Lepr Rev (2003) **74,** 133–140

Estimating the relative recurrence risk ratio for leprosy in Karonga District, Malawi

CHRIS WALLACE*, DAVID CLAYTON** & PAUL FINE*
**London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, UK*
***Cambridge Institute for Medical Research*

Accepted for publication 7 March 2003

Summary Leprosy is a chronic disease caused by infection with *Mycobacterium leprae*. Susceptibility to leprosy is influenced by both genetic and non-genetic factors and the disease is known to cluster in families. One measure of genetic effect is the relative recurrence risk ratio, λ_R. Estimates of this parameter can be inflated if environmental risk factors which also cluster in families, such as household contact, are not properly accounted for. We present the results of fitting a cross ratio model that allows estimation of the odds ratio of disease conditional on disease or no disease in a given relative, given measured covariates. From this model, we can predict fitted values for λ_R that represent the familial risk not accounted for by other covariates including observed household contact. If all covariates could be measured, this would be the 'genetic relative risk ratio'. We find $\lambda_R > 1$ for all relative pairs except grandparent-grandchild, and $\lambda_R > 2$ for siblings. Though not in itself evidence for a strong genetic susceptibility to leprosy, this result is consistent with much other evidence which suggests susceptibility to leprosy is under the control of many factors, the strongest of which may be non-genetic, with host genetics playing a small but significant role.

Introduction

Leprosy is a disease caused by infection with *Mycobacterium leprae*. Infection is necessary for disease, but it is thought that only a small proportion (10%?) of infections lead to clinical manifestations, which may be manifested across a spectrum from paucibacillary (PB) to multibacillary (MB) disease. The exact mode of transmission of *M. leprae* is unknown, but it is favoured by prolonged contact with an infected individual.

Correspondence to: C. Wallace, London School of Hygiene and Tropical Medicine, IDEU (2nd floor), Keppel Street, London WC1E 7HT, UK (e-mail: chris.wallace@lshtm.ac.uk)

0305-7518/03/064053+08 $1.00

Development of disease depends not only on infection, but also age and Bacille Calmette-Guerin (BCG) vaccination history. Age-specific risk peaks in teenagers and young adults, in many populations, and BCG vaccination reduces the risk of disease in all populations and in Malawi by approximately 50%.[1] There is also evidence that host genetics affect the development of disease. Many linkage and association studies have indicated the involvement of genes in the Major Histocompatability Complex (MHC) region on chromosome 6.[2,3] In addition, a recent linkage analysis of sibling pairs from an Indian population found strong evidence for susceptibility loci on chromosomes 10 and 20.[4,5]

Leprosy is known to cluster in families. This may be attributable to a combination of shared environmental risk factors and close contact facilitating transmission of *M. leprae* between family members, as well as to shared genetic susceptibility.

In this paper, we attempt to quantify the role that genetic factors play in family clustering after accounting for other, non-genetic, factors. We do this by estimating the relative recurrence risk ratio, λ_R, for cumulative incidence of leprosy, for first, second and third degree relative pairs. λ_R is defined as the ratio of the risk of disease in individuals who have an affected relative of type R to the risk of disease in the general population.

Mathematically, let X_1 and X_2 denote the disease state of two relatives (X_i if affected, 0 otherwise). Then

$$\lambda_R = \frac{P(X_1 = 1 \mid X_2 = 1)}{P(X_1 = 1)} = \frac{P(X_1 = 1 \text{ and } X_2 = 1)}{P(X_1 = 1)P(X_2 = 1)}.$$

$\lambda_R > 1$ indicates that relatives of affected individuals are at greater risk than the general population, which may be due to risk factors that cluster in families, including genetics.

Interest in λ_R was promoted by three seminal papers by Risch.[6–8] He showed that the power to detect genes responsible for variation in disease outcome (using linkage analysis methods) could be expressed as a function of λ_R. This enables estimation in advance of the power of a genetic linkage study for given sample sizes. Risch also showed how λ_R was expected to vary between relatives under different genetic models, which has led to studies which have used λ_R to discriminate between different genetic models.[9] But perhaps the most widespread use of this measure is in exclusion mapping, by which a region of genome can be 'excluded' from a search for regions that confer a particular locus-specific λ_R (for example, 1·5 or greater).[10]

Generally, λ_R has been estimated by sampling relatives of cases and comparing the estimated risk in the sample to an established population risk. This method is prone to ascertainment bias[11,12] and the estimates may be inflated when environmental factors that also affect risk of disease are ignored.[13]

In this paper, we model the marginal risk of disease for members of relative pairs for given covariates using a logistic model. We use a global cross-ratio model (CRM) to simultaneously model the ratio of the odds of disease, given that someone has an affected relative, to the odds of disease, given that s/he does not have an affected relative. We denote the marginal risks (u_1, u_2) and the odds ratio θ. From these estimated risks, we can estimate the probability that both members of a pair are affected, δ, and then λ_R using δ/u_1u_2. This model is an extension of a discrete realisation of Plackett's copula[14] and is a member of Dale's family of CRMs.[15] It is described in more depth in this context in another paper (Wallace and Clayton, in preparation). A stata package 'rrrest' for fitting this model has been written and is available from http: //www-gene.cimr.cam.ac.uk/clayton/software/stata/.

Data and methods

The data used here come from the Karonga Prevention Study (KPS), a large epidemiological study in Northern Malawi. They were collected between 1979 and 1989 during two total population surveys (LEP1 and LEP2) in Karonga district.[16]

All individuals were visited in their homes and were given a general examination. Those showing signs of possible current or past leprosy were given a further detailed examination. All new cases were biopsied. Slit-skin smears were taken from anyone with skin lesions which were considered as possibly due to MB leprosy and from anyone with signs of leprosy and a history of anti-leprosy treatment not already known to the control project. Diagnostic certainty (as 'certain', 'probable' or 'possible') was determined by an algorithm which has been described in detail.[17] Classification as MB was based upon bacteriological index > 1 on biopsy, slit-skin smear or nasal swab, or on historical grounds if no smear or biopsy was taken. All certain and probable cases are used in this analysis—i.e. we are considering cumulative incidence: was an individual currently affected, or had they ever been affected, by leprosy.

Careful recording of parents during the surveys enables us to connect extended pedigrees. We identified all possible relative pairs up to 3rd degree. We excluded great grandparent-grandchild and great aunt/uncle-niece/nephew pairs from the analysis because there were very few doubly affected pairs (2 and 13, respectively).

In this population, as in others, there is strong evidence that close contact is a risk factor for transmission of *M. leprae* and that sharing a household with an affected individual substantially increases risk of disease, particularly when a household is shared with an MB case.[18] Whether an individual shares a household with a leprosy case is likely to be strongly correlated between members of relative pairs, particularly among closer genetic relationships, e.g. siblings. Household contact is measured here by two covariates—whether an individual was observed to have shared a household with a PB or an MB leprosy case during either of the population surveys.

Some individuals do not know their exact date of birth; 15·4% (8·1% of men, 22·1% of women) gave an estimate based on a local events calendar.[19] These individuals were assigned at random to a year within their estimated age band, according to the distribution of birth years of those who gave a precise year. We included a variable in the model to record whether someone had given a birth year estimate or not because giving an estimated year of birth is more likely among people with lower socio-economic status, as is leprosy.

The copula is a function of distribution functions, which by definition are non-decreasing. Cumulative incidence is non-decreasing up to age 75 (as we would expect) but drops sharply after this. There is no evidence that members of this population aged over 75 years were less likely to have ever had leprosy (this is in principle unlikely); it is far more plausible that this is due to a combination of under-ascertainment, recovery and early mortality. Hypo-pigmented patches are not so visible on older skin, and neuromuscular deficit may be obscured by other conditions associated with ageing. Also, although leprosy itself is unlikely to cause early death (in this population, the risk ratio for early mortality is not significantly different from 1),[19] it is associated with low socio-economic status which is itself associated with higher mortality.[20] We therefore excluded the 2,473 individuals aged 75 or over because it is likely that cumulative leprosy incidence was not accurately measured in this group. The remaining 170,279 individuals were used in this analysis.

In the analysis, we used age bands broader than 5 years. These are shown by the

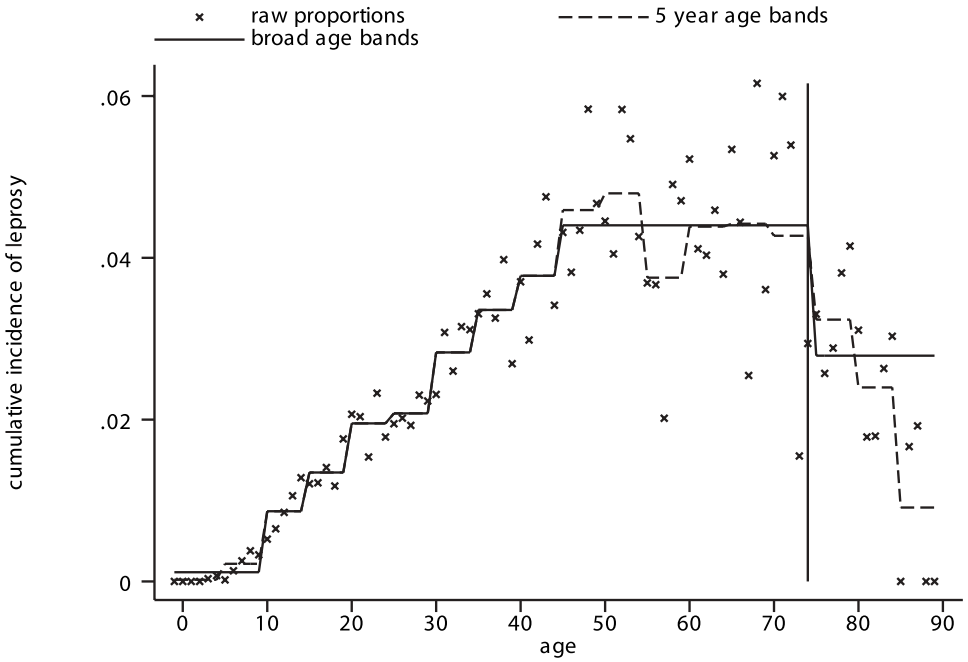

Figure 1. Cumulative leprosy incidence by age when last seen by the KPS. The vertical line shows the cutoff, and individuals 75 years or over were excluded from analysis.

continuous line in Figure 1. If an individual was seen more than once, we used the most recent observation to determine the age at which they were last seen by the KPS.

Initially, we used logistic regression to examine the effect of the covariates chosen. Non-genetic covariates are defined in Table 1. Then the CRM was fitted under the following four sets of conditions:

1. No covariates.
2. All covariates except household contact; θ constant
3. All covariates; θ constant.
4. All covariates; θ allowed to vary.

Under condition 4, θ was allowed to vary between those pairs with high and low non-genetic risk factors by dichotomising individuals according to their marginal predicted risks

Table 1. Definitions of covariates used in fitting the logistic and cross-ratio models

Covariate	Definition
bcg	BCG scar positive or negative
age	broad age group when last seen by the KPS
birest	whether a birth estimate or exact year of birth was given
sex	binary variable, 0 = male, 1 = female
pbcon	household contact with PB leprosy case
mbcon	household contact with MB leprosy case

Table 2. Number of relative pairs by affection status and relative type

	Number affected			
Relatives	0	1	2	Total
1st degree				
Siblings	189,620	5,122	207	194,949
Parent-child	215,770	9,777	236	225,783
2nd degree				
Half siblings	166,910	3,782	99	170,791
Aunt/uncle-niece/nephew	250,320	9,461	112	259,893
Godparent-godchild	156,070	8,622	73	164,765
3rd degree				
Cousins	678,620	16,225	230	695,075
Half aunt/uncle-niece/nephew	369,940	11,423	128	381,491

under condition 3 (within relative types). In part, this was to check whether the assumption implicit under condition 3 (θ constant) was valid. If it was not, it was hoped this would enable us to distinguish whether the genetic effect was stronger among those with lower or higher non-genetic risk factors. The cutoff chosen was the median marginal risk among affecteds estimated under condition 3. It was hoped this would provide most power by placing equal numbers of affecteds in each group. Pairs were then categorised as low-low, low-high or high-high. When estimating the model, no significant difference could be found between the low-high and high-high pairs, so the two groups were combined into a single high risk group.

Results

The numbers of relative pairs after exclusions are shown in Table 2. The graph in Figure 1 shows the relationship between age and cumulative leprosy incidence. Raw counts vary considerably, but grouping into 5 year bands (dashed line) allows a clearer picture to be seen.

LOGISTIC REGRESSION

Incidence of leprosy varies by sex, and in this population there is a slight excess risk for women (RR = 1·15, $P < 0{\cdot}001$).[1] After adjusting for age and birth estimates (which are given more often by women), this excess was no longer significant, and so sex has not been included in the final λ_R analysis. There was a significant interaction between age and birest, but not between any other covariates. The results (not shown) of fitting the logistic model

$$Y = \text{age} \times \text{birest} + \text{pbcon} + \text{mbcon} + \text{bcg} + \text{sex}$$

(where Y is disease status) were as expected: the risk of ever having leprosy increased with age, and after adjusting for age, increased risk of disease was associated with not giving an exact year of birth (which is associated with lower socio-economic status) (OR = 2, 95% CI 1·8–2·2), sharing a household with a PB case (OR = 2·1, 95% CI 1·9–2·2) or an MB case (OR = 3·0, 95% CI 2·6–3·5); BCG was protective (OR = 0·6, 95% CI 0·5–0·7). These results are consistent with previous studies in this population.[1,18]

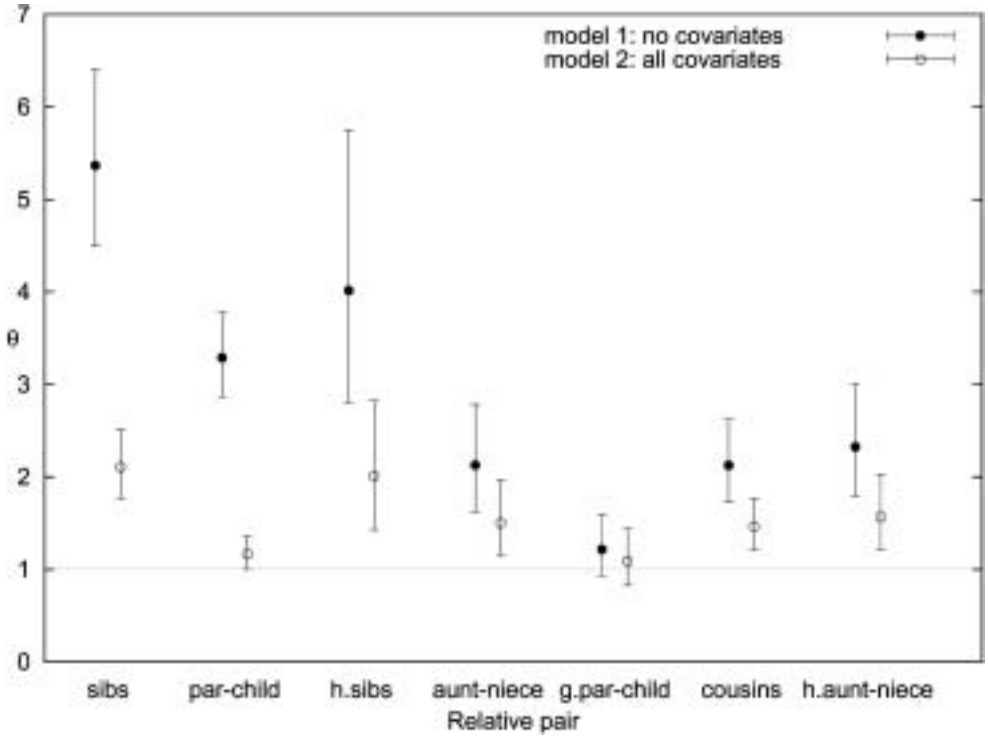

Figure 2. Estimates of θ and 95% confidence intervals from the cross ratio model (CRM) by relative type and inclusion/exclusion of covariates.

CROSS-RATIO MODEL

There was no evidence that θ varied when estimated under condition 4, so results are presented for conditions 1 to 3 only, under which θ is held constant. The parameters in the marginal logistic model were similar to the parameters in the logistic model fitted to the entire population, which indicated the marginal model was fitting well. It is easier to present results of the cross-ratio model in terms of θ rather than λ_R. This is because while θ is a parameter in the model, and thus can take only a limited number of values, λ_R is a fitted value. Note that this makes sense: if a disease is due to a combination of genetic and non-genetic factors, the risk of disease, given a relative has disease, will depend on an individual's covariates and so will vary for each individual. The estimates of θ and the 95% confidence intervals under conditions 1 to 3 are shown in Figure 2.

Discussion

Fitting the CRM and ignoring covariates (condition 1) gives quite large estimates of θ (and hence λ_R). These do not fall under condition 2, indicating that the apparent familial aggregation is not due to clustering of covariates such as BCG vaccination between relatives. They fall under condition 3, indicating that a large part of the apparent aggregation can be explained by common exposure and that the high values are the results of unmeasured

Table 3. Correlation of covariates between relative pairs and proportion of relative pairs observed to share a household during LEP1 or LEP2

Relative pair	Pairwise correlation coefficient (ρ) for						% sharing same household
	Sex	Bcg	Birest	Age	mbcon	pbcon	
Siblings	0·003	0·186	0·372	0·811	0·604	0·592	67·0
Parent	0·005	0·016	0·204	0·196	0·581	0·614	77·9
Half-siblings	0·018	0·088	0·003	0·483	0·485	0·443	44·3
Aunt/uncle	0·004	0·015	0·018	0·474	0·145	0·162	12·7
Godparent	0·005	0·016	0·061	0·115	0·239	0·240	21·9
Cousins	0·003	0·057	0·118	0·377	0·081	0·082	4·0
Half-aunt/uncle	0·004	0·025	0·007	0·362	0·130	0·122	7·1

non-genetic as well as genetic risk factors. Estimates of θ are constant within relative pairs under condition 4 (not shown), indicating that the fit of the model cannot be improved by allowing θ to be non-constant.

The significant fall in θ (and hence λ_R) values as more covariates are included is due largely to the clustering of these covariates within relative pairs. The pairwise correlation coefficients for these covariates are given in Table 3, together with the proportion of relative pairs observed to share a household with each other. Note that it is the covariates which measure household contact with a leprosy case which are are most strongly correlated, explaining why the majority of the familial aggregation of disease found under model 1 can be explained by household sharing of close relatives.

In all analyses, θ is higher for sibling pairs than for any other relationship, as predicted by theoretical genetic models.[21] If θ is measuring purely genetic risk, we would also expect to see θ decline consistently across degree of relationship, as the expected proportion of genes shared decreases from 1/2 to 1/4 and then 1/8. In model 1, this does not happen, presumably because of the unmeasured non-genetic factors. In model 3, λ_R is much higher for siblings than parents which indicates epistasis (interaction) between different genetic effects. It is possible that the residual risk observed could be accounted for by other, unmeasured, non-genetic factors, which are also more likely to be shared by siblings and half siblings who tend to share all their early lives that by parents and their children.

These data included the major risk factors that are likely to cluster in families given that the household contact variable should correlate with environmental factors associated with index cases. However, we realise there was likely to be undetected household contact,[19] given that our measure of contact was based of just two surveys over ten years. Given the potentially long incubation period of leprosy (measured in years and even decades) and that we are considering cumulative incidence we will certainly have missed earlier household contact, which will be important in older cases. The results must thus be interpreted with some caution, and it is likely that our estimates of λ_R would be further reduced if complete life contact histories were available. This may also explain why we do not see the expected pattern of falling λ_R with degree of relationship.

These results do not in themselves provide evidence for a strong genetic susceptibility to leprosy. However, they are consistent with much other evidence that susceptibility to leprosy is under the control of many factors, the strongest of which may be non-genetic, with host genetics playing a small but significant role.

Acknowledgements

The data analysed here were collected within the context of the Karonga Prevention Study KPS), which has been supported by the British Leprosy Relief Association (LEPRA) and the Wellcome Trust, with contributions from the International Federation of Anti-Leprosy Organisations (ILEP) and the WHO/UNDP/World Bank Special Programme for Research and Training in Tropical Diseases. We wish to acknowledge the special contributions of Drs J. M. Ponnighaus, D. K. Warndorff and A. C. Crampin, and more than 50 field staff for the fieldwork, Professor A. V. S. Hill and Dr J. Fitness for helpful discussions and the Health Sciences Research Committee of Malawi for encouraging the KPS programme. C. W. was supported by the British Medical Research Council.

References

[1] Pönnighaus JM, Fine PE, Sterne JA *et al*. Incidence rates of leprosy in Karonga District, Northern Malawi: patterns by age, sex BCG status and classification. *Int J Lepr Other Mycobact Dis*, 1994; **62:** 10–23.
[2] Todd JR, West BC, McDonald JC. Human leukocyte antigen and leprosy: study in Northern Louisiana and review. *Rev Infect Dis*, 1990; **12:** 63–74.
[3] Fine PE. Implications of genetics for the epidemiology and control of leprosy. *Philos Trans R Soc Lond B*, 1988; **321:** 365–376.
[4] Siddiqui MR, Meisner S, Tosh K *et al*. A major susceptibility locus for leprosy in India maps to chromosome 10p13. *Nat Genet*, 2001; **27:** 439–441.
[5] Tosh K, Meisner S, Siddiqui MR *et al*. A region of chromosome 20 is linked to leprosy susceptibility in a South Indian population. *J Infect Dis*, 2002; **186:** 1190–1193.
[6] Risch N. Linkage strategies for genetically complex traits. I. Multilocous models. *Am J Hum Genet*, 1990; **42:** 222–228.
[7] Risch N. Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *Am J Hum Genet*, 1990; **42:** 229–241.
[8] Risch N. Linkage strategies for genetically complex traits. III. The effect of marker polymorphism on analysis of affected relative pairs. *Am J Hum Genet*, 1990; **42:** 242–253.
[9] Rich SS. Mapping genes in diabetes: genetic epidemiological perspective. *Diabetes*, 1990; **39:** 1315–1319.
[10] Duffy DL, Montgomery GW, Hall J *et al*. Human twinning is not linked to the region of chromosome 4 syntenic with the sheep twinning gene *fecb*. *Am J Med Genet*, 2001; **100:** 182–186.
[11] Olson JM, Cordell HJ. Ascertainment bias in the estimation of sibling genetic risk parameters. *Genet Epidemiol*, 2000; **18:** 217–235.
[12] Guo SW. Inflation of sibling recurrence-risk ratio, due to ascertainment bias and/or overreporting. *Am J Hum Genet*, 1998; **63:** 252–258.
[13] Guo SW. Familial aggregation of environmental risk factors and familial aggregation of disease. *Am J Epidemiol*, 2000; **151:** 1121–1131.
[14] Plackett R. A class of bivariate distributions. *J Am Statist Assoc*, 1965; **60:** 516–522.
[15] Dale JR. Global cross-ratio models for bivariate, discrete, ordered responses. *Biometrics*, 1986; **42:** 909–917.
[16] Pönnighaus J, Fine PE, Bliss L *et al*. The Lepra Evaluation Project (LEP) and epidemiological study of leprosy in Northern Malawi. I: methods. *Lepr Rev*, 1987; **52:** 359–375.
[17] Pönnighaus JM, Fine PE, Bliss L. Certainty levels in the diagnosis of leprosy. *Int J Lepr Other Mycobact Dis*, 1987; **55:** 454–462.
[18] Fine PE, Sterne J, Pönnighaus JM *et al*. Household and dwelling contact as risk factors for leprosy in Northern Malawi. *Am J Epidemiol*, 1997; **146:** 91–102.
[19] Chirwa TF. *Effect of household dynamics on risk of disease associated with household contact*. PhD thesis, University of London, 2001.
[20] Pönnighaus JM, Fine PE, Sterne JA *et al*. Extended schooling and good housing conditions are associated with reduced risk of leprosy in rural Malawi. *Int J Lepr Other Mycobact Dis* 1994; **62:** 345–352.
[21] Koivisto M, Mannila H. Offspring risk and siblings risk for multilocus traits. *Hum Hered*, 2001; **51:** 209–216.