ORIGINAL ARTICLE

Estimating risks of common complex diseases across genetic and environmental factors: the example of Crohn disease

C M Lewis, S C L Whitwell, A Forbes, J Sanderson, C G Mathew, T M Marteau

🔓 **UNLOCKED**

This paper is freely available online under the BMJ Journals unlocked scheme, see http://jmg.bmj.com/info/unlocked.dtl

See end of article for authors' affiliations
. . . .

Correspondence to:
Professor Cathryn M Lewis, Department of Medical and Molecular Genetics, King's College London School of Medicine, 8th Floor, Guy's Tower, Guy's Hospital, London SE1 9RT, UK; cathryn.lewis@kcl.ac.uk

Background: Progress has been made in identifying mutations that confer susceptibility to complex diseases, with the prospect that these genetic risks might be used in determining individual disease risk.
Aim: To use Crohn disease (CD) as a model of a common complex disorder, and to develop methods to estimate disease risks using both genetic and environmental risk factors.
Methods: The calculations used three independent risk factors: *CARD15* genotype (conferring a gene dosage effect on risk), smoking (twofold increased risk for smokers), and residual familial risk (estimating the effect of unidentified genes, after accounting for the contribution of *CARD15*). Risks were estimated for high-risk people who are siblings, parents and offspring of a patient with CD.
Results: The CD risk to the sibling of a patient with CD who smokes and carries two *CARD15* mutations is approximately 35%, which represents a substantial increase on the population risk of 0.1%. In contrast, the risk to a non-smoking sibling of a patient with CD who carries no *CARD15* mutations is 2%. Risks to parents and offspring were lower.
Conclusions: High absolute risks of CD disease can be obtained by incorporating information on smoking, family history and *CARD15* mutations. Behaviour modification through smoking cessation may reduce CD risk in these people.

Crohn disease (CD) is a severe inflammatory disorder of the gastrointestinal tract. The disorder can affect any part of the gut, but primarily the ileum and colon are affected. Symptoms vary between patients and include abdominal pain, diarrhoea and weight loss. The inflammation is controlled by drugs (steroids, immune modulators) and surgery. The cause of CD is unknown, but is likely to be an inappropriate or exaggerated mucosal immune response to constituents of gut microflora. The prevalence of CD varies across geographic region and time, with incidence and prevalence rates increasing recently in the developed world. Prevalence rates are between 54 and 214 per 100 000 population,[1] and the onset of CD generally occurs between 15 and 30 years of age.

The aetiology of CD is complex. Monozygotic twins show approximately 50–60% disease concordance, with much lower rates in dizygotic twins (~10%), highlighting the role of both environmental and genetic components in the development of CD.[2 3] The strongest and best replicated environmental risk factor is smoking, which increases both the risk of developing CD and the severity of CD after diagnosis. Other environmental factors that may play a role in CD are appendectomy, oral contraception, diet and domestic hygiene, but the evidence for each of these is much weaker than for smoking, and many contradictory studies exist (reviewed by Loftus[1]).

Familial studies confirm that first degree relatives have a substantially increased risk of CD compared with the general population.[4–8] The lack of a consistent pattern of inheritance in families and the failure of linkage studies to identify strong linkage to any single region indicate that genetic susceptibility to CD is multifactorial, with many genes contributing a small effect towards underlying disease risk. Major progress on the identification of genes contributing to CD has been made recently. The first susceptibility gene to be identified was *CARD15* (*NOD2*);[9–11] three mutations increase risk of CD in a

gene dosage model, with CD risk increasing with the number of mutations (0, 1, 2) carried.[12 13] However, these mutations are also present in 12% of the general Western population, and therefore do not form a very specific test for CD risk. Two further genes have been identified through genome-wide association studies: *IL23R*[14] and *ATG16L1*,[15] and both findings have been independently replicated.[16–18]

In this paper, we present a model for estimating an individual's risk of developing CD, incorporating genetic and environmental risk factors. The model combines *CARD15* genotypes with smoking and family history to calculate CD risk. Although no single factor forms a test of high predictive value, risk factors in combination with each other may raise risks of CD substantially. This provides the potential to identify population subgroups that are at high risk of developing CD, with a view to intervening to prevent disease.[19]

METHODS
Risk components
The probability of developing a multifactorial disease depends on the risk conferred by each genetic and environmental risk factor modelled, and the assumed interaction between these factors. We modelled risk of a disease, D, on three factors: an environmental variable, a genotype at a susceptibility gene and family history.

1. For an environmental risk factor, E, assume a proportion s of the population are exposed to E, increasing their risk of disease by odds ratio r compared with the non-exposed member of the population (E'). So,

$$P(D|E) = r \, P(D|E')$$
and
$$P(E) = s.$$

2. Define a gene, G, with three genotypes (AA, AB, BB), where A is the wildtype allele having frequency p, and B is a mutation increasing risk of disease and having frequency q ($= 1−p$). Assume an arbitrary model for disease risk, with the genotype relative risks for disease risk by genotype being g_0 ($= 1$), g_1 and g_2, where

$P(D|AB) = g_1 P(D|AA)$,
and
$P(D|BB) = g_2 P(D|AA)$.

Let the disease penetrances be f_0, f_1 and f_2, where $f_0 = P(D|AA)$, and similarly for f_1 and f_2, so that $g_2 = f_2/f_0$ and $g_1 = f_1/f_0$. Penetrances can be calculated directly from genotype counts in a case control study using Bayes rule,

$P(D|AA) = P(AA|D) P(D)/P(AA)$,

where $P(D)$ is disease prevalence, and $P(AA)$ is the population frequency of genotype AA. $P(AA)$ is estimated from control genotype frequency if controls are a random population sample of people, or if the disease is rare. If controls are selected to be unaffected and the disease is common, then $P(AA)$ is determined from the weighted case and control frequencies.

Alternatively, the summary measures of risk presented in meta-analysis of genetic association studies may be used to estimate penetrances, enabling risk calculations to be based on more robust estimates than a single case–control study. Risks are usually presented as genotype relative risks (g_1, g_2) for genotypes AB and BB, relative to the wildtype AA genotype. Expressing the disease prevalence as $K = f_0 p^2 + 2pq f_1 + q^2 f_2$ and solving for f_0, gives penetrances in terms of the commonly available measures of allele frequency and genotype relative risk.

3. The final component of the model is the residual familial risk, after accounting for the contribution of gene G to the familial aggregation of disease. Familial risk can be assessed through the sibling relative risk (λ_S), which measures the increase in disease risk to the sibling of a case, compared with a random member of the population.[20] For a gene, G, which has a multiplicative relationship with the remaining familial risks, λ_S can be partitioned into the relative risk due to this gene ($\lambda_{S,G}$), and to the residual risk, after accounting for the gene ($\lambda_{S,G'}$), by

$\lambda_S = \lambda_{S,G} \lambda_{S,G'}$
$\lambda_{S,G}$ can be calculated from the disease risk due to G above,[20 21] as
$\lambda_{S,G} = 1 + (½ V_A + ¼ V_D)/K^2$
where
K is disease prevalence,
$V_A = 2pq (q (f_2 − f_1) + p (f_1 − f_0))^2$
is the additive genetic variance at G and
$V_D = p^2 q^2 (f_2 − 2 f_1 + f_0)^2$
is the dominant genetic variance at G.

Similar expressions using genotype relative risks (1, g_1, g_2) in place of penetrances (f_0, f_1, f_2) can be derived, and then no assumption of disease prevalence is needed in calculating $\lambda_{S,G}$. The sibling relative risk, λ_S, can also be extended to multiple genes, provided no epistasis exists between the genes modelled and between the genes and the residual family history. For estimating disease risk to parents or offspring, λ_P and λ_O are used, with

$\lambda_{P,G} = \lambda_{O,G} = 1 + ½ V_A/K^2$.

Disease risk calculation

To model the disease risk of an individual based on their genotype at G, exposure to the enrivonmental factor E, and their family history, we assume independence between the

environmental factor, genotype at gene G and residual family history (after accounting for G). For the environmental factor E,

$P(D|E) = r P(D|E')$
and
$P(D) = P(D|E) P(E) + P(D|E') (1−P(E))$,
therefore
$P(D|E) = rK/(rs+(1-s))$,
and
$P(D|E') = K/(rs+(1-s))$.
$P(D|G = g_i)$ are given above for $g_i = (AA, AB, BB)$, using genotype relative risks and disease prevalence.

For the sibling of a case, the increase in risk unaccounted for by gene G is $\lambda_{S,G'}$. Piecing together both genetic and environmental contributions to risk, assuming they act independently, gives the probability of disease in the sibling of a case as:

$P(D|G, E, \text{sib affected}) = P(D|G)P(D|E) \lambda_{S,G'}/P(D) =$
$\text{GRR}(G)\text{OR}(E)\lambda_{S,G'} K$

where $\text{GRR}(G)$ and $\text{OR}(E)$ are the genotype relative risk and odds ratios of disease conferred by exposure or non-exposure to the environmental risk factor.

Application to CD
Smoking

Smoking is a well-established risk factor for CD, which increases the risk and severity of CD. A series of epidemiological papers estimated the CD risk associated with smoking, matching patients with CD to controls by relevant factors such as age, sex and birth year, then calculating the odds ratio of CD in smokers compared with never-smokers.[22–24] Many of these papers considered small series of cases and controls, and some used imprecise temporal data of current smoking status (especially in controls), rather than smoking at age of diagnosis of CD in the matched case. A meta-analysis of seven studies estimated a pooled odds ratio (OR) of 2.0 (95% CI 1.65 to 2.47) for CD in smokers compared with never-smokers.[25] No consistent effect on CD risks was seen when data were analysed by sex, by smoking duration, or numbers of cigarettes smoked per day;[25] however, cigarette smoking may specifically increase risk in late-onset CD patients (age of diagnosis >40 years).[26]

Several studies also calculated CD risks in former smokers. The meta-analysis showed an increased risk of CD compared with lifetime non-smokers, with an odds ratio of 1.80 (95% CI 1.33 to 2.51), only slightly lower than the risk for current smokers. Detailed assessment of the time since smoking cessation in the ex-smokers suggests that CD risks remain raised immediately after smoking cessation, and reduce after 2–4 years.[22 24] A more recent case control study of CD and both genetic and epidemiological risk factors estimated an OR for CD of 3.3 in smokers, and 1.7 in ex-smokers (95% CI 1.1 to 2.7).[27] However, no study has systematically examined the risks of CD in long-term ex-smokers, and the potential role of smoking cessation in reducing risks of CD remains unclear. Smoking is currently the only potential modifiable risk factor for CD and such information merits being communicated to smokers at risk of CD. Whether such communication is effective in achieving smoking cessation in this group remains to be seen (www.iop.kcl.ac.uk/grab).

Family history of CD

The familial risks of CD are estimated through ascertaining probands diagnosed with CD, and requesting information on which relatives are affected and unaffected with CD. Four Northern European studies were identified as having sufficient information from a large cohort of CD probands to estimate

Table 1 Summary of family history studies

Reference	Population	CD probands, n	CD siblings, n	Total siblings, n	CD risk, %	Sibling relative risk*
Kuster et al (1989)[4]	Germany	265	13	453	2.87	28.7
Probert et al (1993)[6]	UK	424	19	984	1.93	19.3
Peeters et al (1996)[7]	Belgium	640	57	1728	3.3	33.0
Satsangi et al (1994)[8]	UK	433	20	782	2.56	25.6
Total		2003				
Average risk					2.67	26.7
Average, weighted by study size					2.72	27.2

*Using a population prevalence of 0.001 for all studies.
Estimates of sibling relative risk for each study, mean value across studies, and a mean value weighting by study size (number of probands) are given.

familial relative risks[4–8] (table 1). CD twin studies were omitted because of the low numbers of CD probands.[2 3] Other studies could not be used because CD relatives were classified only as first or second degree, without information on the precise relationship to the proband.[28]

Kuster et al[4] performed a segregation analysis of 265 probands with CD from Dusseldorf, Germany, finding 13 siblings with CD (from a total of 453 siblings), and 4 parents with CD (from 530 parents). Satsangi et al[8] studied 433 adult CD patients in Oxford, UK, showing that 33 siblings, 18 parents and 4 offspring were affected with inflammatory bowel disease. Of the 33 siblings, 20 were affected with CD, and assuming that the ratio of relatives affected with CD to those affected with ulcerative colitis remains constant across all first degree relatives, we estimate that 11.5 parents and 2.4 offspring were affected with CD. Peeters et al[7] ascertained 640 probands with CD in Belgium, showing that 13.6% had a first degree relative with CD, and classifying the affected relatives into siblings (n = 57/1728), parents (n = 21/1280) or offspring (n = 17/835). Probert et al[6] collected family history information on 424 patients with CD from Leicester, UK, finding 19 of 984 siblings affected with CD, 8 of 825 parents and 8 of 493 offspring.

Calculating the sibling relative risk requires a value for CD prevalence. This has been estimated in several UK studies, with rates per 100 000 population of 76 in Leicestershire[6] and 147 in NE Scotland.[29] The incidence of CD is increasing in all Western populations, so more recent estimates are higher, but these studies are from the same time period as the familial studies in table 1. In this study, we assumed 100 patients with CD per 100 000 of population, giving a disease prevalence of $K = 0.001$.

The sibling relative risk (λ_S) is defined as the ratio of the observed risk to siblings of patients with CD to the population prevalence (so $\lambda_S = 1$ provides a baseline of no increased familial risk). Study-specific sibling relative risks ranged from 19[6] to 33,[7] and weighting by the number of probands per study gives a mean sibling relative risk of 27.2 (table 1). Similar calculations from these papers show the estimated relative risk for parents to be $\lambda_P = 12.8$ (from four studies), and relative risk for offspring of probands to be $\lambda_O = 16.6$ (excluding the German study, which did not study the offspring of probands). The lower relative risks of parents and offspring may be due to the greater shared environment of siblings compared with offspring and parents. Alternatively, risks may reflect the lower prevalence of CD relevant for the parental cohort, and the young age of the offspring, who may not have lived through the full risk period for CD.

CARD15 risks

CARD15 (NOD2) was localised in 2001, with the identification of three different mutations (R702W, G908R, L1007fs) conferring increased risk of CD.[9–11] Other, much rarer mutations in CARD15 exist, with frequency <0.1%,[30 31] but are found in only a small

proportion of patients with CD. In common with other studies, these variants are not considered here.

Many research groups have published estimates of the population frequency and genotype relative risks of these CARD15 mutations, and a meta-analysis of 42 studies showed that people carrying one mutation had 2.3-fold increased odds of CD (95% CI 2.00 to 2.86), and people carrying two mutations had a 17.1-fold increased risk of CD (95% CI 10.7 to 27.2).[13] Differences across populations exist, with no mutations present in Asian populations,[32] and a trend across Europe, with higher frequencies of the 1007fs frameshift in central and southern Europe than in northern Europe.[33] Differences in risk conferred by each mutation were detected in the meta-analysis and a large French family study,[13 34] but few individual studies have sufficient power to detect this effect.

In this study, we used data from our large UK study of 1639 patients with CD and 1808 controls, genotyped for the 3 CARD15 mutations as described by Prescott et al.[16] Individuals were categorised by the number of CD mutations carried (0, 1, 2), as mutations occur only rarely on the same haplotype, and without discriminating between the different mutations (table 2). Genotype relative risks are 2.25 (95% CI 1.88 to 2.68) for people carrying one mutation and 9.25 (95% CI 5.44 to 15.7) for people carrying two mutations, compared with those carrying no CARD15 mutations. Using these figures in the equations for the sibling relative risk due to CARD15 above gives $\lambda_{S,CARD15} = 1.16$.

RESULTS

The estimated risks of developing CD, based on the model developed above, are shown in tables 3 and 4. The parameter estimates for the risk calculation are summarised in table 3; in addition to the values derived and discussed above, we assumed that the smoking prevalence in the UK is 24%.[35] The CD risks for the sibling, parent or offspring of a patient with CD, based on CARD15 mutation status (0, or 2 mutations or ungenotyped) and smoking status (smoker, non-smoker or unknown), are given in table 4. The baseline relative risk for the

Table 2 CARD15 genotype counts for UK patients with CD and controls

| | Total | Number of CARD15 mutations carried | | |
		0	1	2
Cases, n (%)	1639	1128 (68.8)	403 (24.6)	108 (6.6)
Controls, n (%)	1808	1546 (85.5)	246 (13.6)	16 (0.9)
Genotype relative risk (95% CI)		1	2.25 (1.88 to 2.68)	9.25 (5.44 to 15.7)

Table 3 Parameter estimates used in CD risk calculation model

Parameter	Notation	Estimate
Prevalence of CD	K	0.001
Smoking prevalence	s	0.24
Increased CD risk for smokers	g	2
Sibling relative risk	λ_S	27.2
CARD15 sibling relative risk	$\lambda_{S,CARD15}$	1.16
GRRs for CARD15 mutations		
No mutation	GRR_0	1
1 mutation	GRR_1	2.24
2 mutations	GRR_2	9.25

sibling of a patient with CD is a sibling relative risk of $\lambda_S = 27.2$, which gives a 2.7% risk of developing CD (assuming a disease prevalence of 0.001). A person who carries no mutations in CARD15 has the risk of CD reduced slightly to 2.3%, person with a single CARD15 mutation has a slightly increased risk of 5.3%, and people with two mutations have a greatly increased risk of 21.7%. Siblings who also smoke increase their risks still further to 3.8% (0 mutation), 8.5% (1 mutation), or 35.0% (2 mutations). Non-smokers have risks equal to approximately half the risks for siblings who smoke. Without information on CARD15, the risks for non-smoking and smoking siblings of patients with CD are 2.2% and 4.4%, respectively. Adding CARD15 mutation status therefore provides considerable resolution of these risks, and permits the identification of a cohort of people (smokers who carry two CARD15 mutations and have a sibling with CD) who have a high probability of developing CD.

For siblings of a patient with CD, a non-smoker with no CARD15 mutations decreases their risk by approximately one-third, whereas a smoker with two CARD15 mutations increases their risk 13-fold, compared with risks of 2.7% based solely on family history information. A non-smoker with a single CARD15 mutation is at approximately the same risk of developing CD as a smoker who carries no CARD15 mutations (as the OR for smoking and for one CARD15 mutations are 2 and 2.24 respectively). The major contribution to disease risk is carrying two CARD15 mutations (OR = 9.24), and the residual family history risk, after accounting for CARD15 ($\lambda_{S,CARD15} = 23$). Similar patterns of risk are seen for the parents and offspring of patients with CD. A program for risk calculation is available on request.

DISCUSSION

We have developed a model to calculate individual-specific risks for complex genetic diseases, based on our current knowledge of environmental and genetic risk factors. This model shows how inclusion of genotype information from only a single gene and environmental risk factor can substantially change disease risks. For an individual with all risk factors present, a moderately high disease risk was obtained (34% for smoking siblings of a patient with CD, with two CARD15 mutations).

Few complex genetic diseases have sufficient information on both genetic and environmental factors to estimate disease risks. In age-related macular degeneration, Maller et al[36] estimated risks of up to 50% from single-nucleotide polymorphism genotypes at three genes. Smoking is also a major risk factor for this condition and, in contrast to CD, an interaction with the LOC387715 locus exists, with risks for smokers who also carry the high-risk genotypes being much higher than predicted by the marginal risks conferred by smoking and LOC387715.[37] A population study of CD in Manitoba[27] showed that people with a family history of CD who also smoked and carried two CARD15 mutations had a much increased risk of CD, with an OR of 257 (95% CI 63 to 1054). The methodology of the Manitoba study was different from the current study, with risk information assessed from a series of healthy controls (n = 336) and patients with CD (n = 232), giving an OR of 13.9 for people carrying two CARD15 mutations, 3.0 for smokers, and 6.2 for the presence of CD in a first degree relative. Despite the disparity of the designs and the specific risk estimates calculated, both this study and the Manitoba study illustrate how combining risks across several factors of modest effect can produce high predicted disease risks for specific subgroups of the population.

One limitation of our model is that it uses no information on the age of the affected and unaffected siblings, or the time since diagnosis of the patient with CD. Siblings are correlated in age at diagnosis, with most affected siblings diagnosed within 10 years of each other.[7 38] Therefore a sibling who remains unaffected with CD more than a decade after the diagnosis of their sibling has lived through the majority of their at-risk period, and must be considered at lower risk. Joint distributions of age at diagnosis of CD in family members could be used to develop further methodology. For example, age-dependent methods for risk estimation were developed in the REVEAL study of Alzheimer's disease, incorporating APOE genotypes and family history and age, using Kaplan–Meier survival methods.[39 40] Our risk model assumes that risk factors are independent, with multiplicative relationship across risks conferred. In CD, a joint study on CARD15 and variants in IL23R and on chromosome 5q31 showed that risks across genes were multiplicative,[16] and smoking prevalence is similar in CD patients with and without CARD15 mutations, implying the lack of any interaction between these risk factors.[27 30] Current studies therefore support the assumption

Table 4 Absolute risks of developing CD

| Relationship to patient with CD | Smoking status | Risk (%) related to number of CARD15 mutations | | | |
		0	1	2	Not genotyped
Sibling	Non-smokers	1.9	4.2	17.5	2.2
	Smokers	3.8	8.5	35.0	4.4
	Unknown	2.3	5.3	21.7	2.7
Offspring	Non-smokers	1.2	2.6	10.7	1.3
	Smokers	2.3	5.2	21.4	2.7
	Unknown	1.4	3.2	13.2	1.7
Parents	Non-smokers	0.9	2.0	8.2	1.0
	Smokers	1.8	4.0	16.5	2.1
	Unknown	1.1	2.5	10.2	1.3

Estimated from risk model, by relationship to patient with CD (sibling, offspring, parent), by smoking status and by CARD15 genotype.

of independent contribution to risk from each genetic and environmental source. The model can be expanded to multiple genes (eg *IL23R* and *ATG16L1*) and additional environmental risk factors, provided the assumption of multiplicative risks across factors holds.

The method for risk estimation developed here is flexible in allowing risk factors estimated from independent studies to be combined. This is a valuable property for CD, as few studies with good information on family history and environmental and genetic risk factors exist. Other methods for risk estimation require more extensive data. For example, the BOUDICEA model was developed for familial breast cancer using segregation analysis on extended families, and incorporating information from a population series of breast cancer cases.[41] Extended pedigrees for CD are rarely available, and genetic studies have focused on collections of affected people (affected sibling pairs, case series), so this tool could not be used here. For studies in which genetic and environmental risk factor information is available for all participants, regression models may be used to estimate risks, and then these risks validated in an independent dataset. Again, suitable datasets for CD are not currently available, and no validation for the current model has yet been performed.

The utility of screening unaffected relatives for mutations in *CARD15* to assess their risk of developing CD has been much debated.[42–44] Such screening is considered to be of limited value because of the low predictive value of the test (14% of the UK population carries at least one mutation in *CARD15*). However, the utility in screening smokers at risk who have the potential to reduce their risk is unknown. The risk estimates published here are being used to provide smokers who are first degree relatives of patients with CD with estimates of the likelihood that they will develop CD, together with information about the potentially risk-reducing effects of smoking cessation. Of interest is whether providing precise disease risk estimates based on DNA analysis together with information about reducing these risks motivates smoking cessation, a question we are investigating in a randomised controlled trial.

ACKNOWLEDGEMENTS

. . . .

Authors' affiliations
C M Lewis, C G Mathew, Department of Medical and Molecular Genetics, Division of Genetics and Molecular Medicine, King's College London School of Medicine, Guy's Campus, London, UK
C M Lewis, Social, Genetic and Developmental Psychiatry Research Centre, Institute of Psychiatry, King's College London, London, UK
S C L Whitwell, T M Marteau, Psychology and Genetics Research Unit, King's College London School of Medicine, Guy's Campus, London, UK
A Forbes, Department of Gastroenterology, Institute for Digestive Diseases, University College London Hospitals Trust, London, UK
J Sanderson, Department of Gastroenterology, Guy's and St Thomas' NHS Foundation Trust, London, UK

REFERENCES
1 **Loftus EV**. Clinical epidemiology of inflammatory bowel disease: incidence, prevalence, and environmental influences. *Gastroenterol* 2004;**126**:1504–17.
2 **Orholm M**, Binder V, Sorensen TI, Rasmussen LP, Kyvik KO. Concordance of inflammatory bowel disease among Danish twins. Results of a nationwide study. *Scand J Gastroenterol* 2000;**35**:1075–81.
3 **Halfvarson J**, Bodin L, Tysk C, Lindberg E, Järnerot G. Inflammatory bowel disease in a Swedish twin cohort: a long-term follow-up of concordance and clinical characteristics. *Gastroenterol* 2003;**124**:1767–73.
4 **Kuster W**, Pascoe L, Purrmann J, Funk S, Majewski F. The genetics of Crohn disease: complex segregation analysis of a family study with 265 patients with Crohn disease and 5,387 relatives. *Am J Med Genet* 1989;**32**:105–8.
5 **Meucci G**, Vecchi M, Torgano G, Arrigoni M, Prada A, Rocca F, Curzio M, Pera A, de Franchis R. Familial aggregation of inflammatory bowel disease in northern Italy: a multicenter study. The Gruppo di Studio per le Malattie Infiammatorie Intestinali (IBD Study Group). *Gastroenterol* 1992;**103**:514–19.
6 **Probert CSJ**, Jayanthi V, Hughes AO, Thompson JR, Wicks CAN, Mayberry JF. Prevalence and family risk of ulcerative colitis and Crohn's disease - an epidemiologic study among Europeans and South Asians in Leicestershire. *Gut* 1993;**34**:1547–51.
7 **Peeters M**, Nevens H, Baert F, Hiele M, de Meyer AM, Vlietinck R, Rutgeerts P. Familial aggregation in Crohn's disease: increased age-adjusted risk and concordance in clinical characteristics. *Gastroenterol* 1996;**111**:597–603.
8 **Satsangi J**, Rosenberg WMC, Jewell DP. The prevalence of inflammatory bowel disease in relatives of patients with Crohn's disease. *Eur J Gastroent Hepat* 1994;**6**:413–16.
9 **Hampe J**, Cuthbert A, Croucher PJ, Mirza MM, Mascheretti S, Fisher S, Frenzel H, King K, Hasselmeyer A, MacPherson AJ, Bridger S, van Deventer S, Forbes A, Nikolaus S, Lennard-Jones JE, Foelsch UR, Krawczak M, Lewis C, Schreiber S, Mathew CG. Association between insertion mutation in NOD2 gene and Crohn's disease in German and British populations. *Lancet* 2001;**357**:1925–8.
10 **Hugot JP**, Chamaillard M, Zouali H, Lesage S, Cezard JP, Belaiche J, Almer S, Tysk C, O'Morain CA, Gassull M, Binder V, Finkel Y, Cortot A, Modigliani R, Laurent-Puig P, Gower-Rousseau C, Macry J, Colombel JF, Sahbatou M, Thomas G. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* 2001;**411**:599–603.
11 **Ogura Y**, Bonen DK, Inohara N, Nicolae DL, Chen FF, Ramos R, Britton H, Moran T, Karaliuskas R, Duerr RH, Achkar JP, Brant SR, Bayless TM, Kirschner BS, Hanauer SB, Nunez G, Cho JH. A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* 2001;**411**:603–6.
12 **Cuthbert AP**, Fisher SA, Mirza MM, King K, Hampe J, Croucher PJ, Mascheretti S, Sanderson J, Forbes A, Mansfield J, Schreiber S, Lewis CM, Mathew CG. The contribution of NOD2 gene mutations to the risk and site of disease in inflammatory bowel disease. *Gastroenterol* 2002;**122**:867–74.
13 **Economou M**, Trikalinos TA, Loizou KT, Tsianos EV, Ioannidis JP. Differential effects of NOD2 variants on Crohn's disease risk and phenotype in diverse populations: a meta-analysis. *Am J Gastroenterol* 2004;**99**:2393–04.
14 **Duerr RH**, Taylor KD, Brant SR, Rioux JD, Silverberg MS, Daly MJ, Steinhart AH, Abraham C, Regueiro M, Griffiths AM, Dassopoulos T, Bitton A, Yang H, Targan SR, Datta LW, Kistner EO, Schumm LP, Lee AT, Gregersen PK, Barmada MM, Rotter JI, Nicolae DL, Cho JH. A genome-wide association study identifies IL23R as an inflammatory bowel disease gene. *Science* 2006;**314**:1461–3.
15 **Hampe J**, Franke A, Rosenstiel P, Till A, Teuber M, Huse K, Albrecht M, Mayr G, de la Vega FM, Briggs J, Gunther S, Prescott NJ, Onnie CM, Hasler R, Sipos B, Folsch UR, Lengauer T, Platzer M, Mathew CG, Krawczak M, Schreiber S. A genome-wide association scan of nonsynonymous SNPs identifies a susceptibility variant for Crohn disease in ATG16L1. *Nat Genet* 2007;**39**:207–11.
16 **Prescott NJ**, Fisher SA, Franke A, Hampe J, Onnie CM, Soars D, Bagnall R, Mirza MM, Sanderson J, Forbes A, Mansfield J, Lewis CM, Schreiber S, Mathew CG. A non-synonymous SNP in ATG16L1 is a risk factor for ileal Crohn's disease and is independent of CARD15 and IBD5. *Gastroenterol* 2007;**132**:1665–71.
17 **Rioux JD**, Xavier RJ, Taylor KD, Silverberg MS, Goyette P, Huett A, Green T, Juballa P, Barmada MM, Datta LW, Shugart YY, Griffiths AM, Targan SR, Ippoliti AF, Bernard EJ, Mei L, Nicolae DL, Regueiro M, Schumm LP, Steinhart AH, Rotter JI, Duerr RH, Cho JH, Daly MJ, Brant SR. Genome-wide association study identifies new susceptibility loci for Crohn disease and implicates autophagy in disease pathogenesis. *Nat Genet* 2007;**39**:596–604.
18 **Tremelling M**, Cummings F, Fisher SA, Mansfield J, Gwilliam R, Keniry A, Nimmo ER, Drummond H, Onnie CM, Prescott NJ, Sanderson J, Bredin F, Berzuini C, Forbes A, Lewis CM, Cardon LR, Deloukas P, Jewell DP, Mathew CG, Parkes M, Satsangi J. IL23R variation determines susceptibility but not disease phenotype in Inflammatory Bowel Disease. *Gastroenterol* 2007;**132**:1657–64.
19 **Cummings SA**, Rubin DT. The complexity and challenges of genetic counselling and testing for inflammatory bowel disease. *J Genet Couns* 2006;**15**:465–76.
20 **Risch N**. Linkage strategies for genetically complex traits. II. Multilocus models. *Am J Hum Genet* 1990;**46**:222–8.
21 **Camp NJ**. Genomewide transmission/disequilibrium testing – consideration of the genotypic relative risks at disease loci. *Am J Hum Genet* 1997;**61**:1424–30.
22 **Franceschi S**, Panza E, La Vecchia C, Parazzini F, Decarli A, Bianchi Porro G. Nonspecific inflammatory bowel disease and smoking. *Am J Epidemiol* 1987;**125**:445–52.
23 **Tobin MV**, Logan RF, Langman MJ, McConnell RB, Gilmore IT. Cigarette smoking and inflammatory bowel disease. *Gastroenterol* 1987;**93**:316–21.
24 **Lindberg E**, Tysk C, Andersson K, Järnerot G. Smoking and inflammatory bowel disease. A case control study. *Gut* 1988;**29**:352–57.
25 **Calkins BM**. A meta-analysis of the role of smoking in inflammatory bowel disease. *Dig Dis Sci* 1989;**34**:1841–54.
26 **Regueiro M**, Kip KE, Cheung O, Hegazi RA, Plevy S. Cigarette smoking and age at diagnosis of inflammatory bowel disease. *Inflamm Bowel Dis* 2005;**11**:42–7.
27 **Brant SR**, Wang MH, Rawsthorne P, Sargent M, Datta LW, Nouvet FJ, Shugart YY, Bernstein CN. A population-based case-control of CARD15 and

other risk factors in Crohn's disease and ulcerative colitis. *Am J Gastroenterol* 2007;**102**:313–23.

28 **Orholm M**, Munkholm P, Langholz E, Nielsen OH, Sorensen IA, Binder V. Familial occurrence of inflammatory bowel disease. *N Engl J Med* 1991;**324**:84–8.

29 **Kyle J**. Crohn's disease in the northeastern and northern Isles of Scotland: an epidemiological review. *Gastroenterol* 1992;**103**:392–9.

30 **Lesage S**, Zouali H, Cezard JP, Colombel JF, Belaiche J, Almer S, Tysk C, O'Morain C, Gassull M, Binder V, Finkel Y, Modigliani R, Gower-Rousseau C, Macry J, Merlin F, Chamaillard M, Jannot AS, Thomas G, Hugot JP. CARD15/NOD2 mutational analysis and genotype-phenotype correlation in 612 patients with inflammatory bowel disease. *Am J Hum Genet* 2002;**70**:845–57.

31 **King K**, Sheikh MF, Cuthbert AP, Fisher SA, Onnie CM, Mirza MM, Pattni RC, Sanderson J, Forbes A, Mansfield J, Lewis CM, Roberts RG, Mathew CG. Mutation, selection, and evolution of the Crohn disease susceptibility gene CARD15. *Hum Mutat* 2006;**27**:44–54.

32 **Croucher PJ**, Mascheretti S, Hampe J, Huse K, Frenzel H, Stoll M, Lu T, Nikolaus S, Yank SK, Krawczak M, Kim WH, Schreiber S. Haplotype structure and association to Crohn's disease of CARD15 mutations in two ethnically divergent populations. *Eur J Hum Genet* 2003;**11**:6–16.

33 **Gaya DR**, Russell RK, Nimmo ER, Satsangi J. New genes in inflammatory bowel disease: lessons for complex diseases? *Lancet* 2006;**367**:1271–84.

34 **Pascoe L**, Zouali H, Sahbatou M, Hugot JP. Estimating the odds ratios of Crohn disease for the main CARD15/NOD2 mutations using a conditional maximum likelihood method in pedigrees collected via affected family members. *Eur J Hum Genet* 2007;**15**:864–71.

35 National Statistics – General Household Survey, 2005. *Smoking and drinking among adults, 2005*, 2006.http://www.statistics.gov.uk/downloads/theme_compendia/GHS05/GHS2005_SmokingandDrinking_Report.pdf.

36 **Maller J**, George S, Purcell S, Fagerness J, Altshuler D, Daly MJ, Seddon JM. Common variation in three genes, including a noncoding variant in CFH, strongly influences risk of age-related macular degeneration. *Nat Genet* 2006;**38**:1055–9.

37 **Schmidt S**, Hauser MA, Scott WK, Postel EA, Agarwal A, Gallins P, Wong F, Chen YS, Spencer K, Schnetz-Boutaud N, Haines JL, Pericak-Vance MA. Cigarette smoking strongly modifies the association of LOC387715 and age-related macular degeneration. *Am J Hum Genet* 2006;**78**:852–64.

38 **Satsangi J**, Grootscholten C, Holt H, Jewell DP. Clinical patterns of familial inflammatory bowel disease. *Gut* 1996;**38**:738–41.

39 **Cupples LA**, Farrer LA, Sadovnick AD, Relkin N, Whitehouse P, Green RC. Estimating risk curves for first-degree relatives of patients with Alzheimer's disease: the REVEAL study. *Genet Med* 2004;**6**:192–6.

40 **Roberts JS, Cupples LA, Relkin NR, Whitehouse PJ, Green RC, REVEAL (Risk Evaluation and Education for Alzheimer's Disease) Study Group**. Genetic risk assessment for adult children of people with Alzheimer's disease: the Risk Evaluation and Education for Alzheimer's Disease (REVEAL) study. *J Geriat Psychiat Neurol* 2005;**18**:250–5.

41 **Antoniou AC**, Pharoah PD, McMullan G, Day NE, Stratton MR, Peto J, Ponder BJ, Easton DF. A comprehensive model for familial breast cancer incorporating BRCA1, BRCA2 and other genes. *Br J Cancer* 2002;**86**:76–83.

42 **Brant SR**, McGovern DPB. NOD2, not yet. *Con Inflamm Bowel Dis,* 2005;**11**:507–9.

43 **Rubin DT**. To test or ''Nod-2'' test: What are the questions? The balanced viewpoint. *Inflamm Bowel Dis* 2005;**11**:510–12.

44 **Silverberg MS**. The time has come for NOD2/CARD15 testing for families with Crohn's disease. *Pro Inflamm Bowel Dis* 2005;**11**:504–6.

Submit an e*Letter*, and join the debate

eLetters are a fast and convenient way to register your opinion on topical and contentious medical issues. You can find the ''submit a response'' link alongside the abstract, full text and PDF versions of all our articles. We aim to publish swiftly, and your comments will be emailed directly to the author of the original article to allow them to respond. eLetters are a great way of participating in important clinical debates, so make sure your voice is heard.