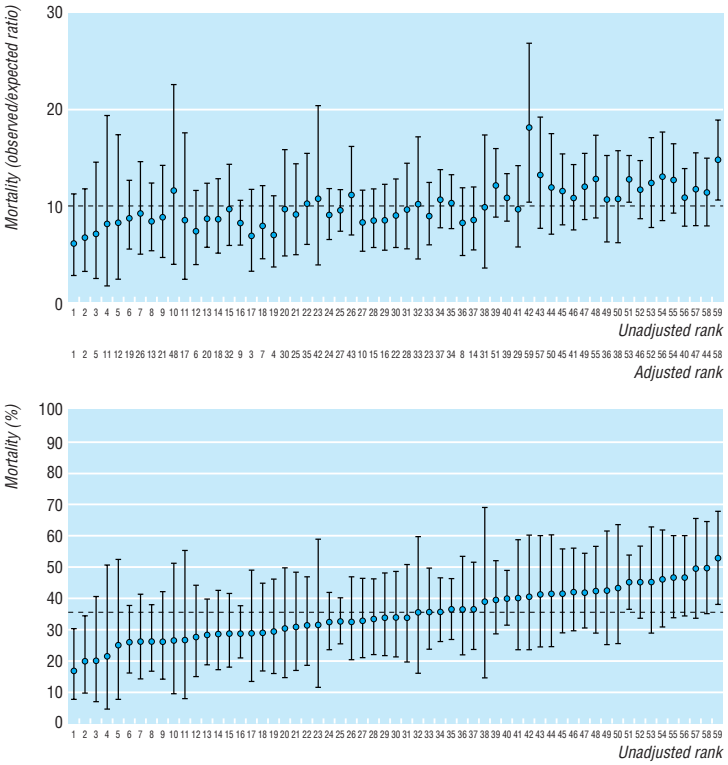
Before 31 December 1996, 1512 patients (48.2%) had died; table 1 shows their characteristics. In 340 cases (10.8%) inadequate clinical information was retrieved for Duke's stage to be assigned, and in 183 cases (5.8%) insufficient address information precluded postcoding (and Townsend scoring).

In univariate analysis, significant predictors of survival at 2 years were age, Duke's stage, the degree of tumour differentiation, whether the liver was deemed clear by the surgeon, and the type of surgery (table 2). As we observed no significant effect on survival for tumour site, we combined the sites for further analysis.

After adjustment for case mix in a single level logistic regression analysis, there was only weak evidence of surgeon specific heterogeneity in patient survival ($\chi^2=76.1$; df=58; P=0.06). Likewise, there was only weak evidence of significant hospital specific heterogeneity in patient outcome. ($\chi^2=27.1$; df=18; P=0.08). We found no detectable effect of consultant workload or "experience" on outcome. Table 3 shows the relative odds of death at 2 years derived from the standard and multilevel logistic regression models and the relative hazards from a Cox's proportional hazards model. The base model contains the biological predictors before we simultaneously forced in the hospital and clinician variables.

In both the standard and multilevel models, the hospitals with higher caseloads—particularly those dealing with between 33 and 54 cases per year—had a slightly worse survival.

Using the entire 5 year cohort of patients, we ranked the performance of the surgeons firstly on the basis of unadjusted 2 year survival and then on survival adjusted for all the significant case mix variables in the logistic model (figure). Using an excess mortality of about 15% (roughly equivalent to the 75th centile of the surgeon's distribution), we compared the ranking of surgeons over the first 3 years with that over the last 2 years. Fourteen surgeons would have failed the test if a health authority had designated its specialist colorectal cancer surgeons by selecting those who achieved a mortality within the acceptable limits for patients treated between 1990 and



Consultants ranked (unadjusted, or adjusted for case mix) by mortality of patients after 2 years

Table 3 Odds ratios and hazard ratios in multivariate model of survival

	Relative odds of death (95% CI)*	P value	Multilevel model†	P value	Relative hazard (95% CI)‡	P value
Type of procedure (base=elective and curative intent)		<0.0001		<0.0001		<0.0001
Elective and palliative	4.26 (3.35 to 5.41)	<0.0001	4.26 (3.35 to 5.42)	<0.0001	2.59 (2.25 to 2.97)	<0.0001
Emergency and curative	1.52 (1.13 to 2.04)	0.006	1.51 (1.12 to 2.03)	0.007	1.36 (1.12 to 1.65)	0.002
Emergency and palliative	5.68 (3.88 to 8.30)	<0.0001	5.65 (3.87 to 8.27)	<0.0001	3.27 (2.69 to 3.98)	<0.0001
Not known	1.09 (0.78 to 1.54)	0.61	1.10 (0.78 to 1.54)	0.60	1.15 (0.92 to 1.44)	0.21
Liver clear (base=yes)		0.004		0.004		0.008
No *v* yes	1.60 (1.18 to 2.17)	0.002	1.61 (1.19 to2.18)	0.002	1.31 (1.1 to 1.55)	0.002
Not known	1.29 (0.97 to 1.73)	0.08	1.30 (0.97 to 1.74)	0.08	1.01 (0.84 to 1.21)	0.93
Age (per year)	1.27 (1.18 to 1.37)	<0.0001	1.27 (1.18 to 1.37)	<0.0001	1.18 (1.13 to 1.24)	<0.0001
Duke's stage (base=A)		<0.0001		<0.0001		<0.0001
B *v* A	1.52 (0.77 to 3.04)	0.23	1.51 (0.76 to 3.01)	0.24	1.41 (0.9 to 2.22)	0.14
C *v* A	2.94 (1.46 to 5.9)	0.002	2.92 (1.45 to 5.87)	0.003	2.28 (1.44 to 3.61)	0.0001
D *v* A	4.48 (2.17 to 9.27)	0.0001	4.45 (2.15 to 9.2)	0.0001	2.64 (1.64 to 4.23)	0.0001
Not known	3.63 (1.58 to 8.32)	0.002	3.58 (1.56 to 8.24)	0.003	2.68 (1.57 to 4.57)	0.0001
Tumour differentiation (base=well differentiated)		0.0001		0.0001		0.004
Moderate *v* well differentiated	1.07 (0.77 to 1.48)	0.70	1.07 (0.77 to 1.48)	0.70	1.02 (0.84 to 1.23)	0.88
Poor *v* well differentiated	1.80 (1.23 to 2.64)	0.003	1.82 (1.24 to 2.66)	0.002	1.31 (1.05 to 1.65)	0.02
Not known *v* well differentiated	1.26 (0.71 to 2.25)	0.43	1.27 (0.71 to 2.27)	0.42	1.06 (0.75 to 1.51)	0.74
Hospital workload (per year; Q1 ≤23 years)		0.01		0.02		0.002
(24-32) Q2 *v* Q1	0.93 (0.7 to 1.25)	0.65	0.93 (0.69 to 1.26)	0.66	0.89 (0.74 to 1.06)	0.19
(33-46) Q3 *v* Q1	1.47 (1.04 to 2.08)	0.03	1.48 (1.03 to 2.13)	0.03	1.14 (0.92 to 1.4)	0.23
(47-54) Q4 *v* Q1	1.53 (1.11 to 2.12)	0.01	1.52 (1.08 to 2.13)	0.02	1.29 (1.07 to 1.57)	0.01
(≥55) Q5 *v* Q1	1.18 (0.86 to 1.63)	0.30	1.18 (0.83 to 1.68)	0.35	1.05 (0.87 to 1.27)	0.59
Consultant experience (years; Q1 ≤13 years)		0.45		0.46		0.70
(14-17) Q2 *v* Q1	1.09 (0.81 to 1.48)	0.55	1.10 (0.8 to 1.52)	0.56	1.09 (0.91 to 1.31)	0.36
(18-22) Q3 *v* Q1	0.88 (0.64 to 1.22)	0.44	0.88 (0.63 to 1.24)	0.47	0.99 (0.81 to 1.21)	0.93
(23-30) Q4 *v* Q1	1.09 (0.78 to 1.52)	0.62	1.07 (0.75 to 1.53)	0.70	1.08 (0.88 to 1.32)	0.48
(≥31) Q5 *v* Q1	1.05 (0.75 to 1.47)	0.77	1.06 (0.75 to 1.51)	0.73	1.07 (0.87 to 1.31)	0.53
Not known *v* Q1	1.50 (0.87 to 2.6)	0.14	1.54 (0.88 to 2.72)	0.13	1.27 (0.9 to 1.8)	0.18
Consultant workload (per year; Q1 ≤9.7 years)		0.65		0.70		0.35
(9.8-12.7) Q2 *v* Q1	1.15 (0.87 to 1.52)	0.33	1.14 (0.85 to 1.53)	0.37	1.03 (0.87 to 1.22)	0.71
(12.8-16.1) Q3 *v* Q1	0.97 (0.73 to 1.29)	0.84	0.97 (0.72 to 1.3)	0.82	1.01 (0.85 to 1.2)	0.93
(16.2-24.9) Q4 *v* Q1	1.13 (0.83 to 1.55)	0.44	1.13 (0.81 to 1.57)	0.47	1.10 (0.91 to 1.34)	0.31
(≥25.0) Q5 *v* Q1	0.92 (0.66 to 1.27)	0.59	0.91 (0.65 to 1.29)	0.60	0.88 (0.73 to 1.07)	0.21
Not known *v* Q1	0.63 (0.16 to 2.44)	0.50	0.61 (0.16 to 2.38)	0.48	0.95 (0.41 to 2.18)	0.90

*Ordinary logistic regression model. †Multilevel logistic regression. ‡Cox's proportional hazards model.

1993. However, over the next 2 years, six of them would have become a "safe" pair of hands and 11 others would no longer satisfy the criterion.

Discussion

We have shown that mortality within 3-4 years of diagnosis of colorectal cancer is not associated with annual caseload for surgeons in Northern Ireland dealing with up to 34 cases per year. Only 44% of the cases in this region were managed by surgeons dealing with 15 or more cases per year. This probably differs little from elsewhere in the United Kingdom.

The register we used ascertains its cases mostly through the pathology laboratories, so even if cases given only a clinical diagnosis were missed, their exclusion would not have affected the conclusions for surgical performance.

Other studies have investigated intermediate outcomes,[12] and although we have not directly studied the surgeons' technical craft, we speculated whether for patients with similar Duke's stage, the type of surgery differed among surgeons with high and low workloads. Indeed, there was no difference in the proportions of curative versus palliative intent or elective versus emergency procedures. The proportion of cases with insufficient data for staging was broadly comparable across workload categories and we do not think this has biased our conclusions. Imprecise case mix adjustment, particularly if it was differential, may have biased the expected mortality. The collection of good quality information by surgeons is thus an indirect measure of outcome quality.

Although McArdle and Hole showed some significant difference between surgeon variation in patients' survival,[12] they did not specifically investigate volume effects. Initial analyses from the Trent and Wales audits given in the report of the Clinical Outcomes Group show no case mix adjusted survival advantage (at 2 years) for patients treated by self reported specialists. Differences have been shown, for example, between survival after treatment in teaching and non-teaching hospitals,[13] but whether they originate from selection bias or workload is not known.

Significant effects of operator workload have been shown in substantial literature from the United States, but most of these studies have focused on colectomy rather than on colorectal cancer.[14–16] The necessity to adjust simultaneously for effects of both hospital and operator specific workload is increasingly acknowledged. Poor outcomes in a specific hospital may be the result of the good outcomes of one expert surgeon being swamped by the poor results of several "occasional" surgeons.

Present policy suggests that we use audit to identify the "safe pair of hands." The practice of surgeons in

some districts has already been curtailed on the basis of samples of less than a year's work.[17] While audit can identify a learning curve for surgical performance,[18] the definition of an expert is rather fragile. If we assumed that the minimum clinically important attrition of survival for any given surgeon was 10%, then at least 150 cases would be required for this to be reliably detected, with 80% power and an α error (one sided) of 0.05.[19] The particular outcome event and how its risk varies over time has a bearing on the type of case mix adjustment required.[20] Clearly, audit focusing on the relation between volume and survival will, in most instances, be misplaced.

We can offer no explanation for the slightly worse outcome in the hospitals with medium volume workload. The relation is not a simple linear one. In a preliminary analysis we found it sensitive to the number of categories used for grouping. The grade of the operator was unknown to us but under emerging arrangements for clinical governance we think it justifiable to assign outcomes to the consultant in charge. We do not regard the level of findings from our hospital as conclusive, and they may reflect inadequate adjustment for unspecified case mix variables.

Whatever is responsible for the heterogeneity in outcome, such variation is of importance in multicentre trials. The potential benefits of treatments being studied may be swamped by such background variability.[21] For example, preoperative radiotherapy has been shown to reduce local recurrence by 20% in university hospitals and by 60% in non-university hospitals.[22]

Several lessons emerge from our study. Firstly, making survival data available to wider audiences to rank the performance of surgeons may be unwise. It will often not identify the real problems. Even for breast cancer the differences in survival of patients between surgeons with high caseload and low caseload is not related to the type of surgery.[23] Rather, the outcomes from colorectal cancer may depend on the clinical organisation of the service, the specialist surgeon being one trained to work effectively in a team and committed to ongoing quality assurance. Bringing together teams of surgeons, oncologists, pathologists, and palliative care specialists inevitably requires a commissioner to consider issues of economy of scale.

For the time being we must recognise that the volume-outcome debate may be a diversion. The rationale for providing multidisciplinary care and the need to enhance the power of clinical trials already provide compelling arguments for centralisation of cancer care.

Key messages

- Various benefits have been described for multidisciplinary cancer care, but the precise relation between a surgeon's or hospital's caseload and the outcome for the patient is not known
- Any investigation of a caseload effect at the hospital or practitioner level has to simultaneously account for each factor and adjust adequately for case mix
- Surgeon had no significant effect on caseload, but patients treated in hospitals with low caseloads (< 33 cases per year) had a slightly better survival at 2 years than those treated in hospitals with a higher caseload
- Defining surgical expertise in terms of volume of activity may be a misdirected and imprecise yardstick for the quality of cancer care; other aspects of the organisation of services may be far more important

Contributors: FK and RHW formulated the study hypotheses and along with KMcC prepared the first draft of the manuscript. RW oversaw core data collection and collation. CH was responsible for collating survival data. CCP performed all the statistical analyses. JMS, RJM, RFH, and BJR initiated the Northern Ireland colorectal cancer register, and all discussed and edited the final manuscript. FK and RHW will act as guarantors for the paper.

Funding: The Northern Ireland colorectal cancer registry is supported by the Friends of Montgomery House, an independent charity that supports work in the Northern Ireland Centre for Clinical Oncology at Belvoir Park Hospital.

Competing interests: None declared.

1 Expert Advisory Group on Cancer to the Chief Medical Officers of England and Wales. *Consultative document: a policy framework for commissioning cancer services*. London: Department of Health, 1994.
2 Royal College of Surgeons of England and the Association of ColoProcology of Great Britain and Ireland. *Guidelines for the management of colorectal cancer*. London: RCSE and the Association of ColoProcology of Great Britain and Ireland, 1996.
3 Department of Health. *Guidance for purchasers. improving outcomes in colorectal cancer. The manual*. London: DoH, 1997.
4 NHS Centre for Reviews and Dissemination, University of York. The management of colorectal cancer. *Effective Health Care* 1997;13:11-2.
5 Taylor I. Debate: is volume important for outcome following surgery for common solid tumours? *Eur J Surg Oncol* 1996;22:205-15.
6 Houghton A. Variation in outcome of surgical procedures. *Br J Surg* 1994;81:653-60.
7 Kee F, Wilson R, Currie SJ, Patterson CC, Houston R, Sloan J, et al. Socioeconomic circumstances and the risk of bowel cancer in Northern Ireland. *J Epidemiol Community Health* 1996;50:640-4.
8 Kee F, Wilson RH, Gilliland R, Sloan JM, Rowlands BJ, Moorehead RJ. Changing site distribution of colorectal cancer. *BMJ* 1992;305:158.
9 Christiansen C, Morris CN. Improving the statistical approach to health care provider profiling. *Ann Intern Med* 1997;127:764-8.
10 Goldstein H, Spiegelhalter D. League tables and their limitations: statistical issues in comparison of institutional performance. *J R Stat Soc (A)* 1996;159:1-25.
11 Goldstein H, Rasbash J, Plewis I, Draper D, Browne W, Yang M, et al. *A users guide to MLwinN*. London: Institute of Education, 1998.
12 McArdle CS, Hole D. Impact of variability among surgeons on post-operative morbidity and mortality and ultimate survival. *BMJ* 1991;302;1501-5.
13 Hakama M, Karjalainen S, Hakulinen T. Outcome based equity in the treatment of colon cancer patients in Finland. *Int J Tech Assess Health Care* 1989;5:619-30.
14 Luft H, Bunker JP, Enthoven AC. Should operations be regionalised? The empirical relation between surgical volume and mortality. *N Engl J Med* 1979;301;1364-9.
15 Hughes R, Hunt S, Luft H. Effects of surgeon volume on quality of care in hospitals. *Med Care* 1987;25:489-503.
16 Hannan E, O'Donnell JF, Kilburn H, Bernard H, Yazici A. Investigation of the relationship between volume and mortality for surgical procedures performed in New York state hospitals. *JAMA* 1989;262:503-10.
17 Pearce K. *Review of colorectal cancer services in a district health authority*. Presentation to the autumn conference of the Faculty of Public Health Medicine, November, 1997. (Abstract in press.)
18 Lovegrove J, Valencia O, Treasure T, Sherlaw-Johnson C, Gallivan S. Monitoring the results of cardiac surgery by variable life adjusted display. *Lancet* 1997;350:1128-30.
19 Machin D, Campbell MJ. *Statistical tables for the design of clinical trials*. Oxford: Blackwell Scientific, 1987.
20 Sagar P, Hartley M, MacFie J, Talyor BA, Copeland GP. Comparison of individual surgeon's performance. Risk adjusted analysis with POSSUM scoring system. *Dis Colon Rectum* 1996;39:654-8.
21 Fielding LP, Stewart-Brown S, Dudley HAF. Surgeon related variables and the clinical trial. *Lancet* 1978;ii:778-81.
22 Holm T, Johansson H, Cedermark B, Ekelund G, Rutqvist LE. Influence of hospital and surgeon related factors on outcome after treatment of rectal cancer with or without preoperative radiotherapy. *Br J Surg* 1997;84:657-63.
23 Sainsbury R, Haward B, Rider L, Johnston C, Round C. Influence of clinician workload and patterns of treatment on survival from breast cancer. *Lancet* 1995;345:1265-70.

(Accepted 21 January 1999)

Commentary: How experienced should a colorectal surgeon be?

Robert Shields

Department of Surgery, Royal Liverpool University Hospital, Liverpool L69 3GA
Robert Shields, *emeritus professor*

A question that a patient with colorectal cancer will naturally ask is, "Will my chances of being alive in two years' time be enhanced by my choice of surgeon?" Intuitively we feel that this is so, and that the patients of a more experienced surgeon will have less chance of dying or having complications after the operation and greater long term survival. If this is the case, there are wider implications in terms of training in the subspecialty of colorectal surgery and the concentration of treatment in larger centres with high volumes of patients. Such concentration of activity, however, has its downside because these centres may be more remote, making treatment less accessible and increasing the cost for patients and relatives.

Most, but not all, publications support a positive relation between increased activity (volume) and outcome for colorectal cancer. Not all studies, however, adjust for case mix or severity of illness, but Kee et al's paper attempts to do so.

The relation between the number of operations that a surgeon performs and the outcome is unlikely to be linear throughout. The threshold for experience to influence outcome (learning curve) must vary according to the operative procedure—lower for colonic cancer and higher for rectal cancer, where the surgical requirements are more demanding, to preserve anal sphincters in low lesions and to reduce the chances of local recurrence by mesorectal excision. The authors found no relation between survival and the site of the tumour.

The authors are the prisoners of their experimental design because inherent variability is such that there is little chance of showing statistical differences between the groups and subgroups that they analyse. For example, by focusing on mortality at 2 years the authors have combined the 30 day hospital mortality (which will be largely influenced by the type of surgery and postoperative complications—for example, anastomotic leakage) and longer term survival, which will probably be influenced by such factors as the biology of the tumour and the use of adjuvant treatment, of which radiotherapy, in the case of rectal cancer, is important. Combining surgical and non-surgical factors, which could influence mortality at 2 years, introduces too much variability into the analysis. The fact, however, that no statistical difference can be shown does not mean that a real difference does not exist.

No information is given on how many of the surgeons had received specialist training in colorectal surgery. Indeed, since in each case it was the name of the consultant surgeon in charge of the case that was recorded, we do not know if the operations were performed by consultants, by surgeons in training, or by both. This could be a factor of importance in emergency operations.

The experience of the surgeons was measured by the number of years on the medical register, surely a rather crude indicator of specialist experience. The number of years as a consultant surgeon would perhaps be a more accurate marker of post-training experience and could presumably be easily obtained in a circumscribed community such as that in Northern Ireland.

The worst survival in those hospitals that treated 33-54 cases a year defies explanation and must be a statistical oddity. The value of analysing hospital data is questionable. A separate study is required to determine if hospital volume can serve as a surrogate for surgeon volume for achieving good outcomes in colorectal cancer.

The problem with Kee et al's study is that, with the comparatively small numbers in the various groups and high variability, it becomes statistically impossible to show that surgeons with small caseloads are better or worse than those with higher caseloads. The results must not give the green light to those who wish to defend small volume workload or to encourage the surgeon who does colorectal surgery only occasionally.

The paper is interesting because it tackles an important subject and invites further study. Much depends on the resolution of the volume-outcome controversy, not only for the individual patient but also for surgical training and the provision of colorectal cancer services.

Corrections and clarifications

Double blind, cluster randomised trial of low dose supplementation with vitamin A or β carotene on mortality related to pregnancy in Nepal

This general practice paper by Keith P West Jr and colleagues (27 February, pp 570-5) contains some minor errors that do not affect the validity of the conclusions. The last paragraph of the results section (p 573) should have read: "The maternal mortality ratio was 630 [not 645] (42 deaths/6670 live births), 407 (29/7120 [not 29/7074]), and 346 [not 361] (23/6643) per 100 000 live births in the placebo, vitamin A, and β carotene groups, respectively (P = 0.11 [not 0.08] for vitamin A and 0.04 for β carotene *v* placebo). The ratio for women receiving either vitamin A or β carotene was 378 [not 385]." In the section of table 5 (p 574) labelled "infection" the values for the column for vitamin A should have been 2, 5, 3, 4, 14, 181, and 0.87 (0.39 to 1.96) [not 2, 5, 5, 3, 15, 194, and 0.94 (0.42 to 2.05)]; the values for the column for β carotene should have been 3, 3, 1, 3, 10, 139, and 0.67 (0.28 to 1.62) [not 3, 2, 1, 3, 9, 125, and 0.60 (0.24 to 1.51)]; the values for the column for vitamin A or β carotene should have been 5, 8, 4, 7, 24, 161, and 0.78 (0.39 to 1.58) [not 5, 7, 6, 6, 24, 161, and 0.78 (0.39 to 1.58)]; and the footnote for "Other" infections should have also included tuberculosis.