Open Access: free full text available online
For the full versions of these articles see bmj.com

Appendicitis, mesenteric lymphadenitis, and subsequent risk of ulcerative colitis: cohort studies in Sweden and Denmark

Morten Frisch,[1] Bo V Pedersen,[1] Roland E Andersson[2,3]

EDITORIAL by Timmer and Obermeier

[1]Department of Epidemiology Research, Division of Epidemiology, Statens Serum Institut, DK-2300 Copenhagen S, Denmark
[2]Department of Surgery, County Hospital Ryhov, SE-551 85 Jönköping, Sweden
[3]Department of Surgery, University Hospital, SE-581 85 Linköping, Sweden
Correspondence to: M Frisch mfr@ssi.dk

Cite this as: *BMJ* 2009;338:b716
doi:10.1136/bmj.b716

ABSTRACT

Objective To determine whether the repeatedly observed low risk of ulcerative colitis after appendicectomy is related to the appendicectomy itself or the underlying morbidity, notably appendicitis or mesenteric lymphadenitis.

Design Nationwide cohort studies.

Setting Sweden and Denmark.

Participants 709 353 Swedish (1964-2004) and Danish (1977-2004) patients who had undergone appendicectomy were followed up for subsequent ulcerative colitis. The impact of appendicectomy on risk was also studied in 224 483 people whose parents or siblings had inflammatory bowel disease.

Main outcome measures Standardised incidence ratios and rate ratios as measures of relative risk.

Results During 11.1 million years of follow-up in the appendicectomy cohort, 1192 patients developed ulcerative colitis (10.8 per 100 000 person years). Appendicectomy without underlying inflammation was not associated with reduced risk (standardised incidence ratio 1.04, 95% confidence interval 0.95 to 1.15). Before the age of 20, however, appendicectomy for appendicitis (0.45, 0.39 to 0.53) or mesenteric lymphadenitis (0.65, 0.46 to 0.90) was associated with significant risk reduction. A similar pattern was seen in those with affected relatives, whose overall risk of ulcerative colitis was clearly higher than the background risk (1404 observed *v* 446 expected; standardised incidence ratio 3.15, 2.99 to 3.32). In this cohort, appendicectomy without underlying appendicitis did not modify risk (rate ratio 1.04, 0.66 to 1.55, *v* no appendicectomy), while risk after appendicectomy for appendicitis was halved (0.49, 0.31 to 0.74).

Conclusions In individuals with or without a familial predisposition to inflammatory bowel disease, appendicitis and mesenteric lymphadenitis during childhood or adolescence are linked to a significantly reduced risk of ulcerative colitis in adulthood. Appendicectomy itself does not protect against ulcerative colitis.

INTRODUCTION

Many case-control studies have linked appendicectomy to a significantly reduced risk of ulcerative colitis.[1] Two cohort studies also provide support for an inverse association,[2 3] but relative risk estimates were less extreme than those reported in case-control studies. We combined national cohort data for all recorded appendicectomies in Sweden and Denmark up to 2004 to give a cohort of over 700 000 patients. We therefore had unprecedented statistical power to address the central question: is it appendicectomy itself or rather the underlying morbidity (notably, appendicitis or mesenteric lymphadenitis) that is responsible for the reduced incidence of ulcerative colitis in people with a history of appendicectomy?

METHODS

We used data from population based hospital discharge registries in Sweden and Denmark to identify patients who underwent appendicectomy and characterise them according to underlying diagnoses, sex, and age at the time of operation.

Appendicectomy cohort

We identified 446 968 patients in Sweden who underwent appendicectomy during the 41 year period 1964-2004. After exclusions the final cohort consisted of 443 761 patients (245 623 women and 198 138 men), representing 99.3% of all recorded appendicectomy patients in Sweden for 1964-2004.

We identified 273 099 patients in Denmark who underwent appendicectomy during the 28 year period 1977-2004. After exclusions the final cohort consisted of 265 592 patients (152 256 women and 113 336 men), representing 97.3% of all appendicectomy patients in Denmark for the period 1977-2004. See bmj.com for details of registries and coding in Sweden and Denmark.

Familial predisposition

We linked hospital admission data with family information in Statistics Sweden and the Danish Civil Registration System, continuously updated demographic databases. We identified 164 955 Swedes and 59 528 Danes for whom a mother, a father, or a sibling had a record of inflammatory bowel disease (ulcerative colitis, Crohn's disease, or both) at any time between 1964 and 2004 in Sweden or between 1977 and 2004 in Denmark.

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2009;338:b716

Ulcerative colitis outcomes

We identified ulcerative colitis outcomes in the hospital discharge registries. Because associations with appendicectomy differ considerably between ulcerative colitis and Crohn's disease,[2-4] we restricted our analysis to patients with unambiguous records of ulcerative colitis. We identified 31 577 first inpatient hospital contacts for ulcerative colitis in Sweden in 1964-2004 and 16 808 in Denmark in 1977-2004. Based on the distribution of sex, age, and calendar year for patients with ulcerative colitis and person time at risk in the underlying general population, we generated a set of incidence rates for ulcerative colitis for each country in strata of sex, age, and calendar year.

Statistical analysis

We used two measures for the relative risk of ulcerative colitis in the appendicectomy cohort and the cohort with familial predisposition to inflammatory bowel disease. We used the standardised incidence ratio to compare the incidence of ulcerative colitis in these cohorts with the incidence in the underlying general population (external comparison) and rate ratios to study differences in rates between cohort categories (internal comparison). The standardised incidence ratio was calculated as the ratio of the observed number of diagnoses of ulcerative colitis to the number expected based on background rates in the general population.

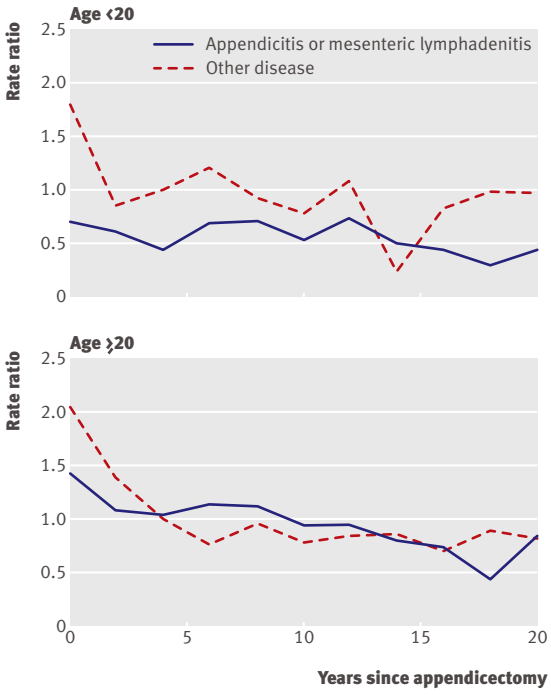
We performed multiple Poisson regressions to obtain rate ratios for comparisons between groups of appendicectomy cohort members with different profiles of explanatory variables and confounders. We also used multiple Poisson regression to obtain rate ratios for the evaluation of the impact of appendicectomy and the disease underlying the appendicectomy on risk in the cohort with familial predisposition to inflammatory bowel disease. See bmj.com.

RESULTS

The Swedish and Danish cohorts comprised 709 353 patients who had undergone appendicectomy. The combined cohort was followed for the occurrence of ulcerative colitis for 11.1 million person years after appendicectomy. Overall, with a total of 1192 cases occurring in the two countries, the crude rate was 10.8/100 000 person years. The incidence of ulcerative colitis in people who had undergone appendicectomy was 12% lower in Sweden (standardised incidence ratio 0.88, 95% confidence interval 0.82 to 0.94) and 11% lower in Denmark (0.89, 0.80 to 0.99), with no significant difference between the two countries (P=0.76). We present all subsequent results for the combined cohort.

Standardised incidence ratios of ulcerative colitis after appendicectomy

By age at appendicectomy—Appendicectomies performed in childhood or adolescence were associated with almost 50% reduction in incidence of ulcerative colitis (0.51, 0.37 to 0.70, and 0.54, 0.47 to 0.62, for appendicectomies before the age of 10 and from age 10 to 19, respectively). Standardised incidence ratios gradually increased with age at appendicectomy (P<0.001) with no indication of a reduced incidence of ulcerative colitis in cohort members aged 30 or more at appendicectomy.



Rate ratios for ulcerative colitis by time since appendicectomy in two year intervals according to underlying disease in patients aged <20 (P<0.001) or ≥20 (P=0.48) at appendicectomy. Rate ratios adjusted for country (Sweden *v* Denmark), sex, attained age (10 year intervals), calendar period (<1970, 1970-9, 1980-9, 1990-9, 2000-4), and age at appendicectomy. Reference rate was rate of ulcerative colitis four to five years after appendicectomy among patients with other disease. P values reflect significance of effect of underlying disease (appendicitis or mesenteric lymphadenitis *v* other disease) obtained in slightly reduced regression models with similar effect of time since appendicectomy in compared groups

By time since appendicectomy—Standardised incidence ratios depended significantly on the time interval since the appendicectomy (P<0.001). The risk of ulcerative colitis was higher in the first five years after appendicectomy—notably, in the first six months after the operation (2.04, 1.63 to 2.56)—whereas it was consistently reduced in time intervals 10 or more years after appendicectomy.

By underlying disease—The incidence ratio also varied considerably according to the underlying disease leading to the appendicectomy. About 68% of cohort members who had surgery because of appendicitis had significantly reduced incidence (0.83, 0.77 to 0.89). Likewise, cohort members with a diagnosis of mesenteric lymphadenitis were at reduced risk (0.65, 0.49 to 0.86). In contrast, those who had surgery for reasons other than appendicitis or mesenteric lymphadenitis experienced no unusual risk (1.04, 0.95 to 1.15).

Rate ratios of ulcerative colitis by age <20 *v* ≥20 at appendicectomy

Details of Poisson regression analysis and rate ratios for ulcerative colitis by age at appendicectomy and time since appendicectomy are on bmj.com. We used Poisson regression to evaluate the role of the underlying disease in subsets of young (<20 years) and older (≥20) patients (figure). Among the younger patients, risk was significantly lower in those with appendicitis or mesenteric lymphadenitis compared with those with other disease (P<0.001). In contrast, among older patients, there was no significant difference in rates between those with underlying appendicitis or mesenteric lymphadenitis and those without (P=0.48).

Role of appendicectomy in people with familial predisposition

We also studied risk among 224 483 individuals whose parents or siblings had inflammatory bowel disease (table). The number of cases during 4.2 million person years of follow-up exceeded the expected number based on rates in the underlying general populations of Sweden and Denmark (1404 observed *v* 446 expected; standardised incidence ratio 3.15, 2.99 to 3.32). Poisson regression analysis controlled for age, calendar period, and type of relative showed no difference between rates in relatives who retained their appendix intact (reference group) and relatives who underwent appendicectomy without appendicitis. In contrast, risk in relatives who underwent appendicectomy for appendicitis was significantly reduced, a pattern seen whether the affected relative was a parent or a sibling. By introducing interaction terms in the Poisson regression models between appendicectomy status on one side and country (P=0.67) or sex (P=0.16) on the other, we found that the observed risk reduction associated with appendicectomy for appendicitis (*v* no appendicectomy) was consistent for Swedish (0.45, 0.25 to 0.73) and Danish (0.63, 0.27 to 1.24) relatives and identical for female (0.49, 0.25 to 0.87) and male (0.49, 0.26 to 0.85) relatives of patients with inflammatory bowel disease.

DISCUSSION

Key findings

We observed significantly fewer subsequent diagnoses of ulcerative colitis in patients who had undergone appendicectomy, the association being restricted to appendicectomies for appendicitis or mesenteric lymphadenitis before the age of 20. Others have observed this age restriction,[5-7] but we also showed that, without appendicitis or mesenteric lymphadenitis, appendicectomy has no impact on subsequent risk, even when done in childhood or adolescence. The analysis of risk in individuals with a familial predisposition to inflammatory bowel disease corroborates this view.

Possible mechanisms linking appendicitis in childhood to low risk of ulcerative colitis

One mechanism could be that inflammatory responses elicited during the course of childhood appendicitis or mesenteric lymphadenitis somehow induce longlasting immunological changes in the colonic mucosa, which eventually protect these individuals from developing ulcerative colitis (beneficial inflammation hypothesis).

A second mechanism could be that an, as yet, uncharacterised genetic trait that confers protection against ulcerative colitis might be closely linked to susceptibility genes for appendicitis or, in reverse, susceptibility genes for ulcerative colitis might be closely linked to an unknown genetic trait that protects against appendicitis (linkage disequilibrium hypothesis).

Thirdly, an aetiological mechanism might involve an environmental or microbial factor associated with increased risk of appendicitis and reduced risk of ulcerative colitis or the reverse, a factor associated with protection against appendicitis and increased risk (antagonistic risk factor hypothesis).

Finally, constitutional factors associated with a preference for immune responses orchestrated by Th1 or Th2 cells might differ between patients developing appendicitis and those developing ulcerative colitis (constitutional immunity hypothesis). See bmj.com.

Strengths of the study

The prospective nature of our historical cohort analyses eliminated potential information and selection biases and other methodological problems that are often encountered in case-control studies. Other strengths include the size of our cohort and the truly population based data sources we used to

Rate ratios* (RR) with 95% confidence intervals for ulcerative colitis according to appendicectomy status in cohort of 224 483 people with familial predisposition to inflammatory bowel disease, Sweden (1964-2004) and Denmark (1977-2004). Figures are numbers of cases of ulcerative colitis per person years at risk

Relative with inflammatory bowel disease	No appendicectomy		Appendicectomy, no appendicitis		Appendicectomy + appendicitis	
	Cases/person years	RR (95% CI)	Cases/person years	RR (95% CI)	Cases/person years	RR (95% CI)
Parent or sibling†	1361/4 047 047	1‡	22/52 608	1.04 (0.66 to 1.55)	21/102 211	0.49 (0.31 to 0.74)
Parent	854/2 803 351	1‡	12/30 195	1.01 (0.54 to 1.71)	11/61 385	0.43 (0.22 to 0.75)
Sibling	551/1 281 094	1‡	10/23 118	1.01 (0.51 to 1.79)	10/42 050	0.54 (0.27 to 0.96)

*Adjusted for age, calendar period (both in one year intervals with restricted cubic splines), and type of relative with inflammatory bowel disease (parent, sibling, or both).
†Some cases of ulcerative colitis (n=44) occurred in patients with familial predisposition to inflammatory bowel disease through both parent and sibling.
‡Reference category.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Appendicectomy has been associated with low risk of ulcerative colitis, but the reason for this inverse relation remains controversial

Appendicectomy has been suggested as a possible prophylactic procedure in individuals with a predisposition to ulcerative colitis

WHAT THIS STUDY ADDS

Appendicectomy for appendicitis or mesenteric lymphadenitis in childhood or adolescence, but not after the age of 20, is linked to a reduced risk of ulcerative colitis

Appendicectomy itself does not protect against the development of ulcerative colitis

characterise cohort members and identify ulcerative colitis outcomes.

In addition to presenting standardised incidence ratios we also compared rates between strata of the appendicectomy cohort by means of Poisson regression, which showed that both young age and appendicitis or mesenteric lymphadenitis had to be present to confer a low risk of ulcerative colitis.

Limitations of the study

We relied on routinely collected data from health and administrative registers, which are not primarily set up for research purposes. Also hospital discharge data account only for those treated as inpatients. Like most previous studies we were unable to adjust for smoking, the only behavioural factor that has been linked consistently to risk. Major confounding, however, is unlikely because a strong inverse association between appendicectomy and risk was also observed in previous studies that took smoking into account.[8 9] Also, in light of the considerably higher prevalence of smokers in Denmark than in Sweden (35% *v* 19% among men and 43% *v* 25% among women aged ≥25 years), it is reassuring that we obtained almost identical results in country specific analyses.

Clinical implications

Our findings should put an end to speculations about possible prophylactic capabilities of appendicectomy.

Contributors: See bmj.com.
Funding: The study was supported by unrestricted research grants from Aase and Ejnar Danielsen's Foundation, Civil engineer Frode V. Nyegaard and Wife's Foundation, and the Gangsted Foundation. The funders had no role in the data analysis or interpretation of the data.
Competing interests: None declared.
Ethical approval: The study involved no patient contact. The National Board of Health and Welfare in Sweden delivered anonymised data from the Swedish Hospital Discharge Registry, and the National Board of Health in Denmark provided corresponding Danish data for the study. All analyses were carried out in Denmark (Danish Data Inspection Board approval No 2001-41-0576).

1 Koutroubakis IE, Vlachonikolis IG, Kouroumalis EA. Role of appendicitis and appendectomy in the pathogenesis of ulcerative colitis: a critical review. *Inflamm Bowel Dis* 2002;8:277-86.
2 Andersson RE, Olaison G, Tysk C, Ekbom A. Appendectomy and protection against ulcerative colitis. *N Engl J Med* 2001;344:808-14.
3 Frisch M, Johansen C, Mellemkjær L, Engels EA, Gridley G, Biggar RJ, et al. Appendectomy and subsequent risk of inflammatory bowel diseases. *Surgery* 2001;130:36-43.
4 Andersson RE, Olaison G, Tysk C, Ekbom A. Appendectomy is followed by increased risk of Crohn's disease. *Gastroenterology* 2003;124:40-6.
5 Kurina LM, Goldacre MJ, Yeates D, Seagroatt V. Appendicectomy, tonsillectomy, and inflammatory bowel disease: a case-control record linkage study. *J Epidemiol Community Health* 2002;56:551-4.
6 Feeney MA, Murphy F, Clegg AJ, Trebble TM, Sharer NM, Snook JA. A case-control study of childhood environmental risk factors for the development of inflammatory bowel disease. *Eur J Gastroenterol Hepatol* 2002;14:529-34.
7 Baron S, Turck D, Leplat C, Merle V, Gower-Rousseau C, Marti R, et al. Environmental risk factors in paediatric inflammatory bowel diseases: a population based case control study. *Gut* 2005;54:357-63.
8 Uzan A, Jolly D, Berger E, Diebold MD, Geoffroy P, Renard P, et al. Effet protecteur de l'appendicectomie contre la rectocolite hémorragique. Etude cas-témoins [Protective effect of appendectomy on the development of ulcerative colitis. A case-control study]. *Gastroenterol Clin Biol* 2001;25:239-42.
9 Russel MG, Dorant E, Brummer RJ, van de Kruijs MA, Muris JW, Bergers JM, et al. Appendectomy and the risk of developing ulcerative colitis or Crohn's disease: results of a large case-control study. South Limburg Inflammatory Bowel Disease Study Group. *Gastroenterology* 1997;113:377-82.

Accepted: 26 November 2008

Predicting risk of type 2 diabetes in England and Wales: prospective derivation and validation of QDScore

Julia Hippisley-Cox,[1] Carol Coupland,[1] John Robson,[2] Aziz Sheikh,[3] Peter Brindle[4]

EDITORIAL by Schwarz and colleagues

[1]Division of Primary Care, Tower Building, University Park, Nottingham NG2 7RD
[2]Centre for Health Sciences, Queen Mary's School of Medicine and Dentistry, London E1 2AT
[3]Centre for Population Health Sciences: GP Section, University of Edinburgh, Edinburgh EH8 9DX
[4]Avon Primary Care Research Collaborative, Bristol Primary Care Trust, Bristol BS2 8EE
Correspondence to:
J Hippisley-Cox
Julia.hippisley-cox@nottingham.ac.uk

Cite this as: *BMJ* 2009;338:b880
doi:10.1136/bmj.b880

ABSTRACT

Objective To develop and validate a new diabetes risk algorithm (the QDScore) for estimating 10 year risk of acquiring diagnosed type 2 diabetes over a 10 year time period in an ethnically and socioeconomically diverse population.

Design Prospective open cohort study using routinely collected data from 355 general practices in England and Wales to develop the score and from 176 separate practices to validate the score.

Participants 2 540 753 patients aged 25-79 in the derivation cohort, who contributed 16 436 135 person years of observation and of whom 78 081 had an incident diagnosis of type 2 diabetes; 1 232 832 patients (7 643 037 person years) in the validation cohort, with 37 535 incident cases of type 2 diabetes.

Outcome measures A Cox proportional hazards model was used to estimate effects of risk factors in the derivation cohort and to derive a risk equation in men and women. The predictive variables examined and included in the final model were self assigned ethnicity, age, sex, body mass index, smoking status, family history of diabetes, Townsend deprivation score, treated hypertension, cardiovascular disease, and current use of corticosteroids; the outcome of interest was incident diabetes recorded in general practice records. Measures of calibration and discrimination were calculated in the validation cohort.

Results A fourfold to fivefold variation in risk of type 2 diabetes existed between different ethnic groups. Compared with the white reference group, the adjusted hazard ratio was 4.07 (95% confidence interval 3.24 to 5.11) for Bangladeshi women, 4.53 (3.67 to 5.59) for Bangladeshi men, 2.15 (1.84 to 2.52) for Pakistani women, and 2.54 (2.20 to 2.93) for Pakistani men. Pakistani and Bangladeshi men had significantly higher hazard ratios than Indian men. Black African men and Chinese women had an increased risk compared with the corresponding white reference group. In the validation dataset, the model explained 51.53% (95% confidence interval 50.90 to 52.16) of the variation in women and 48.16% (47.52 to 48.80) of that in men. The risk score showed good discrimination, with a D statistic of 2.11 (95% confidence interval 2.08 to 2.14) in women and 1.97 (1.95 to 2.00) in men. The model was well calibrated.

Conclusions The QDScore is the first risk prediction algorithm to estimate the 10 year risk of diabetes on the basis of a prospective cohort study and including both social deprivation and ethnicity. The algorithm does not need laboratory tests and can be used in clinical settings and also by the public through a simple web calculator (www.qdscore.org).

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2009;338:b880

INTRODUCTION

Evidence from randomised controlled trials shows that behavioural or pharmacological interventions can prevent type 2 diabetes in up to two thirds of high risk cases.[1-4] Cost effectiveness modelling suggests that screening programmes aid earlier diagnosis and help to prevent type 2 diabetes or improve outcomes in people who develop the condition.[5 6] Early detection is important, as up to half of people with newly diagnosed type 2 diabetes have one or more complications at the time of diagnosis.[7]

No widely accepted diabetes risk prediction score has been developed and validated for use in routine clinical practice. Previous studies have been limited by size,[8] and some have performed inadequately when tested in ethnically diverse populations.[9] A diabetes risk prediction tool with weightings for both social deprivation and ethnicity is needed given the prevalence of type 2 diabetes, particularly among minority ethnic communities.[10]

We present the derivation and validation of a new risk prediction algorithm for assessing the risk of developing type 2 diabetes with appropriate weightings for ethnicity and social deprivation. We based the algorithm (the QDScore) on variables that are readily available in electronic health records, enabling it to be readily and cost effectively implemented in routine clinical practice.

METHODS

Study design and data source

We did a prospective cohort study in a large population of primary care patients from the QResearch database. This is a large, validated primary care electronic database containing the health records of 11 million patients registered with 551 general practices. Practices and patients are nationally representative for England and Wales.[11]

Practice selection—We included all practices in England and Wales who had been using the Egton Medical Information System (EMIS) computer system for at least a year. We randomly allocated two thirds of practices to the derivation dataset and the remaining third to the validation dataset.

Cohort selection—We identified an open cohort of patients aged 25-79 years drawn from patients registered with practices between 1 January 1993 and 31 March 2008. We excluded patients with a prior recorded diagnosis of diabetes, and those who did not have a postcode related Townsend deprivation score (about 4% of the population). Censor points were diagnosis of type 2 diabetes, death, deregistration with the practice, last upload of computerised data, or the study end date.

Adjusted hazard ratios (95% confidence interval) for QDScore in derivation cohort (see fig 1 in full version for graphical representation of interaction terms)

	Women	Men
White/not recorded	1	1
Indian	1.710 (1.488 to 1.965)	1.929 (1.700 to 2.189)
Pakistani	2.152 (1.839 to 2.517)	2.538 (2.202 to 2.925)
Bangladeshi	4.071(3.242 to 5.112)	4.532 (3.673 to 5.591)
Other Asian	1.264 (0.943 to 1.695)	1.894 (1.492 to 2.404)
Black Caribbean	0.798 (0.695 to 0.915)	0.955 (0.824 to 1.108)
Black African	0.805 (0.661 to 0.979)	1.695 (1.421 to 2.023)
Chinese	1.961 (1.385 to 2.777)	1.414 (0.928 to 2.154)
Other	0.889 (0.738 to 1.07)	1.199 (1.005 to 1.431)
Townsend score (per increase of 1 SD)	1.201 (1.188 to 1.214)	1.140 (1.129 to 1.152)
Family history of diabetes in a first degree relative	2.358 (2.278 to 2.441)	2.725 (2.638 to 2.815)
Current smoker	1.268 (1.225 to 1.312)	1.249 (1.214 to 1.285)
Treated hypertension	1.787 (1.738 to 1.837)	1.711 (1.665 to 1.759)
Diagnosis of cardiovascular disease	1.458 (1.402 to 1.517)	1.500 (1.455 to 1.546)
Current treatment with corticosteroids	1.412 (1.339 to 1.489)	1.259 (1.181 to 1.342)

Model also included fractional polynomial terms for age and body mass index and interactions between age terms and body mass index terms, age terms and family history of diabetes, and age terms and smoking status (see fig 1 in full version).

Primary outcomes

Our primary outcome measure was the first diagnosis of type 2 diabetes mellitus as recorded on the general practice computer records.

Diabetes risk factors

We examined the following variables for inclusion in our analysis, all of which are known or thought to affect risk of developing diabetes [8 12-18]: self assigned ethnicity; age; body mass index; smoking status; Townsend deprivation score; history of diabetes in a first degree relative; cardiovascular disease at baseline; treated hypertension at baseline; systemic corticosteroids at baseline. For body mass index and smoking status, we used the values recorded closest to the study entry date.

Model derivation and development

We used a Cox proportional hazards model to estimate the coefficients and hazard ratios associated with each potential risk factor for the first ever recorded diagnosis of diabetes for men and women separately. We tested for interactions between each variable and age and between smoking and deprivation and included significant interactions in the final model. We used multiple imputation to replace missing values for smoking status and body mass index.

We took the regression coefficient for each variable from the final model and used these as weights for the new disease risk equations for type 2 diabetes. We combined these weights with the baseline survivor function for diagnosis of diabetes evaluated at 10 years to derive a risk equation for 10 years' follow-up.

We compared our final model with three other models in order to determine the additional contribution to the fit and performance of the model of including both ethnicity and deprivation in the algorithm.

Validation of the QDScore

We tested the performance of the final algorithm (the QDScore) in the validation dataset. We calculated the mean predicted risk and the observed risk of diabetes at 10 years and compared these by 10th of predicted risk. The observed risk at 10 years was obtained by using the 10 year Kaplan-Meier estimate. We calculated the Brier score (a measure of goodness of fit),[19] D statistic (a measure of discrimination),[20] and an R^2 statistic (a measure of explained variation for survival data).[21] We also calculated the area under the receiver operator curve. We compared the performance of the QDScore with the Cambridge risk score.[8] We calculated the proportion of patients in the validation sample who had an estimated 10 year risk of diagnosed diabetes of ≥10%, ≥15%, ≥20%, ≥30%, ≥40%, and ≥50% by age, sex, ethnic group, and deprivation according to the QDScore.

RESULTS

Description of the derivation and validation dataset

Overall, 531 UK practices met our inclusion criteria, of which 355 were randomly assigned to the derivation dataset and 176 to the validation dataset. We excluded 20 practices from Scotland and Northern Ireland or with incompletely uploaded data.

The derivation cohort contained 2 594 578 patients, of whom 53 825 had type 1 or type 2 diabetes before the start of the study and were excluded leaving 2 540 753 patients (1 283 135; 50.50% women) aged 25-79 years. The validation cohort contained 1 261 419 patients aged 25-79, of whom 28 587 had a previous diagnosis of type 1 or type 2 diabetes leaving 1 232 832 patients (50.49% women).

Overall, we studied 3 773 585 patients contributing 24 079 172 person years, of whom 115 616 patients (78 081 in the derivation cohort and 37 535 in the validation cohort) had a new diagnosis of type 2 diabetes during follow-up. The baseline characteristics of the validation cohort were very similar to those for the derivation cohort.

Patterns of missing data

Overall, 22.97% of women and 29.88% of men had either smoking or body mass index imputed by multiple imputation. Similar figures were observed for men and women in the validation cohort, where multiple imputation was also used.

Incidence of diabetes

The age standardised rates of type 2 diabetes for the white reference group were 4.13 (95% confidence interval 4.08 to 4.17) per 1000 person years for women and 5.31 (5.26 to 5.36) per 1000 person years for men. Age standardised rates were higher for men and women in every ethnic group compared with the white reference group, except for Chinese men. The highest age standardised rates were in South Asians, and significant differences existed between the South Asian groups. The rate for Bangladeshi women was 18.20 (12.93 to 23.47) per 1000 person years and that

for Bangladeshi men was 19.34 (14.28 to 24.4) per 1000 person years. For Pakistanis, the corresponding rates per 1000 person years were 11.19 (9.16 to 13.21) for women and 13.22 (11.24 to 15.21) for men.

We also found a marked difference in the age standardised incidence rates of type 2 diabetes by deprivation, with a more than twofold difference for women when comparing the most deprived fifth (6.39 (6.25 to 6.54) per 1000 person years) with the most affluent fifth (3.00 (2.93 to 3.08) per 1000 person years). A similar, but less steep gradient was seen for men.

Prevalence of risk factors by ethnicity

Substantial differences were evident in the age standardised prevalence of smoking among men of Bangladeshi (46.04%, 95% confidence interval 43.16% to 48.92%), Caribbean (40.45%, 38.99% to 41.91%), Pakistani (32.82%, 31.29% to 34.35%), white/not recorded (33.49%, 33.40% to 33.58%), Chinese (26.63%, 24.23% to 29.03%), Indian (22.71%, 21.60% to 23.81%), and black African (17.95%, 16.76% to 19.14%) origin. Smoking rates were lower for women in each ethnic group compared with men but varied widely between women from different groups.

Treated hypertension was highest among black Caribbean and black African men and women and more than twice as high as that for the white reference group. Recorded family history of diabetes was highest among black Caribbean women (32.63%, 31.41% to 33.85%) and Indian men (29.95%, 28.78% to 31.11%), which was more than three times that for the white reference group who had the lowest rates (11.32%, 11.27% to 11.38% for women and 8.07%, 8.02% to 8.12% for men).

Bangladeshi men and women had the highest age standardised mean deprivation scores, followed by those of black African and black Caribbean origin. Indians and the white reference group had the lowest mean deprivation scores.

The highest mean body mass index was seen among black African women (age standardised mean 28.44, 28.29 to 28.58) compared with 25.47 (25.46 to 25.48) for women in the white reference group. The lowest value was in Chinese women (age standardised mean 22.87, 22.68 to 23.06). Similar patterns were seen for men across the ethnic groups. Finally, 9.70% (7.76% to 11.65%) of Bangladeshi men had a recorded diagnosis of cardiovascular disease at baseline, which was more than twice that for men in the white reference group (4.54%, 4.50% to 4.57%) and more than four times that found in Chinese men (2.26%, 1.15% to 3.37%).

Model development

The table shows the results of the Cox regression analysis for the QDScore. After adjustment for all other variables in the model, we found significant associations with risk of type 2 diabetes in both men and women for age, body mass index, family history of diabetes, smoking status, treated hypertension, use of corticosteroids, diagnosed cardiovascular disease, social deprivation, and ethnicity. We therefore included these variables in the final model and risk prediction algorithm.

We found significant heterogeneity of risk of type 2 diabetes by ethnic group compared with the white reference population, having adjusted for age, body mass index, deprivation, family history of diabetes, smoking status, treated hypertension, diagnosed cardiovascular disease, use of corticosteroids, and diagnosed cardiovascular disease (see bmj.com). For example, among Bangladeshis, the adjusted hazard ratio for women was 4.07 (95% confidence interval 3.24 to 5.11) and that for men was 4.53 (3.67 to 5.59). These were significantly higher than the increased hazard ratios in Pakistani women and men (2.15, 1.84 to 2.52; and 2.54, 2.20 to 2.93). Both Pakistani and Bangladeshi men had significantly higher hazard ratios than Indian men. Black African men and Chinese women had increased risks compared with the corresponding white reference group. The only groups to have significantly lower risks than the white reference group were black African women (0.81, 0.66 to 0.98) and black Caribbean women (0.80, 0.70 to 0.92).

We identified significant interactions between age and body mass index, age and family history of diabetes, and age and smoking status. We included these interactions in the final model, and the general direction of the effects was that body mass index and family history of diabetes tended to have a greater impact on risk of diabetes at younger ages. Smoking had a more complex relation with age; the risk peaked in middle age for both men and women.

Calibration and discrimination of QDScore

Validation statistics in the validation dataset show higher levels of discrimination in the QDScore than the Cambridge risk score (see bmj.com). The QDScore also explained a higher proportion of the variation—it explained 51.53% of the variation in women and 48.16% in men. The corresponding values for the Cambridge risk score were 45.77% and 41.82%. The Brier score, however, was slightly lower for the Cambridge risk score in both men and women.

Comparison of the mean predicted scores from the QDScore with observed risks show close correspondence between predicted and observed 10 year risks within each model 10th, suggesting that the model was well calibrated.

Predictions with age, sex, deprivation, and ethnicity

At the 10% threshold, 10.60% of women and 15.06% of men had a 10% or higher predicted risk of being diagnosed as having type 2 diabetes over 10 years. This varied markedly by age such that 21.43% of women aged 55-59 and 30.99% of women aged 65-69 had a 10% or greater risk of being diagnosed as having type 2 diabetes over 10 years. The corresponding figures for men were 33.28% and 44.08%.

Across ethnic group and deprivation, 33.83% of Bangladeshi women had a 10 year risk of being diagnosed as having diabetes of 10% or more compared with 10.48% of women in the white reference group,

and 15.03% of women in the most deprived fifth had a 10% or higher risk of developing diabetes over the next 10 years compared with 6.52% of women in the most affluent fifth. The difference between affluent and deprived fifths was more marked for women than for men; the corresponding figures were 15.65% for men in the most deprived fifth and 13.21% for men in the most affluent fifth. Overall, almost half (15 545/32 450; 47.9%) of cases of diabetes occurred in the top 10th of the distribution (risk of ≥10.38%) and almost 70% (22 476/32 450) occurred in the top fifth (risk of ≥5.98%).

DISCUSSION

The QDScore is the first diabetes prediction algorithm developed and validated by using routinely collected data to predict the 10 year risk of developing type 2 diabetes. Our final model includes both deprivation and ethnicity as well as age, sex, smoking, treated hypertension, body mass index, family history of diabetes, current treatment with corticosteroids, and previous diagnosis of cardiovascular disease. The QDScore does not require any laboratory testing or clinical measurements and so can be used in many settings.

Simple clinical models using readily available data can offer similar discrimination to more complex models using laboratory data or biomarkers,[12] and may have a further utility in settings where clinical measurements are not available or are too costly.[22] UK datasets derived from family practices have the advantage of having large and representative populations with historical data on many of the key variables known to be associated with risk of type 2 diabetes tracking back over a decade

Strengths and weaknesses

Sampling and generalisability

Particular strengths of our study are the use of a large representative population from a validated database, our prospective cohort design, and the substantial numbers of patients in the analysis. We have modelled interactions with age and included these in the final model.

Another important strength of the QDScore is that all the variables used in the algorithm will either be known to an individual patient or are collected as part of routine clinical practice. This means that the algorithm can be used by patients for self assessment in a web based calculator (www.qdscore.org). Alternatively, it can be implemented within clinical computer systems to stratify the practice population for risk on a continuing basis without the need for manual entry of data.

Potential sources of misclassification, bias, and confounding

One limitation of our study is that the main outcome of type 2 diabetes was not formally validated. However, other studies of similar databases have shown good levels of accuracy for common chronic conditions.[23] Undiagnosed diabetes is a well recognised problem and is not specifically considered by our study.

Our study might have been affected by recording bias if a patient diagnosed as having diabetes was not recorded as having diabetes on the practice computer system. Any misclassification bias of the outcome, if non-differential, would tend to bias the hazard ratio towards one and reduce discrimination.

Our study might have been affected by an ascertainment bias caused by differential testing of patients for diabetes by ethnic group or in those with specific risk factors. This could lead to increased rates of detection among patients with specific risk factors. Our hazard ratios for the risk factors in the model are generally of a similar magnitude to those found in other studies which tested for diabetes in the entire study cohort.[24]

Validation of risk prediction algorithm

We validated the QDScore in a separate sample of general practices from those used to develop the score. The QDScore has good discrimination and explains approximately 50% of the total variation in times to diagnosis of diabetes. An important limitation of our validation is that the general practices used for the validation use the same clinical computer system (EMIS) as those used to derive the algorithm. Nonetheless, our previous algorithm for cardiovascular disease, developed with similar methods and the same database,[25] has subsequently performed well on another database containing primary care data.[11]

Comparison with other diabetes risk scores

Routinely collected data from electronic primary healthcare records have been used to develop other risk prediction algorithms. For example, data from 531 general practices was used to develop and validate the QRISK2 cardiovascular disease risk tool, which is being implemented in clinical settings in the UK.[25,26] The Cambridge diabetes risk score was developed by combining data from patients from one general practice in Cambridge

WHAT IS ALREADY KNOWN ON THIS TOPIC

Good evidence shows that behavioural or pharmacological interventions can prevent type 2 diabetes in up to two thirds of patients at high risk and that early diagnosis is likely to improve outcomes

In 2009 the Department of Health will start a major vascular screening programme, which includes identification and management of patients at high risk of diabetes for preventive care

No widely accepted and validated risk prediction score takes account of both social deprivation and ethnicity and can be applied in primary care in the UK

WHAT THIS STUDY ADDS

The QDScore is a new risk prediction algorithm for type 2 diabetes developed in a very large and unselected family practice derived population, with appropriate weightings for ethnicity and social deprivation

The final algorithm includes self assigned ethnicity, age, sex, body mass index, smoking status, family history of diabetes, Townsend deprivation score, treated hypertension, cardiovascular disease, and current use of corticosteroids

The performance of the QDScore in an independent sample of practices showed good discrimination and calibration

with half of the cases of incident diabetes from another 41 practices in a different region of England.

One advantage of the QDScore is the use of a larger and more representative cohort. Another advantage is the inclusion of both deprivation and self assigned ethnicity, which are independently associated with risk of incident diabetes; this is likely to help with the problems identified with the Cambridge risk score in its performance in ethnically diverse populations.[9] The QDScore explained significantly more of the variation and had improved discrimination compared with the Cambridge risk score.

Other diabetes scores have been developed within specific ethnic groups, such as Mexican or Japanese Americans, but we have too few patients in the UK in these ethnic groups to allow a meaningful comparison to be made within this analysis. Nonetheless, our receiver operator curve statistic of 0.85 for women and 0.83 for men is substantially higher than those in many studies.[8 16 17 27 28]

Conclusions

This algorithm to predict risk of type 2 diabetes has the unique advantage of including both ethnicity and social deprivation, can be derived without laboratory measurements, and thus is suitable for use both in clinical settings and for self assessment. The QDScore could be used to identify patients at high risk of diabetes who might benefit from interventions to reduce their risk.

We acknowledge the contribution of Egton Medical Information System (EMIS) and practices using EMIS and contributing to the QResearch database.

Contributors: See bmj.com.

Funding: This study received no external funding. The authors did the work either in their personal time or during the course of their normal employment. The corresponding author (JH-C) and CC had access to all the data in the study, and all authors agreed and share responsibility for the decision to submit for publication.

Competing interests: JH-C is co-director of QResearch, a not for profit organisation, which is a joint partnership between the University of Nottingham and EMIS. JH-C is also director of ClinRisk Ltd, which produces software to ensure the reliable and updatable implementation of clinical risk algorithms within clinical computer systems to help to improve patients' care. EMIS is the leading supplier of information technology for 60% of UK general practices and may implement the QDScore within its clinical computer system. AS chairs the Equality and Diversity Forum of the National Clinical Assessment Service and is co-investigator on an MRC/NPRI funded randomised controlled trial aiming to prevent onset of type 2 diabetes in South Asians in the UK; he is also a co-investigator on the MRC Edinburgh Translational Medicine Methodology Hub. QResearch does analyses for the Department of Health and other government organisations. All research using QResearch is peer reviewed and published. This work and any views expressed within it are solely those of the co-authors and not of any affiliated bodies or organisations.

Ethical approval: The proposal was approved by the Trent Multi Centre Research Ethics Committee.

1 Diabetes Prevention Program Research Group. Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. *N Engl J Med* 2002;346:393-403.
2 Tuomilehto J, Lindstrom J, Eriksson JG, Valle TT, Hamalainen H, Ilanne-Parikka P, et al. Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. *N Engl J Med* 2001;344:1343-50.
3 Chiasson J-L, Josse RG, Gomis R, Hanefeld M, Karasik A, Laakso M. Acarbose for prevention of type 2 diabetes mellitus: the STOP-NIDDM randomised trial. *Lancet* 2002;359:2072.
4 Ramachandran A, Snehlatha C, Mary S, Mukesh B, Bhaskar A, Vijay V. The Indian diabetes prevention programme shows that lifestyle modification and metformin prevent type 2 diabetes in Asian Indian subjects with impaired glucose tolerance (IDPP-1). *Diabetologia* 2006;49:289-97.
5 Gillies CL, Lambert PC, Abrams KR, Sutton AJ, Cooper NJ, Hsu RT, et al. Different strategies for screening and prevention of type 2 diabetes in adults: cost effectiveness analysis. *BMJ* 2008;336:1180-5.
6 Waugh N, Scotland G, McNamee P, Gillett M, Brennan A, Goyder E, et al. Screening for type 2 diabetes: literature review and economic modelling. *Health Technol Assess* 2007;11(17).
7 Harris M. Undiagnosed NIDDM: clinical and public health issues. *Diabetes Care* 1993;16:642-52.
8 Griffin SJ, Little PS, Hales CN, Kinmonth AL, Wareham NJ. Diabetes risk score: towards earlier detection of type 2 diabetes in general practice. *Diabetes Metab Res Rev* 2000;16:164-71.
9 Spijkerman A, Yuyun M, Griffin S, Dekker J, Nijpels G, Wareham N. The performance of a risk score as a screening test for undiagnosed hyperglycaemia in ethnic minority groups: data from the 1999 health survey for England. *Diabetes Care* 2007;27:116-22.
10 Smith JP. Economics of health and mortality special feature: nature and causes of trends in male diabetes prevalence, undiagnosed diabetes, and the socioeconomic status health gradient. *Proc Natl Acad Sci* 2007;104:13225-31.
11 Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, Brindle P. Performance of the QRISK cardiovascular risk prediction algorithm in an independent UK sample of patients from general practice: a validation study. *Heart* 2008;94:34-9.
12 Wilson PWF, Meigs JB, Sullivan L, Fox CS, Nathan DM, D'Agostino RB Sr. Prediction of incident diabetes mellitus in middle-aged adults: the Framingham offspring study. *Arch Intern Med* 2007;167:1068-74.
13 Lindstrom J, Tuomilehto J. The diabetes risk score: a practical score to predict risk of type two diabetes. *Diabetes Care* 2003;26:725-31.
14 Glumer C, Carstensen B, Sandbaek A, Lauritzen T, Jorgensen T, Borh-Johnsend K. A Danish diabetes risk score for targeted screening: the inter99 study. *Diabetes Care* 2004;27:727-33.
15 Simmons RK, Harding AH, Wareham NJ, Griffin SJ. Do simple questions about diet and physical activity help to identify those at risk of type 2 diabetes? *Diabet Med* 2007;24:830-35.
16 Kanaya AM, Fyr CLW, de Rekeneire N, Shorr RI, Schwartz AV, Goodpaster BH, et al. Predicting the development of diabetes in older adults: the derivation and validation of a prediction rule. *Diabetes Care* 2005;28:404-8.
17 McNeely MJ, Boyko EJ, Leonetti DL, Kahn SE, Fujimoto WY. Comparison of a clinical model, the oral glucose tolerance test, and fasting glucose for prediction of type 2 diabetes risk in Japanese Americans. *Diabetes Care* 2003;26:758-63.
18 Conen D, Ridker PM, Mora S, Buring JE, Glynn RJ. Blood pressure and risk of developing type 2 diabetes mellitus: the women's health study. *Eur Heart J* 2007;28:2937-43.
19 Graf E, Schmoor C, Sauerbrei W, Schumacher M. Assessment and comparison of prognostic classification schemes for survival data. *Stat Med* 1999;18:2529-45.
20 Royston P, Sauerbrei W. A new measure of prognostic separation in survival data. *Stat Med* 2004;23:723-48.
21 Royston P. Explained variation for survival models. *Stata J* 2006;6:1-14.
22 Gaziano TA, Young CR, Fitzmaurice G, Atwood S, Gaziano JM. Laboratory-based versus non-laboratory-based method for assessment of cardiovascular disease risk: the NHANES I follow-up study cohort. *Lancet* 2008;371:923-31.
23 Doran T, Fullwood C, Gravelle H, Reeves D, Kontopantelis E, Hiroeh U, et al. Pay-for-performance programs in family practices in the United Kingdom. *N Engl J Med* 2006;355:375-84.
24 Kumari M, Head J, Marmot M. Prospective study of social and other risk factors for incidence of type 2 diabetes in the Whitehall II study. *Arch Intern Med* 2004;164:1873-80.
25 Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, May M, Brindle P. Derivation and validation of QRISK, a new cardiovascular disease risk score for the United Kingdom: prospective open cohort study. *BMJ* 2007;335:136.
26 Hippisley-Cox J, Coupland C, Vinogradova Y, Robson J, Minhas R, Sheikh A, et al. Predicting cardiovascular risk in England and Wales: prospective derivation and validation of QRISK2. *BMJ* 2008;336:1475-82.
27 Schmidt MI, Duncan BB, Bang H, Pankow JS, Ballantyne CM, Golden SH, et al. Identifying individuals at high risk for diabetes: the atherosclerosis risk in communities study. *Diabetes Care* 2005;28:2013-8.
28 Rahman M, Simmons R, Harding A, Wareham N, Griffin S. A simple risk score identifies individuals at high risk of developing type 2 diabetes: a prospective cohort study. *Fam Pract* 2008;25:191-6.

Accepted: 19 January 2009

Evidence of methodological bias in hospital standardised mortality ratios: retrospective database study of English hospitals

Mohammed A Mohammed,[1] Jonathan J Deeks,[1] Alan Girling,[1] Gavin Rudge,[1] Martin Carmalt,[2] Andrew J Stevens,[1] Richard J Lilford[1]

EDITORIAL by Wright and Shojania

[1]Unit of Public Health, Epidemiology and Biostatistics, University of Birmingham, Birmingham B15 2TT
[2]Royal Orthopaedic Hospital, Birmingham B31 2AP
Correspondence to:
M A Mohammed
M.A.Mohammed@Bham.ac.uk

Cite this as: *BMJ* 2009;338:b780
doi:10.1136/bmj.b780

ABSTRACT

Objective To assess the validity of case mix adjustment methods used to derive standardised mortality ratios for hospitals, by examining the consistency of relations between risk factors and mortality across hospitals.

Design Retrospective analysis of routinely collected hospital data comparing observed deaths with deaths predicted by the Dr Foster Unit case mix method.

Setting Four acute National Health Service hospitals in the West Midlands (England) with case mix adjusted standardised mortality ratios ranging from 88 to 140.

Participants 96 948 (April 2005 to March 2006), 126 695 (April 2006 to March 2007), and 62 639 (April to October 2007) admissions to the four hospitals.

Main outcome measures Presence of large interaction effects between case mix variable and hospital in a logistic regression model indicating non-constant risk relations, and plausible mechanisms that could give rise to these effects.

Results Large significant (P≤0.0001) interaction effects were seen with several case mix adjustment variables. For two of these variables—the Charlson (comorbidity) index and emergency admission—interaction effects could be explained credibly by differences in clinical coding and admission practices across hospitals.

Conclusions The Dr Foster Unit hospital standardised mortality ratio is derived from an internationally adopted/adapted method, which uses at least two variables (the Charlson comorbidity index and emergency admission) that are unsafe for case mix adjustment because their inclusion may actually increase the very bias that case mix adjustment is intended to reduce. Claims that variations in hospital standardised mortality ratios from Dr Foster Unit reflect differences in quality of care are less than credible.

INTRODUCTION

The need to measure quality of care in hospitals has led to publication of league tables of standardised mortality ratios for hospitals in several countries, including England, the United States, Canada, the Netherlands, and Sweden.[1-6] These data have been derived with methods influenced by the seminal work of Jarman et al,[1] and by the subsequent methodological developments by Dr Foster Unit.[7 8] The Dr Foster Unit methodology is used by Dr Foster Intelligence, a former commercial company that is now a public-private partnership, to annually publish standardised mortality ratios for English hospitals in the national press.

A consistent, albeit controversial,[9-11] inference drawn from the wide variation in published standardised mortality ratios for hospitals is that this reflects differences in quality of care.

Case mix adjustment is widely used to overcome imbalances in patients' risk factors so that fairer comparisons between hospitals can be made. Methods for case mix adjustment are often criticised because they can fail to include all the important case mix variables and do not adequately adjust for a variable because of measurement error.[10 11] Moreover, Nicholl pointed out that case mix adjustment can create biased comparisons when underlying relations between case mix variables and outcome are not the same in all the comparison groups.[12] This phenomenon has been termed "the constant risk fallacy," because if the risk relations are assumed to be constant, but in fact are not, then case mix adjustment may be more misleading than crude comparisons.[12] Two key mechanisms can give rise to non-constant risk relations. The first mechanism involves differential measurement error (see box), and the second one involves inconsistent proxy measures of risk.

The second mechanism can occur even in the absence of measurement error. Consider emergency admissions to hospitals. Patients admitted as emergencies are usually regarded as being seriously ill, but if an individual hospital often admits the "walking wounded" as emergencies, then the risk associated with being an emergency admission in that hospital will be reduced. Variation in this practice across hospitals leads to a non-constant relation between emergency admission and mortality.

A simple way to screen case mix variables for their susceptibility to non-constant risk relations is to test for interaction effects between hospital and case mix variables in a logistic regression model that predicts death in hospital.[12] If a large interaction effect is found, this indicates a non-constant risk relation. If this is due to inconsistent measurement practices across hospitals, or because the covariate genuinely has different relations with death across hospitals, it will result in a misleading adjustment to standardised mortality ratios. Alternatively, the interaction could occur if different levels of the covariate were associated with different standards of care across hospitals. Unfortunately, no statistical method exists for differentiating explanations, but they can be explored by seeking a likely cause for the observed interaction effect.

In this paper we screened the Dr Foster Unit method,[13] for its susceptibility to the constant risk fallacy.

This article is an abridged version of a paper that was published on bmj.com. Cite this article as: *BMJ* 2009;338:b780

METHODS

Dr Foster Unit case mix adjustment method—The Dr Foster Unit case mix adjustment method uses data derived from routinely collected hospital episode statistics on every inpatient admission in NHS hospitals in England.[13] The standardised mortality ratio is derived from logistic regression models, which are based on 56 primary diagnosis groups accounting for 80% of hospital mortality. Covariates for case mix adjustment in the model are sex, age group, method of admission, deprivation, primary diagnosis, emergency admissions in the previous year, whether the patient was admitted to a palliative care specialty, and the Charlson (comorbidity) index (range 0-6).[14]

Study hospitals and data sources—This study involves four hospitals, representing a wide range of the published case mix adjusted Dr Foster Unit standardised mortality ratios (88-143, for the period April 2005-March 2006). The hospital with the lowest standardised mortality ratio is a large teaching hospital (University Hospital North Staffordshire); those with higher ratios were one large teaching hospital (University Hospitals Coventry and Warwickshire) and two medium sized acute hospitals (Mid Staffordshire Hospitals and George Eliot Hospital). Our analyses are based on data and predictions for the following time periods: April 2005 to March 2006 (year 1), April 2006 to March 2007 (year 2), and April to October 2007 (part of year 3).

Statistical analyses—The Dr Foster Unit dataset includes the predicted risk of death for each patient, which we included as an offset term in a logistic regression model of in-hospital deaths. To this model we added terms for each hospital and then interaction terms for each hospital and case mix variable in turn. We tested the significance of interactions; we deemed P values ≤ 0.01 to be statistically significant.

Selected variables—The following patient level variables included in the Dr Foster Unit adjustment were available and tested: Charlson index, age, sex, deprivation, primary diagnosis, emergency admission, and the number of emergency admissions in the previous year. We excluded less than 1.5% of all the data because of missing data. For two prominent case mix variables—the Charlson index of comorbidity and emergency admission—we did detective work to seek explanations for the presence of large interaction effects.

Investigation of interaction effects seen with Charlson index—We investigated the possibility of systematic undercoding in the Charlson index. Firstly, we investigated changes in the depth of clinical coding (number of ICD-10 codes for secondary diagnoses identified per admission) over time within the hospitals. Secondly, we considered that if clinical coding was similarly accurate in all hospitals, then differences in the Charlson index should reflect genuine differences in case mix profiles.

Investigation of interaction effects seen with emergency admission—The practice of admitting less seriously ill patients as emergency admissions has been increasingly used in the NHS to comply with accident and emergency waiting time targets.[15,16] This potentially leads to a reduction in the risk of mortality associated with emergency admissions. We examined the magnitude of differences in the proportion of emergency admissions with zero length of stay both within hospitals over time and between hospitals, as well as the observed risk associated with zero and non-zero lengths of stay.

RESULTS

The table reports the odds ratios of tests of interactions for six case mix variables. Two variables (sex and deprivation) had no significant interaction with hospitals. However, the remaining variables had significant interactions. The number of previous emergency admissions was significant in year 2; the three hospitals with high standardised mortality ratios had 6% to 10% increases in odds of death with every additional previous emergency admission over and above the allowance made in the Dr Foster Unit model. Age had a significant interaction in year 2, but the effect was small. Primary diagnosis also had significant interactions in all three years (results not shown).

The Charlson index had significant interaction effects in year 1 and year 2 but not in year 3. A unit change in the Charlson index was associated with a wide range of effect sizes—up to a 7% increase in odds of death and an 8% reduction in odds of death over and above that accounted for in the Dr Foster Unit model. Across the full range of the Charlson

Example of differential measurement error

To illustrate the constant risk fallacy we construct hypothetical hospital mortality data with a single case mix variable—a comorbidity index (CMI) that takes values 0 to 6. The relation between in-hospital mortality and CMI value has been modelled for the population, estimating risks of in-hospital death of 0.02, 0.04, 0.08, 0.14, 0.25, 0.40, and 0.57 in the seven CMI categories (equivalent to an odds ratio of two for each unit increase in the index).

Consider two hospitals, A and B, both of which admit 1000 patients a year in each of the seven CMI categories. Assume that the case mix of the groups of patients and the quality of care in the two hospitals are identical and that 1500 deaths are observed in both hospitals. Hospital A correctly codes the comorbidity index, whereas hospital B tends to under-code, such that in hospital B for each true CMI the following are recorded:

- CMI=0: all are coded as 0
- CMI=1: 50% coded 0, 50% coded 1
- CMI=2: 33% coded 0, 33% coded 1, 33% coded 2
- CMI=3: 25% coded 0, 25% coded 1, 25% coded 2, 25% coded 3
- CMI=4: 20% coded 0, 20% coded 1, 20% coded 2, 20% coded 3, 20% coded 4
- CMI=5: 20% coded 1, 20% coded 2, 20% coded 3, 20% coded 4, 20% coded 5
- CMI=6: 20% coded 2, 20% coded 3, 20% coded 4, 20% coded 5, 20% coded 6.

Rather than observing 1000 patients in each of the seven CMI categories, in hospital B the numbers instead are 2283, 1483, 1184, 850, 600, 400, and 200. The expected number of deaths in hospital A is (1000×0.02)+(1000×0.04)+(1000×0.08)+(1000×0.14)+(1000×0.25)+(1000×0.40)+(1000×0.57)=1500, yielding a standardised mortality ratio (observed/expected deaths) of 1500/1500=100. The expected number of deaths in hospital B is (2283×0.02)+(1483×0.04)+(1184×0.08)+(850×0.14)+(600×0.25)+(400×0.40)+(200×0.57)=743, yielding a standardised mortality ratio of 1500/743=202.

It thus wrongly seems that the mortality in hospital B is twice that in hospital A. Modelling the data by using logistic regression reveals that whereas the relation between CMI and mortality in hospital A is the same as in the population (odds ratio=2.0 per category increase), the relation in hospital B is weaker (odds ratio=1.6 per category increase in CMI), and the interaction between hospital B and CMI is clinically and statistically significant (P<0.001).

index, these correspond to increases in odds of death of 50% or decreases of 39%.

We found significant interactions with being an emergency admission in all years across all hospitals. The effect sizes ranged from 38% to 355% increases in odds of death above those accounted for in the Dr Foster Unit equation.

Investigation of interaction effects seen with Charlson index

The 96 948 admissions in the four hospitals for 2005/06 had an overall mean Charlson index of 1.17 (median 1, interquartile range 0-2). The hospital with a low standardised mortality ratio (University Hospital North Staffordshire) had the highest mean Charlson index (1.54), whereas the three hospitals with high standardised mortality ratios had mean Charlson index values near or below the median (1) (see bmj.com).

University Hospital North Staffordshire had the highest mean coding depth and Charlson index in all years; more importantly, as coding depth increased over the years in all hospitals, the interaction between the Charlson index and hospitals became smaller and statistically non-significant (table). The percentage of emergency admissions, readmissions, length of stay, and crude mortality at University Hospital North Staffordshire are at variance with the view that this hospital treats a systematically "sicker" population of patients. The evidence is therefore inconsistent with the explanation that differences in the Charlson index reflect genuine differences in case mix profiles (see bmj.com).

Investigation of interaction effects seen with emergency admission

The proportion of emergency admissions with zero length of stay varied between 10.4% and 20.4% across hospitals. The hospital with the lower case mix adjusted standardised mortality ratio (University Hospital North Staffordshire) had the highest proportion of zero stay emergency patients in years 2 and 3 (20.4% and 17.7%), whereas the hospital with the highest standardised mortality ratio (George Eliot Hospital) had the lowest proportion of zero stay emergency patients in all three years (10.4%, 11.0%, and 12.9%). The large variations in proportions of emergency/non-emergency patients with zero length of stay indicate that systematically different admission policies were being adopted across hospitals (see bmj.com).

DISCUSSION

Our results show that the relation between risk factors used in case mix adjustment and mortality differed across the hospitals, leading to the constant risk fallacy. This phenomenon can increase the very bias that case mix adjustment is intended to reduce.[12] The routine use of locally collected administrative data for case mix variables makes this a real concern.[12] A serious problem is

Interactions between case mix variables and hospital

Variable and year*	Hospital specific odds ratio† (95% CI)				Likelihood ratio test‡; P value
	GEH	MSH	UHC	UHN	
Charlson index (per unit increase in Charlson index)					
April 2005 to March 2006	1.07 (1.02 to 1.13)	1.06 (1.00 to 1.12)	0.99 (0.96 to 1.02)	1.00 (0.96 to 1.03)	χ^2=13.10; P=0.01
April 2006 to March 2007	1.02 (0.98 to 1.07)	1.01 (0.95 to 1.07)	0.99 (0.95 to 1.02)	0.92 (0.90 to 0.95)	χ^2=26.63; P<0.0001
April to October 2007	1.02 (0.95 to 1.09)	0.96 (0.89 to 1.03)	1.00 (0.96 to 1.05)	0.99 (0.95 to 1.03)	χ^2=1.64; P=0.80
Emergency admission					
April 2005 to March 2006	1.68 (1.21 to 2.34)	1.76 (1.23 to 2.52)	1.44 (1.22 to 1.71)	1.79 (1.46 to 2.20)	χ^2=77.18; P<0.0001
April 2006 to March 2007	2.14 (1.39 to 3.29)	4.55 (2.79 to 7.42)	1.75 (1.46 to 2.11)	3.09 (2.58 to 3.69)	χ^2=322.66; P<0.0001
April to October 2007	2.68 (1.45 to 4.96)	1.85 (1.16 to 2.95)	1.38 (1.10 to 1.74)	1.45 (1.18 to 1.80)	χ^2=42.48; P<0.0001
Age (per 10 year age group)					
April 2005 to March 2006	1.01 (1.00 to 1.01)	1.00 (1.00 to 1.01)	1.00 (1.00 to 1.01)	1.00 (1.00 to 1.00)	χ^2=8.11; P=0.09
April 2006 to March 2007	1.00 (1.00 to 1.01)	1.01 (1.00 to 1.01)	1.00 (1.00 to 1.01)	1.01 (1.00 to 1.01)	χ^2=25.00; P=0.0001
April to October 2007	1.00 (0.99 to 1.01)	1.00 (0.99 to 1.01)	1.00 (1.00 to 1.01)	1.00 (0.99 to 1.00)	χ^2=3.01; P=0.56
Previous emergency admissions (per extra admission)					
April 2005 to March 2006	1.02 (0.96 to 1.10)	1.06 (0.98 to 1.15)	1.01 (0.97 to 1.05)	1.00 (0.95 to 1.05)	χ^2=3.20; P=0.53
April 2006 to March 2007	1.06 (0.99 to 1.14)	1.10 (1.02 to 1.19)	1.07 (1.03 to 1.12)	0.99 (0.95 to 1.03)	χ^2=19.61; P=0.0006
April to October 2007	0.92 (0.84 to 1.02)	1.05 (0.95 to 1.16)	1.03 (0.97 to 1.09)	1.00 (0.94 to 1.06)	χ^2=3.97; P=0.41
Sex					
April 2005 to March 2006	1.10 (0.95 to 1.27)	0.91 (0.78 to 1.06)	1.02 (0.92 to 1.13)	0.97 (0.87 to 1.09)	χ^2=3.23; P=0.52
April 2006 to March 2007	1.02 (0.88 to 1.19)	0.89 (0.76 to 1.05)	1.12 (1.01 to 1.25)	1.03 (0.94 to 1.14)	χ^2=7.20; P=0.13
April to October 2007	0.99 (0.80 to 1.22)	0.96 (0.78 to 1.19)	1.07 (0.93 to 1.23)	0.90 (0.79 to 1.03)	χ^2=3.27; P=0.51
Deprivation (per fifth)					
April 2005 to March 2006	1.00 (0.95 to 1.05)	1.02 (0.96 to 1.01)	1.00 (0.97 to 1.04)	1.01 (0.97 to 1.04)	χ^2=0.38; P=0.98
April 2006 to March 2007	1.00 (0.95 to 1.06)	1.02 (0.96 to 1.09)	1.02 (0.96 to 1.09)	0.98 (0.95 to 1.02)	χ^2=3.01; P=0.56
April to October 2007	0.98 (0.91 to 1.06)	0.94 (0.87 to 1.02)	1.02 (0.97 to 1.07)	1.05 (1.00 to 1.09)	χ^2=6.79; P=0.15

GEH=George Eliot Hospital; MSH=Mid Staffordshire Hospitals; UHC=University Hospitals Coventry and Warwickshire; UHN=University Hospital North Staffordshire.
*Year three, April to October 2007, is a part year because these were the most recent data available at the time of study.
†For relation between each case mix variable and mortality over and above that accounted for in Dr Foster Unit case mix adjustment equation.
‡Global test for systematic deviation from odds ratio=1 in any hospital; df=4.

WHAT IS ALREADY KNOWN ON THIS TOPIC

Case mix adjusted hospital standardised mortality ratios are used around the world in an effort to measure quality of care

However, valid case mix adjustment requires that the relation between each case mix variable and mortality is constant across all hospitals (a constant risk relation)

Where this requirement is not met, case mix adjustment may be misleading, sometimes to the degree that it will actually increase the very bias it is intended to reduce

WHAT THIS STUDY ADDS

Non-constant risk relations exist for several case mix variables used by the Dr Foster Unit to derive standardised mortality ratios for English hospitals, raising concern about the validity of the ratios

The cause of the non-constant risk relation for two case mix variables—a comorbidity index and emergency admission—is credibly explained by differences in clinical coding and hospitals' admission practices

Case mix adjustment methods should screen case mix variables for non-constant risk relations

that no statistical fix exists for overcoming the challenges of variables susceptible to this constant risk fallacy.[12]

We screened variables for non-constant risk and found that three of seven—age, sex, and deprivation—were safe in this respect. However, we found that emergency admission, the Charlson (comorbidity) index, primary diagnosis, and the number of emergency admissions in the previous year had clinically and statistically significant interaction effects. For two variables, the Charlson index and emergency admission, we found credible evidence to suggest that the non-constant risks were caused by systematic differences in clinical coding and emergency admission practices across hospitals.

For the Charlson index variable, we showed how the interaction effects seemed to relate to the number of ICD-10 codes per admission—that is, depth of clinical coding.[17] We reasoned that as the increased depth of coding (over time) was accompanied by a decrease in the interaction effect and as differences in the Charlson index did not reflect genuine differences in case mix profiles, we could reasonably conclude that the Charlson index is prone to the constant risk fallacy largely as a result of differential measurement error from clinical coding practices. For the emergency admission variable, we found strong evidence of systematic differences across hospitals in numbers of patients admitted as emergencies who were admitted and discharged on the same day. The higher risk usually associated with emergencies would be diluted by the inclusion of zero length of stay admissions in some hospitals. Thus, we judge these two variables to be unsafe to use in case mix adjustment methods. Further research to understand the mechanisms behind the other variables with large interactions is clearly warranted.

Our analyses are based on a subset of hospitals in the West Midlands, and our study urgently needs to be replicated with more hospitals. Given the widespread use of case mix adjusted outcome comparisons in health care, we urge that all case mix adjustment methods should screen variables for their susceptibility to the constant risk fallacy. A similar analysis could be done within a single hospital to discover which set of the case mix variables has any systematic relation with mortality over and above the original adjustments.

Our findings suggest that the current Dr Foster Unit method is prone to bias and that any claims that variations in standardised mortality ratios for hospitals reflect differences in quality of care are less than credible.[8 13] We urge that screening case mix variables for non-constant risk relations needs to become an integral part of validating case mix adjustment methods.

This independent study was commissioned by the NHS West Midlands Strategic Health Authority. We are grateful for the support of all the members of the steering group, chaired by R Shukla. We especially thank the staff of participating hospitals, in particular P Handslip. Special thanks go to Steve Wyatt for his continued assistance with the project. We also thank our reviewers for their helpful suggestions.

Contributors: See bmj.com.

Funding: The study was part of a study commissioned by the NHS West Midlands Strategic Health Authority. AG is supported by the EPSRC MATCH consortium.

Competing interests: None declared.

Ethical approval: Not needed.

1 Jarman B, Gault S, Alves B, Hider A, Dolan S, Cook A, et al. Explaining differences in English hospital death rates using routinely collected data. *BMJ* 1999;318:1515-20.
2 Institute for Healthcare Improvement. *Moving your dot: measuring, evaluating, and reducing hospital mortality rates*. Cambridge: Institute for Healthcare Improvement, 2003 (available at www.ihi.org/NR/rdonlyres/CC13C05E-9435-4692-8D7B-771BBAF44C9C/3609/FinalMoveDot1.pdf).
3 Institute for Healthcare Improvement. *Reducing hospital mortality rates (part 2)*. Cambridge: Institute for Healthcare Improvement, 2005 (available at www.ihi.org/NR/rdonlyres/CC13C05E-9435-4692-8D7B-771BBAF44C9C/3605/ReducingHospitalMortalityRates2WhitePaper2005.pdf).
4 Canadian Institute for Health Information. *HSMR: a new approach for measuring hospital mortality trends in Canada*. Ottawa: CIHI, 2007.
5 Heijink R, Koolman X, Pieter D, van der Veen A, Jarman B, Westert G. Measuring and explaining mortality in Dutch hospitals; the hospital standardized mortality rate between 2003 and 2005. *BMC Health Serv Res* 2008;8:73.
6 Koster M, Jurgensen U, Spetz C, Rutberg H. [Standardized hospital mortality as quality measurement in healthcare centres and hospitals.] *Lakar Tidningen* 2008:19:8:1391-6. (In Swedish.)
7 Dr Foster Intelligence. What is Dr Foster Intelligence? www.drfosterintelligence.co.uk/aboutUs/.
8 Imperial College London. Dr Foster Unit at Imperial College. 2008. www1.imperial.ac.uk/medicine/research/researchthemes/publicandint/research/drfosters/.
9 Moran JL, Solomon PJ. Mortality and other event rates: what do they tell us about performance? *Crit Care Resusc* 2003;5:292-304.
10 Lilford R, Mohammed MA, Spiegelhalter D, Thomson R. Use and misuse of process and outcome data in managing performance of acute medical care: avoiding institutional stigma. *Lancet* 2004;363:1147-54.
11 Iezzoni LI. The risks of risk adjustment. *JAMA* 1997;278:1600-7.
12 Nicholl J. Case-mix adjustment in non-randomised observational evaluations: the constant risk fallacy. *J Epidemiol Community Health* 2007;61:1010-3.
13 Dr Foster Intelligence. Hospital guide. www.drfoster.co.uk/hospitalguide.
14 Charlson ME, Pompei P, Ales KL, MacKenzie CR. A new method of classifying prognostic comorbidity in longitudinal studies: development and validation. *J Chron Dis* 1987;40:373-83.
15 Locker TE, Mason SM. Are we hitting the target but missing the point? Analysis of the distribution of time patients spend in the emergency department. *BMJ* 2005;330:1188-9.
16 Healthcare Commission. *Acute hospital portfolio review. Management of admissions in acute hospitals: review of the national findings—2006*. London: Healthcare Commission, 2006 (available at www.healthcarecommission.org.uk/_db/_documents/Management_of_admissions_national_report.pdf).
17 Dixon J, Sanderson C, Elliott P, Walls P, Jones J, Petticrew M. Assessment of the reproducibility of clinical coding in routinely collected hospital activity data: a study in two hospitals. *J Public Health Med* 1998;20:63-9.

Accepted: 18 November 2008

pico

Migraines during pregnancy linked to stroke and vascular diseases: US population based case-control study

Cheryl D Bushnell,[1] Margaret Jamison,[2] Andra H James[3]

[1]Department of Neurology, Wake Forest University Health Sciences, Medical Center Boulevard, Winston-Salem, NC 27157, USA
[2]University of North Carolina School of Medicine, Office of Educational Development, Chapel Hill, NC 27599-7530
[3]Division of Maternal Fetal-Medicine, Department of Obstetrics and Gynecology, Duke University Medical Center, Durham, NC 27710
Correspondence to: C D Bushnell cbushnel@wfubmc.edu

Cite this as: *BMJ* 2009;338:b664
doi: 10.1136/bmj.b664

STUDY QUESTION What is the prevalence of peripartum migraine headache and its associated medical conditions and complications during pregnancy?

SUMMARY ANSWER The prevalence of migraine discharge codes is low (185 per 100 000 deliveries). However, this probably represents only pregnant women with active migraine during admission to hospital. Although cause and effect still need to be established, active migraine during pregnancy could be viewed as a marker of vascular diseases, especially ischaemic stroke.

Participants and setting

Our case-control study was based on the nationwide inpatient sample from the Healthcare Cost and Utilization Project, which included a total of 18 345 538 pregnancy related discharges from US hospitals during 2000 to 2003.

Design, size, and duration

We calculated frequencies of discharges with ICD codes for migraine for each maternal age group, ethnic group, timing of pregnancy related discharge, comorbidity, and pregnancy complication. We identified jointly associated factors with multivariable logistic regression modelling developed from statistically and clinically significant common vascular comorbidities.

Primary outcome(s), risks, and exposures

Migraine discharge codes and the jointly associated discharge codes for vascular and pregnancy complications.

Main results and the role of chance

We found 33 956 migraine discharges, or 185 per 100 000 deliveries, in this cohort. In the descriptive analysis, a strong association existed between migraine discharge codes and stroke codes of all types (odds ratio 15.8, 95% confidence interval 11.1 to 22.5), but especially ischaemic stroke (30.7, 17.4 to 34.1). Migraine discharge codes were also significantly associated with codes for myocardial infarction (4.9, 1.7 to 14.2), pulmonary embolus (3.1, 1.7 to 5.6), deep venous thrombosis (2.4, 1.3 to 4.2), thrombophilia (3.6, 2.1 to 6.1), diabetes (2.3, 1.9 to 2.7), hypertension (3.6, 3.1 to 4.2), cigarette smoking (2.7, 2.4 to 3.1), and pre-eclampsia or gestational hypertension (2.3, 2.1 to 2.5). We found no association with most non-vascular diagnoses. A multivariable logistic regression included age and removal of pre-eclampsia diagnoses from records with stroke, vascular diagnoses, hypertension, smoking, and diabetes. Key findings from this analysis are shown in the table.

LOGISTIC REGRESSION ANALYSIS OF ASSOCIATIONS WITH MIGRAINE DISCHARGE CODES

Independent variable	Odds ratio (95% CI)	P value
Age	1.03 (1.02 to 1.03)	<0.001
Pre-eclampsia	2.29 (2.13 to 2.46)	<0.001
All strokes*	15.05 (8.26 to 27.4)	<0.001
Venous thromboembolism or pulmonary embolus*	3.23 (2.06 to 7.07)	<0.001
Acute myocardial infarction or heart disease*	2.11 (1.76 to 2.54)	<0.001
Hypertension*	8.61 (6.43 to 11.54)	<0.001

Pre-eclampsia or eclampsia codes excluded from records

Bias, confounding, and other reasons for caution

The results of this study include only women with migraines that were active during hospital admission and listed as the primary or secondary discharge diagnosis. Therefore, the population does not include women with mild migraines treated as outpatients or women with a history of migraines that are quiescent during pregnancy. Other reasons for caution exist. Firstly, we were unable to separate migraines with aura from migraines without aura. Secondly, peripartum migraine could have been miscoded or inadvertently listed separately in the setting of cerebrovascular complications in which headache is a prominent symptom, such as cerebral venous thrombosis. Thirdly, the headache of severe pre-eclampsia often includes the presence of visual scotomata, which could have been confused with migraine. Fourthly, we were unable to compare these results in non-pregnant women because the database was limited to pregnancy discharge codes. Fifthly, this study does not allow us to establish cause and effect because the timing of migraine and onset of the vascular event is uncertain.

Generalisability to other populations

This analysis was done in US hospitals only. Whether the results are generalisable to non-US pregnancy discharges is unclear.

Study funding/potential competing interests

The study was funded in part by the National Institutes of Health. The authors were independent from the funders in all aspects of the study design, analysis of data, and writing of the manuscript.

This is a summary of a paper that was published on bmj.com as *BMJ* 2009;338:b664

pico

Written informed consent and selection bias in observational studies using medical records: systematic review

Michelle E Kho,[1] Mark Duffett,[2] Donald J Willison,[1] Deborah J Cook,[1 3] Melissa C Brouwers[1 4]

[1]Department of Clinical Epidemiology and Biostatistics, McMaster University, 1280 Main Street West, Hamilton, ON, Canada L8S 4L8
[2]Department of Critical Care, McMaster Children's Hospital, Hamilton, ON, Canada
[3]Department of Medicine, McMaster University, Hamilton, ON, Canada
[4]Program in Evidence-based Care, Cancer Care Ontario, McMaster University, Canada

Correspondence to: M E Kho
khome@mcmaster.ca

Cite this as: *BMJ* 2009;338:b866
doi: 10.1136/bmj.b866

STUDY QUESTION Does informed consent for medical record use introduce selection bias? Are there differences in key demographic variables between participants and non-participants in prospective observational studies requiring informed consent for medical records access? What are the consent rates in these studies?

SUMMARY ANSWER Significant differences between participants and non-participants may threaten the validity of results from observational studies requiring consent for use of medical records. To ensure that privacy legislation does not unduly bias observational studies using medical records, research ethics boards must consider carefully the need for mandatory consent.

Selection criteria for studies

We searched Embase (1980 to week 13, 2008), Medline (1966 to week 3, March, 2008), and the *Cochrane Library* (issue 1, 2008) for English language studies. We sought all studies reporting characteristics of participants and non-participants approached for informed consent to use their medical records for prospective observational studies or registries. We included studies reporting at least one of the following characteristics: age, sex, race, education, income, or health status.

Primary outcome

Comparisons between participants and non-participants by age, sex, race, education, income, or health status.

Main results and role of chance

Of 1650 citations, 17 unique studies met our inclusion criteria and had analysable data. Our inter-rater reliability for included studies was 0.84 (95% CI 0.83 to 0.86). Of 161 604 eligible patients in the 17 studies, 108 033 (66.9% (95% CI 66.6% to 67.1%)) provided active consent for use of their medical records. Consent rates for eligible participants varied across the studies (36.6% to 92.9%). By characteristic, we identified 16 studies reporting age, 14 reporting sex, seven reporting income, and six reporting race, education, or health status. Across all outcomes, differences between participants and non-participants occurred, but there was a lack of consistency in the direction and size of effect.

Bias, confounding, and other reasons for caution

Our review was limited by the published reports—including lack of clarity about the sample size and reporting standards for screening and consent procedures. Not all studies reported data on our outcomes of interest; authors may not have collected data on these outcomes or chose to report only significant differences between enrolled and non-enrolled patients. Because these observational studies were not specifically designed to study differences in consent between participants and non-participants, we may have observed statistically significant differences across our outcomes of interest simply due to chance.

SUGGESTED STRATEGIES TO MINIMISE BIAS FROM INFORMED CONSENT

Request a waiver of consent from research ethics boards with explicit procedures to protect patient confidentiality

If a waiver is not possible then:
- Collect a minimum dataset of key prognostic variables on all eligible people identified through screening
- Complete a preliminary analysis comparing participants and non-participants on key prognostic variables at predetermined times
- Revise the strategy for recruitment as necessary

Educate clinicians, researchers, and research ethics boards on conditions under which studies can proceed without individual consent

Standardise reporting of methods used to seek informed consent

Increase awareness by clinicians and researchers of the potential for selection bias from informed consent and implications for interpretation of result

Study funding/potential competing interests

MEK is funded by a fellowship from the Canadian Institutes of Health Research (CIHR Clinical Research Initiative). DJC is a research chair of the CIHR. The CIHR had no involvement in the conduct of the study or preparation of the manuscript.

BMJ pico: advice to authors

For Research articles, we routinely post the full version only on bmj.com, with open access, and prepare an abridged version for publication in the print journal. We believe that a well written short version in the print *BMJ* can encourage casual readers to read something they might otherwise miss, while the full version on bmj.com allows serious readers the detail they need.

To increase readership of research articles in the print *BMJ* and to give authors more control over the abridging, we are piloting a new way of abridging research articles for the print *BMJ*—publishing what is essentially an evidence abstract called *BMJ* pico. We hope that you will want to take part in this pilot if your research article is accepted. This would involve you producing, using a template from us, a succinct evidence abstract of your own article after acceptance. There is no need to prepare a *BMJ* pico in advance, however—please wait until we have offered to publish your article.

This is a summary of a paper that was published on bmj.com as *BMJ* 2009;338:b866