AD_____

Award Number: DAMD17-00-1-0546

TITLE: Statistical Methods for Analysis of NF Clinical Data

PRINCIPAL INVESTIGATOR: Harry Joe, Ph.D.

CONTRACTING ORGANIZATION: The University of British Columbia
Vancouver, British Columbia
Canada V6T 1Z3

REPORT DATE: August 2003

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 2003	Annual (1 Aug 2002 - 31 Jul 2003)

4. TITLE AND SUBTITLE
Statistical Methods for Analysis of NF Clinical Data

5. FUNDING NUMBERS
DAMD17-00-1-0546

6. AUTHOR(S)
Harry Joe, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
The University of British Columbia
Vancouver, British Columbia
Canada V6T 1Z3

E-Mail: harry@stat.ubc.ca

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

This project describes research in statistical methods that would be useful for statistical modeling and analysis of clinical data from NF1 and NF2 subjects. The statistical methods are classified into the areas:
(a) estimation of familial correlation for different types of data,
(b) assessment of multi_hit mutation models for incidence of tumours.

Some of the statistical methods to be developed are either new or partly new and require further research for computer software implementation.

Clinical data exist in many formats including binary, categorical, count, and continuous information. Furthermore, a common "real life" problem is censored data (where the beginning or end point is not known for all cases but some intermediate data exist). One goal of the project is to produce a software package for familial data analysis for different analysis for different types of data, such as binary, count, and censored survival data.

14. SUBJECT TERMS
Familial associations, genotype_phenotype correlation, multi_hit mutation models.

15. NUMBER OF PAGES
64

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

INTRODUCTION

This project describes research in statistical methods that would be useful for statistical modelling and analysis of clinical data from NF1 and NF2 subjects. The statistical methods are classified into the areas:
(a) estimation of familial correlation for different types of data,
(b) assessment of multi-hit mutation models for incidence of tumours.

Some of the statistical methods to be developed are either new or partly new and require further research for computer software implementation.

Clinical data exist in many formats including binary, categorical, count, and continuous information. Furthermore, a common "real life" problem is censored data (where the beginning or end point is not known for all cases but some intermediate data exist). One goal of the project is to produce a software package for familial data analysis for different types of data, such as binary, count, and censored survival data.

BODY

Purpose of the project:

(A) To develop statistical methods that can be used to characterize the phenotype of individuals with NF1 and NF2.

(B) To develop methods to elaborate on the standard two-hit model of tumour formation taking into account additional pathogenic factors and allelic differences for tumours in NF1 and NF2.

The research accomplishments associated with each objective from the statement of work are summarized below.

Objective 1. Develop statistical methods for interval-censored data, and obtain estimates of age of onset distributions for NF1 and NF2 features, using longitudinal information in the databases.

This objective is being postponed as we currently do not have enough longitudinal information in the databases.

Objective 2. Develop statistical methods for familial correlations for non-continuous and censored data, and obtain estimates of intraclass and interclass correlations for quantitative and binary traits in NF1 and NF2.

Estimation of familial correlations in clinical traits is related to the assessment of familial aggregation in genetic diseases. It is important to our understanding of the causes of variable expressivity in Mendelian diseases.

Much progress in the past year, in the published papers and a near completed PhD thesis has been in this objective. A summary of the statistical aspects of the published papers was given in last year's annual report. Some extracts from the forthcoming thesis of Yinshan Zhao, specifically, Section 1.4 with an outline by chapter, and a section with a summary of the new estimation methods are given in an Appendix. Yinshan Zhao's PhD committee received a first draft of the complete thesis in August 2003. It is planned to have the thesis defended before the end of the year 2003.

We give a short overview of Zhao's thesis research. This work involves the theory for various estimating equation approaches that should lead to more reasonable computing time for estimating familial associations for traits in the form of right-censored survival or count data, etc. An example of a count variable is the number of tumors of a particular type; an example of a

right-censored survival variable is the age of onset of a particular disease feature - those who do not have the feature are right-censored at the age of last clinical observation.

For most of the multivariate models for familial data, such as those with a latent multivariate normal distribution, the maximum likelihood approach is not computationally tractable when high-dimensional numerical integration is involved. It is desirable to develop other estimating approaches which are computationally less demanding, and relatively efficient (in the sense of variance of the sampling distribution of the estimator). Estimation using two types of composite likelihood equations are considered; these are based on the likelihoods of the univariate and bivariate margins of the multivariate model. They are called two-stage and bivariate composite likelihood in the Appendix.

The estimates of the model parameters based on composite likelihood methods are asymptotically consistent and unbiased. There are three types of parameters in the models for familial data: regression or covariate coefficients, dispersion parameters and dependence parameters. The
estimates are much easier to compute under many circumstances compared to the maximum likelihood estimates. The major concern of these approaches is the efficiency. Comparisons of the relative efficiency of these two approaches against the maximum likelihood method have been made, with summaries given in the Appendix. Different models, including multivariate normal, multivariate probit, lognormal-Poisson mixture and multivariate lognormal with right censoring, have been examined analytically or by Monte Carlo simulation.

Other work done by a research assistant Lisa Kuramoto is summarized below. We summarize the new developments in statistical methodology, particularly in the modelling of familial count data.

Genotype-phenotype analyses, accounting for the familial associations, were done for absence/presence of cataracts and some count variables for the NF2 database; count variables included the number of spinal tumors, number of meningiomas, and number of cutaneous Schwannomas for NF2 patients. The main analyses used for a manuscript, to be written and submitted, was based on the negative-binomial gamma mixture model proposed in Zhao et al (2002). The main results of the analyses are summarized in the Abstracts in the Appendix?

In addition, for comparisons, we also fitted a zero-inflated multivariate Poisson-log normal model [the latter from Aitchison and Ho (1989)]. Zero inflation refers to a standard probability distribution for count data, with extra probability assigned to a count of zero. If the count variable refers to a tumor count, zero inflation is reasonable for statistical modelling when the high observed frequency of zero counts includes a subset of subjects with a mild form of the disease.

The zero-inflated multivariate Poisson-log normal model has more flexibility than the negative-binomial gamma mixture model in terms of familial dependence. Without the zero inflation for the multivariate Poisson-log normal model, we found that the fit to the count

variables was not very good; this is because the count variables are heavily right skewed (some large counts of the order of 50) and have a high frequency of zero. The zero inflation for the Poisson-log normal model adds a third univariate parameter to the marginal count distribution, resulting in as many univariate non-regression parameters as the negative-binomial gamma mixture model. Because of the shape of the histogram of the count variables for the NF2 database, univariate marginal distributions with three non-regression parameters are needed to provide an adequate fit.

Objective 3. Fit multi-hit mutation models for the incidence of NF2 and NF1 tumours by age, distinguish whether a two-hit or three-hit model provides a better fit to the data, and adapt the models to account for mutation type and other factors.

Two- and three-hit models are vestibular schwanomas were fit to data for NF2 subjects; this was published in Genetic Epidemiology in 2003 [Woods et al 2003]. With the latest NF2 database with more data on mutation type, we still plan to further check on the strength of the genotype-phenotype correlations when fitting the two-hit and three-hit models for vestibular schwanomas separately for several mutation types.

An additional task, not part of the original objectives but which relates to objective 3, was providing support to MSc student, Bernard Lee, for a project to determine areas of *NF1* involved in transcriptional regulation. This support included providing expertise in interpreting algorithms and bioinformatic analyses as well as sequence alignment analyses for comparative genomics. An understanding of this region of the NF1 gene and mutations that can possibly arise in the transcriptional control region is crucial to adequate classification of mutations for genotype-phenotype mutational studies. Please see the appendix for an abstract describing this research.

Objective 4. Write C code to implement all of these statistical methods and provide a user-friendly interface for the code.

Software written in C/C++, is being developed in Unix/Linux; it runs also in Windows with Cygnus/Gnuwin [see www.cygwin.com], the public domain version of Unix for Windows. The implementation of the interface currently is through control files which specifies parameters and data files. By the end of 2001, methods for binary and quantitative (continuous) traits had been integrated into a computer package and this was used in the statistical analysis in Szudek et al (2002). Later it was used in the analysis for the presence/absence of cataracts in Baser et al (2003).

The additional modules, completed since the last annual report, are a module for the handling of familial survival data based on multivariate normal distribution (this module expanded on work of D Aeschliman reported last year), and modules for familial count data. For count data, the

models being implemented (referred to under Objective 2) are the negative-binomial gamma mixture model, and the zero-inflated multivariate Poisson-log normal models, as these have been found to be the most useful for tumor count data in NF2 patients.

New modules based the estimation methods from Zhao's PhD thesis research will be added after the thesis is completed.

The latest version of the software package will be put in the directory
ftp://ftp.stat.ubc.ca/pub/hjoe/famil/

KEY RESEARCH ACCOMPLISHMENTS

- Comparison of two- and three-hit models for onset time of vestibular schwanomas for NF2 subjects.

- Two-third completion of a software package for analysis of familial data of various types (binary, count, continuous, censored). The software written in the C/C++ programming languages, developed in Unix/Linux, runs also in Windows with Cygnus/Gnuwin (public domain version of Unix for Windows).

- Application of the software package for estimating familial associations for NF1 and NF2 clinical features, with adjustments for the age effect.

- Development of statistical methodology that will be further used to analyze NF databases in the future when there are more quantitative and longitudinal information.

REPORTABLE OUTCOMES

Papers appeared and accepted

Szudek J, Joe H, and Friedman JM (2002). Analysis of intra-familial phenotypic variation in neurofibromatosis 1 (Nf1) Genet Epid, 23, 150-164.

Zhao Y, Kumar RA, Baser ME, Evans DGR, Wallace A, Kluwe L, Mautner VF, Parry DM, Rouleau GA, Joe H, Friedman JM (2002). Intrafamilial correlation of clinical manifestations in neurofibromatosis 2 (NF2) Genet Epid, 23, 245-259.

Baser ME, Friedman JM, Aeschliman D, Joe H, Wallace AJ, Ramsden RT, Evans DGR (2002). Predictors of the risk of mortality in neurofibromatosis 2. Am J Hum Genet, 71, 715-723.

Baser ME, Friedman JM, Wallace AJ, Ramsden RT, Joe H, Evans DGR (2002). Evaluation of clinical diagnostic criteria for neurofibromatosis 2. Neurology, 59(11), 1759-65.

Woods R, Friedman JM, Evans DGR, Baser ME, and Joe H (2003). Exploring the '2-hit hypothesis' in NF2: tests of 2-hit and 3-hit models of vestibular schwannoma development. Genet Epid, 24, 265-272.

Baser ME, Joe H, Kuramoto L, Friedman JM, Wallace AJ, Ramsden RT, Evans DGR (2003). Genotype-phenotype correlations for cataracts in neurofibromatosis 2. J Medical Genetics, accepted Apr 2003. (Preprint in appendix 1)

Palmer V, Szudek J, Joe H, Riccardi VM, and Friedman JM (2003). Analysis of neurofibromatosis 1 (nf1) lesions by body segment. Accepted for publication. (Preprint in appendix 1)

Papers submitted

Joe H and Latif AHMM (2003). Familial analysis of binary traits. Submitted to a statistics journal. (Preprint in appendix 1)

Abstracts accepted in 2003

Baser ME, Woods R, Joe H, Kuramoto L, Friedman JM, Wallace AJ, Bijlsma E, Olschwang S, Papi L, Parry DM, Ramsden RT, Rouleau GA, Evans DGR. The location of constitutional neurofibromatosis 2 (*NF2*) splice-site mutations is associated with the number of intracranial meningiomas: results from an international NF2 database. NNFF International Consortium for the Molecular Biology of NF1 and NF2, 1-4 June 2003, Aspen (CO).

Baser ME, Joe H, Kuramoto L, Friedman JM, Wallace AJ, Ramsden RT, Evans DGR. Genotype-phenotype correlations for cataracts in neurofibromatosis 2. NNFF International Consortium for the Molecular Biology of NF1 and NF2, 1-4 June 2003, Aspen (CO).

Baser ME, Parry DM, Joe H, Kuramoto L, Friedman JM, Gillespie JE, Wallace AJ, Ramsden RT, Evans DGR. Genotype-phenotype correlations for spinal tumors in neurofibromatosis 2. NNFF International Consortium for the Molecular Biology of NF1 and NF2, 1-4 June 2003, Aspen (CO).

Baser ME, Joe H, Kuramoto L, Friedman JM, Gillespie JE, Wallace AJ, Ramsden RT, Evans DGR. Genotype-phenotype correlations for peripheral nerve tumors in neurofibromatosis 2. NNFF International Consortium for the Molecular Biology of NF1 and NF2, 1-4 June 2003, Aspen (CO).

Ramsden RT, Evans DGR, Wallace AJ, Joe H, Baser ME. Revised diagnostic criteria for neurofibromatosis 2. 53rd Annual Meeting, American Society of Human Genetics, 4-8 November 2003, Los Angeles (CA). Accepted.

CONCLUSIONS

As discussed in the previous year's report, progress on objective 1 has been limited by data availability. We will complete this objective if data become available. At the current time, more data are becoming available on NF2 as the new NF2 genotype-phenotype database in the Friedman Lab at UBC, in Vancouver become populated.

For objectives 2 and 3, the theory for simplest cases has been mostly developed. The coding into C programs and use on current NF1/NF2 databases (objective 4) has been done for the statistical methods, but not have been implemented into the software package. Both objectives have also been expanded somewhat from the original plan of work.

To date, seven manuscripts have been accepted for publication, and ~~two~~ one others have been submitted. Presentations were made at the 2000-2002 meetings of the American Society of Human Genetics and are planned for the 2003 meeting. In addition this year, five presentations were made at the NNFF International Consortium for the Molecular Biology of NFl and NF2, at the Fourth International Conference on Vestibular Schwannoma and Other CPA Lesions, Cambridge UK, and at the 10th European Neurofibromatosis Meeting, Turku (Finland). The more statistically theoretical papers based on Zhao's thesis will be written after the completion of her thesis. We are behind schedule because of the PhD research; in advance it is hard to predict how quickly PhD students can accomplish things.

REFERENCES

(other references are in the reportable outcomes section)

Aitchison, J and Ho, CH (1989). The multivariate Poisson-log normal distribution. Biometrika 76: 643-653.

Godambe, VP (ed.) (1991). Estimating Functions. Oxford University Press, Oxford.

Lindsay, BG (1988). Composite likelihood methods. In Statistical Inference from Stochastic Processes, ed. by Prabhu, NU, American Mathematical Society: RI, pp. 221-239.

Xu, JJ (1996). Statistical Modelling and Inference for Multivariate and Longitudinal Discrete Response Data. Ph.D. thesis, Department of Statistics, University of British Columbia.

APPENDICES

1. Preprints
    (a) Genotype-phenotype correlation for cataracts in NF2.
    (b) Familial analysis of binary traits
    (c) Analysis of neurofibromatosis lesions by body segment
2. Outline for Zhao's thesis
3. Excerpt from summary section on estimation methods from Zhao's thesis.
4. 2003 published or accepted abstracts
    (a) Phylogenetic Footprinting of the NF1 5'' Upstream Region (5UR).
    (b) The location of constitutional neurofibromatosis 2 (*NF2*) splice-site mutations is associated with the number of intracranial meningiomas: results from an international NF2 database.
    (c) Genotype-phenotype correlations for cataracts in neurofibromatosis 2.
    (d) Genotype-phenotype correlations for spinal tumors in neurofibromatosis 2
    (e) Genotype-phenotype correlations for peripheral nerve tumors in neurofibromatosis 2
    (f) Revised diagnostic criteria for neurofibromatosis 2.

Appendix 1 (Preprints)

Genotype-phenotype correlation for cataracts in NF2
Familial analysis of binary traits
Analysis of neurofibromatosis lesions by body segment

Genotype-phenotype correlations for cataracts in neurofibromatosis 2

EDITOR - Neurofibromatosis 2 (NF2) is an autosomal dominant disease that is caused by inactivating mutations of the *NF2* tumor suppressor gene.[1,2] Multiple central and peripheral nervous system tumors and ocular abnormalities are common in NF2; bilateral vestibular schwannomas are pathognomonic for the disease. Genotype-phenotype correlations are well-established for NF2-associated tumors. In general, constitutional nonsense or frameshift *NF2* mutations are associated with severe NF2 (i.e., earlier onset of symptoms and more tumors), splice-site mutations with variable disease severity, and missense mutations with mild disease.

Genotype-phenotype correlations have not been demonstrated for the non-tumor manifestations of NF2. The most common of these manifestations is presenile cataracts (posterior subcapsular and cortical), which occur in about 60-80% of people with NF2.[3-5] In animal models, lens fiber cells that are more differentiated express less *Nf2* protein than the epithelial regions of the lens, suggesting that the *Nf2* protein may play a role in lens epithelial cell migration or elongation.[6] The purpose of this study was to determine if there were genotype-phenotype correlations for cataracts in NF2.

The study was based on the United Kingdom NF2 registry in the Department of Medical Genetics, St. Mary's Hospital, Manchester. NF2 patients are ascertained by contacting neurosurgeons, otolaryngologists, neurologists, pediatricians, dermatologists, and geneticists throughout the United Kingdom, augmented in the North West Region by the Regional Cancer Registry. The study was subject to continuing ethics committee evaluation and subjects consented to participation. Patients were screened for constitutional *NF2* mutations using single-strand conformational polymorphism analysis (SSCP) as previously described,[7] and examined

for cataracts using slitlamp biomicroscopy at the time of diagnosis of NF2. For this study, cataracts were defined as present or absent (i.e., posterior subcapsular cataracts and cortical cataracts were aggregated). There were 255 people from 190 families (159 people with new mutations and 96 inherited cases; 132 females and 123 males) (Table 1).

For univariate analyses, Fisher's exact test was used for binary variables and the two-tailed t-test for continuous variables. A multivariate probit model with an exchangeable correlation structure within families was used with various sets of covariates to account for possible familial dependence.[8] From a regression coefficient β, an approximate relative risk (RR = $\exp\{2*\beta\}$) and confidence interval (CI) for presence of cataracts can be calculated. In the probit model, people with classical NF2 (i.e., who met the Manchester clinical diagnostic criteria for NF2[9]) and constitutional nonsense or frameshift *NF2* mutations were the reference group in comparisons between people with different types of *NF2* mutations.

There is a potential bias toward a lower age at onset of symptoms or age at diagnosis in inherited cases due to the family history of the disease. In the study group as a whole, there were no significant differences in these ages between people with new mutations and inherited cases for any type of *NF2* mutation. Also, using a probit model, the RR of cataracts was not significantly associated with age at diagnosis (see below). Therefore, for all mutation categories except unfound mutations, we combined people with new mutations and inherited cases. In the large group of people with unfound mutations, we retained the division between those with new mutations and inherited disease because people with new unfound mutations may be somatic mosaics. We used age at onset of symptoms to categorize people with new unfound mutations by disease severity (severe disease, onset of symptoms at ages < 20 years; mild disease, onset of symptoms at ages ≥ 20 years).[3]

As expected, the mean age at onset of symptoms and age at diagnosis were higher in people with non-truncating mutations and in somatic mosaics than in people with classical NF2 and nonsense or frameshift mutations (Table 1). The overall prevalence of cataracts was 33%, but the prevalence of cataracts was significantly lower in somatic mosaics and in people with new unfound mutations and onset of symptoms at ages > 20 years than in people with classical NF2 and nonsense or frameshift mutations. In people with cataracts, 29% were diagnosed with cataracts at ages < 10 years, and 47% at ages < 20 years (mean $\pm$ SE, 23 $\pm$ 2 years). Seventy per cent were diagnosed with cataracts before their first non-ocular sign or symptom.

In the multivariate probit model summarized in Table 2, the RR of cataracts did not significantly increase with increasing age at diagnosis, after accounting for the type of constitutional *NF2* mutation. In other probit models, the RR of cataracts also did not significantly increase with increasing age, after accounting for the type of constitutional *NF2* mutation (data not shown). This is probably due to the relatively young study population (mean $\pm$ SE age at diagnosis, 28 $\pm$ 1 years; only 5% diagnosed at ages > 55 years and 2% at ages > 60 years), since the prevalence of posterior subcapsular and cortical cataracts in people aged < 55 years in the general population is very low.[10]

The RR and estimated prevalence of cataracts was lower in all mutation groups as compared to people with classical NF2 and nonsense or frameshift mutations. This difference was statistically significant in somatic mosaics (RR = 0.20, 95% CI = 0.10 - 0.40), in people with large deletions (RR = 0.39, 95% CI = 0.16 - 0.98), and in people with new unfound mutations and onset of symptoms at ages $\geq$ 20 years (RR = 0.09, 95% CI = 0.03 - 0.28). The RR of cataracts in people with missense mutations was low but not statistically significant (RR = 0.38,

95% CI = 0.14 - 1.08). The lower RR of cataracts in each of these groups is consistent with the generally mild disease in NF2 patients with these types of mutations or conditions.

The lower RR of cataracts in people with new unfound mutations and mild disease could be due to several types of mutations or conditions that are unlikely to be identified by SSCP, and that are known to be associated or likely to be associated with mild NF2. These mutations or conditions are somatic mosaicism; large deletions, insertions, or other rearrangements; mutations in the 3 or 5 untranslated regions, the promoter region, or untranscribed transcriptional control elements; intronic mutations that are not covered by conventional SSCP primers; or other epigenetic events causing loss of *NF2* expression, such as methylation.

Somatic mosaicism and large deletions are the most likely of these possibilities. In the present study, 17 (18%) of the 92 patients with new mutations and identified constitutional *NF2* mutations were somatic mosaics. The estimated prevalence of somatic mosaicism in NF2 patients with new mutations is 25-30%.[11,12] Some of the 41 NF2 patients with new unfound mutations and mild disease may be somatic mosaics in whom conventional DNA sequencing of lymphocyte DNA PCR product has failed to identify a difference from the normal sequence because the mutant allele is present at too low a level to be detected. Constitutional *NF2* large deletions have been found in 21% of NF2 families using microarray-comparative genomic hybridization,[13] and in 32% of NF2 families using multiple mutation screening methods.[14]

The intrafamilial correlation for cataracts was weak (and statistically insignificant) in all multivariate probit models that were tried, although there were relatively few families with multiple affected relatives. Several other clinical features of NF2 (age at onset of symptoms, age at diagnosis, and number of intracranial meningiomas) have strong familial correlations.[15] The prevalence of cataracts in the present study was lower than in other studies,[3-5] probably because

the population-based United Kingdom NF2 registry is less heavily weighted toward NF2 patients with severe disease than studies that are based on patients from tertiary referral clinics,[4,5] and because some cataract examinations were done by medical specialists other than ophthalmologists. Non-ophthalmologists may miss faint cataracts, but in such cases, it is unlikely that faint cataracts are missed more frequently in people with mild NF2 than in those with severe NF2 (i.e., it will not bias genotype-phenotype correlations). In one large study, all patients were examined using slitlamp biomicroscopy by a non-ophthalmologist, and the prevalence of cataracts was similar in mild cases (35%) and in severe cases (40%).[9]

The genotype-phenotype correlations for cataracts in the present study extend the correlations that have been reported for the tumor manifestations of NF2. The high prevalence of cataracts in young NF2 patients, and their frequent occurrence before the tumor manifestations of NF2, underscore the importance of non-8th nerve signs or symptoms of NF2 in children and adolescents as a useful aid to diagnosis in this age group.[16]

Michael E. Baser[1]

Harry Joe[2]

Lisa Kuramoto[2]

J. M. Friedman[3]

Andrew J. Wallace[4]

Richard T. Ramsden[5]

D. Gareth R. Evans[4]

[1]*Los Angeles, CA, U.S.A.,* [2]*Department of Biostatistics and* [3]*Department of Medical Genetics, University of British Columbia, Vancouver, B.C., Canada,* [4]*Department of Medical Genetics, St. Mary's Hospital, Manchester, U.K.,* [5]*Department of Otolaryngology, Manchester Royal Infirmary, Manchester, U.K.*

Correspondence to:

Michael E. Baser, Ph.D.

2257 Fox Hills Dr.

Los Angeles, CA 90064

baser@earthlink.net

Key words: NF2; neurofibromatosis 2; cataracts; genotype-phenotype

Acknowledgments

Supported in part by the FBT Foundation and U.S. Army grant U.S.A.R.M.C. NF990038.

References

1. Trofatter JA, MacCollin MM, Rutter JL, Murrell JR, Duyao MP, Parry DM, Eldridge R, Kley N, Menon AG, Pulaski K, Haase VH, Ambrose CM, Munroe D, Bove C, Haines JL, Martuza RL, MacDonald ME, Seizinger BR, Short MP, Buckler AJ, Gusella JF. A novel moesin-, ezrin-, radixin-like gene is a candidate for the neurofibromatosis 2 tumor suppressor [published erratum appears in *Cell* 1993;**75**:826]. *Cell* 1993;**72**:791-800.

2. Rouleau GA, Merel P, Lutchman M, Sanson M, Zucman J, Marineau C, Hoang-Xuan K,

Demczuk S, Desmaze C, Plougastel B, Pulst SM, Lenoir G, Bijlsma E, Fashold R, Dumanski J, de Jong P, Parry D, Eldridge R, Aurias A, Delattre O, Thomas G. Alteration in a new gene encoding a putative membrane-organizing protein causes neurofibromatosis type 2. *Nature* 1993;**363**:515-21.

3. Parry DM, Eldridge R, Kaiser-Kupfer MI, Bouzas EA, Pikus A, Patronas N. Neurofibromatosis 2 (NF2): clinical characteristics of 63 affected individuals and clinical evidence for heterogeneity. *Am J Med Genet* 1994;**52**:450-61.

4. Mautner VF, Lindenau M, Baser ME, Hazim W, Tatagiba M, Haase W, Samii M, Wais R, Pulst SM. The neuroimaging and clinical spectrum of neurofibromatosis 2. *Neurosurgery* 1996;**38**:880-5.

5. Ragge NK, Baser ME, Klein J, Nechiporuk A, Sainz J, Pulst SM, Riccardi VM. Ocular abnormalities in neurofibromatosis 2. *Am J Ophthalmol* 1995;**120**:634-41.

6. Claudio JO, Veneziale RW, Menko AS, Rouleau GA. Expression of schwannomin in lens and Schwann cells. *Neuroreport* 1997;**8**:2025-30.

7. Evans DGR, Trueman L, Wallace A, Collins S, Strachan T. Genotype/phenotype correlations in type 2 neurofibromatosis (NF2): evidence for more severe disease associated with truncating mutations. *J Med Genet* 1998;**35**:450-5.

8. Ashford J, Sowden R. Multivariate probit analysis. *Biometrics* 1970;**26**:535-46.

9. Evans DGR, Huson SM, Donnai D, Neary W, Blair V, Newton V, Harris R. A clinical study of type 2 neurofibromatosis. *Q J Med* 1992;**84**:603-18.

10. Klein BE, Klein R, Linton KL. Prevalence of age-related lens opacities in a population. The Beaver Dam Eye Study. *Ophthalmology* 1992;**99**:546-52.

11. Kluwe L, Mautner VF, Heinrich B, Dezube R, Jacoby LB, Friedrich RE, MacCollin M. Molecular study of frequency of mosaicism in neurofibromatosis 2 patients with bilateral vestibular schwannomas. *J Med Genet* 2003;**40**:109-14.

12. Moyhuddin A, Baser ME, Watson C, Purcell S, Ramsden RT, Heiberg A, Wallace A, Evans DGR. Somatic mosaicism in neurofibromatosis 2: prevalence and risk of disease transmission to offspring. *J Med Genet.* In press.

13. Bruder CEG, Hirvelä C, Tapia-Paez I, Fransson I, Segraves, R, Hamilton G, Zhang XX, Evans DG, Wallace AJ, Baser ME, Zucman-Rossi J, Hergersberg M, Boltshauser E, Papi L, Rouleau GA, Poptodorov G, Jordanova A, Rask-Andersen H, Kluwe L, Mautner V, Sainio M, Hung G, Mathiesen T, Möller C, Pulst SM, Harder H, Heiberg A, Honda M, Niimura M, Sahlén S, Blennow E, Alberton DG, Pinkel D, Dumanski JP. High-resolution deletion analysis of constitutional DNA from neurofibromatosis type 2 (NF2) patients using microarray-CGH. *Hum Mol Genet* 2001;**10**:271-82.

14. Zucman-Rossi J, Legoix P, Sarkissian HD, Cheret G, Sor F, Bernardi A, Cazes L, Giraud S, Ollagnon E, Lenoir G, Thomas G. NF2 gene in neurofibromatosis type 2 patients. *Hum Mol Genet* 1998;**7**:2095-101.

15. Zhao Y, Kuman RA, Baser ME, Evans DGR, Wallace A, Kluwe L, Mautner VF, Parry DM, Rouleau GA, Joe H, Friedman JM. Intrafamilial correlation of clinical manifestations of in neurofibromatosis 2 (NF2). *Genet Epidemiol* 2002;**23**:245-59.

16. MacCollin M, Mautner VF. The diagnosis and management of neurofibromatosis 2 in childhood. *Sem Pediatr Neurol* 1998;**5**:243-52.

Table 1. Characteristics of study population, by type of *NF2* mutation (SD = standard deviation)

Characteristic	Type of *NF2* mutation[1]								
	Nonsense or frameshift		Splice-site	Missense	Large deletion	Unfound			Total
	Classical	Somatic mosaic				Inherited cases	People with new mutations Age at onset (years) < 20	People with new mutations Age at onset (years) ≥ 20	
No. of people/families	73/56	17/17	47/25	15/6	25/14	14/11	23/23	41/41	255/190
Age (years, mean ± SD)									
Onset of symptoms[2]	16 ± 9	32 ± 12	23 ± 12	30 ± 14	21 ± 9	25 ± 16	12 ± 6	34 ± 9	22 ± 12
Diagnosis	21 ± 12	38 ± 12	27 ± 15	38 ± 21	23 ± 9	29 ± 17	22 ± 12	40 ± 11	28 ± 15
Current	27± 12	44 ± 12	32 ± 16	45 ± 21	30± 11	35 ± 20	29 ± 12	46 ± 13	34 ± 16
Intracranial meningiomas (%)	56	59	34	27	52	14	70	51	48
Cataracts (%)	45	18	38	27	28	36	39	10	33

[1]All are constitutional mutations except the somatic mosaics
[2]Excludes 15 inherited cases who were asymptomatic at the time of diagnosis of NF2

Comparisons to people with classical NF2 and nonsense or frameshift mutations (P-values are computed based on an assumption of independence, which is violated to a slight degree due to families with multiple affected relatives):

Somatic mosaics: age at onset, age at diagnosis, current age, $P < .001$; cataracts, $P = .053$
Splice-site mutations: age at onset, $P = .001$; age at diagnosis, $P = .023$; current age, $P = .043$; intracranial meningiomas, $P = .024$
Missense mutations: age at onset, $P = .003$; age at diagnosis, $P = .006$; current age, $P < .001$; intracranial meningiomas, $P = .049$
Large deletions: age at onset, $P = .032$
Unfound mutations:
    Inherited cases: intracranial meningiomas, $P = .007$
    People with new mutations and age at onset ≥ 20 years: cataracts, $P < .001$

Table 2. Multivariate probit model for cataracts (RR = relative risk, CI = confidence interval, SE = standard error)

Covariate	Estimated prevalence of cataracts from model (%)	Parameter estimate (SE)	RR	95% CI
Age at diagnosis (per year)		0.15 (0.58)	1.00	0.98 - 1.02
Exchangeable dependence (familial correlation)		0.11 (0.15)		
Type of *NF2* mutation[1]				
Nonsense or frameshift				
Classical NF2	45	Reference group	Reference group	
Somatic mosaic	17	-0.82 (0.18)	0.20	0.10 - 0.40
Splice-site	39	-0.17 (0.19)	0.71	0.34 - 1.50
Missense	28	-0.48 (0.26)	0.38	0.14 - 1.08
Large deletion	28	-0.47 (0.23)	0.39	0.16 - 0.98
Unfound				
People with new mutations				
Age at onset < 20 years	39	-0.15 (0.15)	0.74	0.40 - 1.36
Age at onset ≥ 20 years	9	-1.20 (0.28)	0.09	0.03 - 0.28
Inherited cases	36	-0.23 (0.34)	0.63	0.16 - 2.45

[1]All are constitutional mutations except the somatic mosaics

Familial analysis of binary traits

Harry Joe and A. H. M. Mahbub-ul Latif
Department of Statistics
University of British Columbia

SUMMARY

For familial aggregation of a binary trait, we compare the GEE2 odds ratio regression or multivariate logit model with the multivariate probit model, and report on our computer implementations. One comparison is the conditional probability that one (future) member of a family will have (or develop) the trait given that status of other family members. Similar to the univariate logit and probit models, the inferences are similar in the multivariate case.

KEY WORDS: multivariate probit and logit models, generalized estimating equations, familial aggregation.

1. INTRODUCTION

In quantitative genetics and epidemiology, researchers are often interested in identifying important variables or traits related to a genetic disease and also in familial aggregation. The response variables measured for the disease can be discrete or continuous. In this paper, we focus on binary traits, such as presence/absence of a disease, and presence/absence of a symptom/feature of a genetic disease. For familial aggregation, for genetic hypotheses one would like to know the strength of dependence for different relationships in a family such as parent-offspring, sib-sib, degree 2 relationship (Falconer 1989).

One method for multivariate binary data is the GEE2 odds ratio regression or multivariate logit model. Liang, Zeger and Qaqish (1992), Molenberghs and Lesaffre (1994), Glonek and McCullagh (1995) and Joe (1997) considered this method/model from different points of view. GEE2 odds ratio regression corresponds to the multivariate logit model with a multivariate Plackett distribution (Molenberghs and Lesaffre 1994; Joe 1997). The formulation of Liang and Beaty (1991), and Liang, Zeger and Qaqish (1992) estimates regression parameters and log odds dependence parameters based on estimating equations without considering whether there is a probability model behind their assumptions.

The multivariate probit model (Ashford and Sowden 1970, Mendell and Elston 1974) is motivated for a binary trait based on a (latent) polygenic effect, so for familial analysis of a binary trait, it is more interpretable than a multivariate logit model. Although odds ratios have a convenient interpretation, there is no physical or stochastic model that leads to the odds ratio as a natural dependence parameter. Joe (1997) considers the multivariate probit and multivariate logit models as multivariate analogues of the univariate probit and logit latent variable family of models.

In this paper, we compare maximum likelihood and GEE2 estimates assuming a multivariate logit model and compare conditional probability inferences with the multivariate probit and logit models. The computation of GEE2 estimates includes the novel use of automatic differentiation software to solve the set of estimating equations.

The organization of the remainder of the paper is as follows. The models are specified as latent vector models in Section 2. Computational details are given in Section 3 and comparisons are made in Section 4. Section 5 concludes with a discussion.

2. MODELS AND METHODS FOR A BINARY TRAIT

For a binary response variable Y, with covariate vector $\mathbf{x}$, common statistical methods are logistic and probit regression. Both of these methods are latent variable methods with the probabilistic representation is:

$$Y = I(Z \leq \alpha + \boldsymbol{\beta}'\mathbf{x}), \tag{2.1}$$

where Z is standard normal (logistic) for the probit (logit) regression model. The model (2.1) can be written as

$$Y = I(X \geq \tau), \quad X = -Z + \alpha + \boldsymbol{\beta}'\mathbf{x} + \tau \sim N(\alpha + \boldsymbol{\beta}'\mathbf{x} + \tau, 1),$$

where X is the liability and τ is the threshold. For a binary trait, one can apply the Central Limit Theorem to arrive at the probit model when the liability is influenced by the additive effects of many genes. The logistic density is also bell-shaped, so probabilistic properties of logistic and probit regression are similar.

For familial data, the binary response vector is $(Y_1, \ldots, Y_d)$, where d is the family size. For a model for familial aggregation of a binary trait, one needs to define a joint probability distribution for $(Y_1, \ldots, Y_d)$ for $d \geq 2$, where the dependence parameter of responses (Y_i, Y_j) for two different members of a family, depends on the relation type.

Probit regression model easily extends to the multivariate probit model, with a latent multivariate normal random vector. The extension of logistic regression to its multivariate counterpart requires a way to define a multivariate logistic distribution with suitable dependence parameters. The extensions are explained below.

The multivariate probit model has been known for a long time (e.g., Ashford and Sowden 1970, Mendell and Elston 1974). The stochastic representation of the model, with common regression parameters for each margin, is

$$Y_j = I(Z_j \leq \alpha + \boldsymbol{\beta}'\mathbf{x}_j),\ j = 1, \ldots, d, \quad (Z_1, \ldots, Z_d) \sim N(\mathbf{0}, R_d) \tag{2.2}$$

where R_d is a correlation matrix. For models for familial aggregation, R_d can have one or more correlation parameters. For example, for the exchangeable model there is a single correlation parameter ρ; for a model for nuclear families, one has 3 parameters: correlations ρ_{PP}, ρ_{SS}, and ρ_{PO} for parent-parent, sib-sib,

and parent-offspring respectively; for a model with members in several generations, one may further have correlation parameters for second and higher degree relatives.

To obtain maximum likelihood estmates of the parameters of the model (2.2), multivariate normal rectangle probabilities are required, since

$$\Pr(Y_1 = y_1, \ldots, Y_d = y_d) = \Pr(Z_1 \prec\succ_1 \alpha + \boldsymbol{\beta}'\mathbf{x}_1, \ldots, Z_d \prec\succ_d \alpha + \boldsymbol{\beta}'\mathbf{x}_d)$$

where $\prec\succ_j$ is $\leq$ if $y_j = 1$ and is $>$ if $y_j = 0$.

For logistic regression, the regression parameter β for a binary covariate x has an odds ratio interpretation. For a pair (Y_1, Y_2), one can use the bivariate Plackett distribution which has an odds ratio as a dependence parameter:

$$\gamma = \gamma_{12} = \frac{\Pr(Y_1 = 1, Y_2 = 1; \mathbf{x}_1, \mathbf{x}_2)\ \Pr(Y_1 = 0, Y_2 = 0; \mathbf{x}_1, \mathbf{x}_2)}{\Pr(Y_1 = 1, Y_2 = 0; \mathbf{x}_1, \mathbf{x}_2)\ \Pr(Y_1 = 0, Y_2 = 1; \mathbf{x}_1, \mathbf{x}_2)} \tag{2.3}$$

for all $\mathbf{x}_1, \mathbf{x}_2$. Let $F_{12}(z_1, z_2)$ be the joint distribution of the latent pair (Z_1, Z_2) and $F(z) = (1 + e^{-z})^{-1}$ be the logistic cumulative distribution function. Then (2.3) is the same as

$$\gamma = \frac{F_{12}(\alpha + \boldsymbol{\beta}'\mathbf{x}_1, \alpha + \boldsymbol{\beta}'\mathbf{x}_2)\,[1 - F(\alpha + \boldsymbol{\beta}'\mathbf{x}_1) - F(\alpha + \boldsymbol{\beta}'\mathbf{x}_2) + F_{12}(\alpha + \boldsymbol{\beta}'\mathbf{x}_1, \alpha + \boldsymbol{\beta}'\mathbf{x}_2)}{[F(\alpha + \boldsymbol{\beta}'\mathbf{x}_1) - F_{12}(\alpha + \boldsymbol{\beta}'\mathbf{x}_1, \alpha + \boldsymbol{\beta}'\mathbf{x}_2)]\,[F(\alpha + \boldsymbol{\beta}'\mathbf{x}_2) - F_{12}(\alpha + \boldsymbol{\beta}'\mathbf{x}_1, \alpha + \boldsymbol{\beta}'\mathbf{x}_2)]} \tag{2.4}$$

and this equation can be solved for $F_{12}(\alpha + \boldsymbol{\beta}'\mathbf{x}_1, \alpha + \boldsymbol{\beta}'\mathbf{x}_2)$.

The multivariate Plackett extension is given in Molenberghs and Lessafre (1994), where (2.3) and (2.4) are extended to higher orders; for example, for $d = 3$ dimensions, with $\pi(y_1, y_2, y_3) = \Pr(Y_1 = y_1, Y_2 = y_2, Y_3 = y_3; \mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$,

$$\gamma_{123} = \frac{\pi(1,1,1)\,\pi(1,0,0)\,\pi(0,1,0)\,\pi(0,0,1)}{\pi(1,1,0)\,\pi(1,0,1)\,\pi(0,1,1)\,\pi(0,0,0)}. \tag{2.5}$$

For the d-dimensional product ratio, there are 2^{d-1} probabilities each in the numerator and denominator. Joe (1997) shows that these ratios do not lead to a proper multivariate logistic distribution if γ_{123} and higher order γ's are close to 0 or large. To have the same number of dependence parameters as the multivariate probit model, the third and higher order γ parameters are taken to be 1 (see Joe 1997 for a maximum entropy interpretation in this case). (2.5) and its higher-order equivalents lead to roots of polynomials that must be computed to obtain $F_{1\ldots d}(\alpha + \boldsymbol{\beta}'\mathbf{x}_1, \ldots, \alpha + \boldsymbol{\beta}'\mathbf{x}_d)$, the joint distribution of the latent multivariate logistic random vector.

Liang et al. (1992) and Liang and Beaty (1991) develop a method called odds ratio regression or GEE2 and apply it for familial aggregation of a binary trait. They do not assume any joint distribution for $(Y_1, \ldots, Y_d)$ but estimate interclass and intraclass odds ratios using estimating equations that generalize method of moments equations. These estimating equations use multivariate Plackett probabilities for dimension 2, 3, and 4. Although Liang and Beaty (1991) didn't mention any underlying model for their method,

their method corresponds to an estimation method for the multivariate logit model that is different from maximum likelihood. We will express their estimating equations in the more general context of familial data.

Let $\theta = (\beta, \psi)$, where ψ is a vector of log odds ratios, with different odds ratio parameters for different relation types (similar to the correlations for the probit model). Let $\mathbf{y}_i' = (y_{i1}, \ldots, y_{id_i})$ and let $\mathbf{w}_i' = (y_{i1}y_{i2}, \ldots, y_{i,d_i-1}y_{id_i})$, $i = 1, \ldots, n$, with n families and d_i members in the ith family. Let $\boldsymbol{\mu}_i' = \boldsymbol{\mu}_i'(\beta) = (\mu_{i1}, \ldots, \mu_{id_i})$, where $\mu_{ij} = \mathrm{E}(y_{ij})$, and $\boldsymbol{\eta}_i' = \boldsymbol{\eta}_i'(\theta) = (\mathrm{E}[y_{i1}y_{i2}], \ldots, \mathrm{E}[y_{i,d_i-1}y_{id_i}])$. The estimating equations have the form:

$$U(\theta) = \sum_{i=1}^{n} \frac{\partial(\boldsymbol{\mu}_i', \boldsymbol{\eta}_i')}{\partial\theta} \Sigma_i^{-1}(\theta) \begin{pmatrix} \mathbf{y}_i - \boldsymbol{\mu}_i \\ \mathbf{w}_i - \boldsymbol{\eta}_i \end{pmatrix} = \mathbf{0},$$

where $\Sigma_i(\theta)$ is the covariance matrix of $(\mathbf{y}_i, \mathbf{w}_i)$ based on multivariate Plackett probabilities up to dimension 4.

3. COMPUTATIONAL IMPLEMENTATION

Conceptually, the models in the previous section are straightforward, but computations for the maximum likelihood and GEE2 estimation methods are not straightforward.

For the multivariate probit model, we compute the multivariate normal rectangle probabilities using the fast approximation methods given in Joe (1995); the first order approximation requires only bivariate normal rectangle probabilities and the second order approximation requires multivariate probabilities up to the fourth dimension. The log-likelihood can then be coded and the maximum likelihood estimates (MLE) of β and the ρ parameters can be obtained using an iterative quasi-Newton method; for example, the method in Nash (1990) is convenient as it also computes the inverse Hessian (asymptotic covariance matrix) at the MLE.

For the multivariate logit model, we have coded the computation of multivariate Plackett probabilities by recursively finding the roots of many polynomial equations. Then maximum likelihood estimation and quasi-Newton iterations proceed in a similar way to the multivariate probit model. Because of the recursions, the computational effort for maximum likelihood estimation is exponentially increasing in the dimension or family size d.

The computer program of Liang and Beaty (1991) and Qaqish et al. (1992) cannot handle familial data in general pedigree form; it can only handle familial data in which each pair is either an interclass or intraclass pair. This code was written in Pascal and not easy to modify even after conversion to C with *p2c*. Therefore, we wrote a new implementation of GEE2, in which the equations were coded in C++, and the solutions of β and ψ parameters were obtained using the Newton-Raphson method with automatic differentiation

(Bendtsen and Stauning 1996) for the derivatives of the estimating equations with respect to the parameters. Because this requires multivariate Plackett probabilities in dimensions 4 and less, GEE2 with automatic differentiation is faster than maximum likelihood for family sizes of 5 or more, even with the C++ overhead in automatic differentiation.

All of our programs are written in C/C++, and are part of a developing software package for analysis for familial response data. Information can be obtained from the first author's web page.

The programs are written in a form that allows the user to specify general relation classes (see examples in Section 2) and there is a dependence parameter (latent correlation for probit and odds ratio for logit) for each relation class. For multivariate probit, one can do a variance component decomposition based on the estimated latent correlations. For the simpler use of the programs, the dependence parameters are not functions of covariates. This is mainly due to mathematical or probabilistic consistency of the models; there is no known way of making the correlation or odds ratio dependence parameters be functions of a covariate $\mathbf{x}$ so that the resulting correlation matrix is positive definite for all $\mathbf{x}$ or the resulting set of odds ratios are compatible for all $\mathbf{x}$. For a categorical covariate x, one could split the data into groups for estimates of the parameters or form extra relation classes. For example, if the gender of the parent might be a factor, one could use father-offspring and mother-offspring relation classes in place of the parent-offspring relation class.

4. COMPARISONS OF THE MODELS/METHODS

Latif (2001) has a simulation study to compare maximum likelihood and GEE2 estimates of the multivariate logistic models. The binary familial data are simulated from multivariate probit model with an age covariate. [Note that simulation from the multivariate logit model is much more difficult, since one cannot simulate the latent logistic variables easily because of the implicit equations defining the multivariate distribution functions.]

The simulation results are similar in different cases, so we summarize just one of the simulations in Latif (2001), in which all families have the 3-generation 5-member pedigree. with a sib-sib pair, one parent, one uncle/aunt and one grandparent (see Table I). Other comparisons via simulations in Latif (2001) include cases of different pedigrees for different families.

The table shows that maximum absolute differences of the maximum likelihood and GEE2 parameter estimates as well as the average of each point estimate over the 500 simulations of 200 families; the dependence parameters are correlation for probit and log odds ratio for logit. The maximum likelihood and GEE2 estimates were often the same to 2 or 3 significant digits. The standard deviation of the parameter estimates and the average standard errors in each line are roughly the same. See below for the relation of the probit

and logit regression coefficients, and relation of a latent correlation and odds ratio dependence parameter.

For illustration with some real data, we use a data set of familial binary response data for patients with neurofibromatosis 1 (NF1) which is an autosomal dominant genetic disease (Friedman et al. 1999); the binary variables are indicators of the presence of features, such as café-au-lait spots, plexiform neurofibromas, intertriginous freckling, Lisch nodules, etc. A detailed study of familial aggregration of features is given in Szudek et al. (2002). We just show some results for the presence/absence of peripheral neurofibromas for a subset of the NF1 database. The analysis must be adjusted for age because of a tendency of increased incidence of the feature with age. We use three relation classes: sib-sib, parent-offspring and degree 2 relation (there are no parent-parent pairs as within the families only one parent has Nf1). For two family members with relation of degree 3 or higher, the pairwise dependence parameter is taken to be the value corresponding to pairwise independence. Because our data set has fewer pairs for degree 2 than sib-sib or parent-offspring, the dependence parameter for degree 2 relation can not be estimated accurately. The family sizes range from 1 to 7; missing values for the response were assumed to be missing at random.

Table II has a summary table for the multivariate probit and logit models, with MLEs for probit and GEE2 estimates for logit. The dependence parameters are latent correlations and log odds ratios respectively.

The standard deviation of the standard logistic distribution is $\pi/\sqrt{3} = 1.81$, so for the same data, the regression coefficients of logistic regression are usually roughly 1.8 times the corresponding regression coefficients of probit regression. For two binary variables based on a latent bivariate standard normal distribution with correlation ρ, the odds ratio depends on the cut-off points, but is bounded by

$$B(\rho) = \left\{ \frac{1 + (2/\pi) \arcsin \rho}{1 - (2/\pi) \arcsin \rho} \right\}^2 .$$

For the NF1 example, the estimated latent correlations from multivariate probit are 0.587 and 0.158 for sib-sib and parent-offspring, and the estimated odds ratio from multivariate logit are 6.35 and 1.76 respectively. For comparison, $B(0.587) = 5.43$ and $B(0.158) = 1.50$ so that the dependence estimates of the two models are comparable given the standard errors in the estimates.

Table III shows conditional probabilities $\Pr(Y_d = y_d \mid y_1, \ldots, y_{d-1}, \mathbf{x}_1, \ldots, \mathbf{x}_d)$ for three cases (values taken from our data set) to compare the conditional probabilities for the fitted parameters from the multivariate probit and logit models. The indexing within families is from the oldest to the youngest. In the fourth column, the notation for relation classes is 0 for sib-sib, 1 for parent-offspring, 2 for degree two; relations are given in order for pairs (1,2), (1,3), $\ldots$, $(d-1, d)$. In the third case, the first child is a half-sib of the other two (degree 2 relation).

The similarity of results for the multivariate probit and logit models here is not surprising. Joe (1997, Chapter 11) has examples for multivariate binary and ordinal data which show that inferences for multivariate probit and logit models are very similar.

5. DISCUSSION

Liang and Beaty (1991) mention that their odds ratio regression model is a more general model and avoids the unobservable continuous trait of the multivariate probit model. However we have shown that the probabilistic assumptions behind their method are conceptually very close to that of the multivariate probit model with latent logistic random variables in place of latent normal random variables.

The inferences from the two models are very similar, and we prefer the multivariate probit model because of the physical derivation for a polygenic effect. However, it is still useful to apply the multivariate logit model for a sensitivity analysis. Note that the GEE2 type of estimation equations for the multivariate probit model for familial data could be implemented like in Reboussin and Liang (1998).

The approximations of Joe (1995) have made it easier to do computations for the multivariate probit model. Lesaffre and Molenberghs (1991) mention that lack of software for the multivariate probit model and provide software for bivariate probit. Molenberghs and Lesaffre (1994) mention availability of software for the multivariate logit model. Their software was written in GAUSS. Our software, which is written in C/C++, will be much faster. This is crucial for familial data with large family sizes, as the computational effort increases rapidly with the family size. Even with compiled programs in C/C++, the computational time will be of the order of minutes on fast Pentium computers if there are many families of size 6 or more.

Acknowledgements

This research is supported from an NSERC Canada grant and US Army Medical Research grant NF990038.

REFERENCES

[1] Ashford JR, Sowden RR. Multivariate probit analysis. *Biometrics* 1970; **26**:535–546.

[2] Bendtsen C, Stauning O. FADBAD, a flexible C++ package for automatic differentiation. Technical report 1996-17. http://www.imm.dtu.dk/fadbad.html

[3] Falconer DS. *Introduction to Quantitative Genetics*, 3rd edition. Longman Group, HK, 1989.

[4] Friedman JM, Gutmann DH, MacCollin M, Riccardi VM. *Neurofibromatosis: Phenotype, Natural History, and Pathogenesis, 3rd ed.* Johns Hopkins University Press, 1999.

[5] Joe H. Approximations to multivariate normal rectangle probabilities based on conditional expectations. *Journal of the American Statistical Association* 1995; **90**:957–964.

[6] Joe H. *Multivariate Models and Dependence Concepts*. Chapman & Hall, London, 1997.

[7] Latif AHMM. A comparison of methods for multivariate familial binary response. M.Sc. Thesis 2001, Department of Statistics, University of British Columbia.

[8] Lesaffre E, Molenberghs G. Multivariate probit analysis: a neglected procedure in medical statistics. *Statistics in Medicine* 1991; **10**:1391–1403.

[9] Liang KY, Beaty TH. Measuring familial aggregation by using odds-ratio regression models. *Genetic Epidemiology* 1991; **8**: 361–370.

[10] Liang KY, Zeger SL, Qaqish B. Multivariate regression analyses for categorical data. *Journal of the Royal Statistical Society, Series B* 1992; **54**:3–24.

[11] Glonek GFV, McCullagh P. Multivariate logistic models. *Journal of the Royal Statistical Society, Series B* 1995; **57**:533–546.

[12] Mendell NR, Elston RC. Multifactorial qualitative traits: genetic analysis and prediction of recurrence risks", *Biometrics* 1974; **30**:41–57.

[13] Molenberghs G, Lesaffre E. Marginal modelling of correlated ordinal data using a multivariate Plackett distribution. *Journal of the American Statistical Association* 1994; **89**:633–644.

[14] Nash JC. *Compact Numerical Methods for Computers: Linear Algebra and Function Minimisation*, second edition. Hilger, New York, 1990.

[15] Reboussin BA, Liang KY. An estimating equations approach for the LISCOMP model. *Psychometrika* 1998; **63**:165–182.

[16] Szudek J, Joe H, Friedman JM. Analysis of intra-familial phenotypic variation in neurofibromatosis 1 (Nf1) *Genetic Epidemioly* 2002; **23**:150–164.

Table I: Simulation data from multivariate probit: average of estimates, average absolute differences (ML and GEE2) and average SEs assuming a multivariate logit model.

	True	Parameter Estimates			Standard Errors		
	values	ML	GEE2	Diff.	ML	GEE2	Diff.
Const.	0.5	0.787	0.787	0.003	0.165	0.160	0.007
Age	1.0	1.796	1.798	0.011	0.406	0.394	0.021
SS	0.8	2.885	2.888	0.011	0.294	0.294	0.014
PO	0.6	1.944	1.947	0.021	0.273	0.271	0.018
D2	0.4	1.223	1.212	0.019	0.281	0.279	0.022
Const.	0.8	1.310	1.310	0.004	0.181	0.178	0.008
Age	0.2	0.374	0.375	0.011	0.399	0.393	0.021
SS	0.9	3.815	3.821	0.015	0.343	0.342	0.020
PO	0.5	1.527	1.534	0.018	0.267	0.263	0.020
D2	0.3	0.914	0.909	0.020	0.287	0.279	0.025

Table II: NF1 data. Parameter estimates for multivariate probit and logit models.

parameter	mprobit	SE	mlogit	SE
intercept	-1.236	0.097	-2.248	0.209
age/100	6.436	0.423	12.319	1.353
sib-sib	0.587	0.115	1.849	0.354
parent-child	0.158	0.111	0.564	0.924
degree2	0.031	0.259	0.085	0.763

Table III: NF1 data. Comparisons of some conditional probabilities for multivariate probit and logit fits.

d	y's	ages/100	relations	mprobit	mlogit
3	1,1,0	0.528, 0.230, 0.191	1,1,0	0.340 (0.039)	0.320 (0.028)
3	1,0,0	0.391, 0.096, 0.064	1,1,0	0.881 (0.026)	0.892 (0.019)
4	0,1,0,0	0.379, 0.155, 0.051, 0.038	1,1,1,2,2,0	0.934 (0.047)	0.933 (0.041)

ANALYSIS OF NEUROFIBROMATOSIS 1 (NF1) LESIONS BY BODY SEGMENT

Vairochana Palmer[1,2], Jacek Szudek[1,3], Harry Joe[4], Vincent M. Riccardi[5], and J.M. Friedman[1,*]

[1] *Department of Medical Genetics, The University of British Columbia, Vancouver, CANADA*

[2] *Department of Genetics, Stanford University, Stanford, California, USA*

[3] *Dalhousie Medical School, Halifax, Nova Scotia, CANADA*

[4] *Department of Statistics, The University of British Columbia, Vancouver, CANADA*

[5] *American Medical Consumers, La Crescenta, California, USA*

Running Title: Analysis of NF1 lesions by body segment

Grant sponsor: This work was supported in part by United States Army Medical Research and Materiel Command grant number NF960003.

* Correspondence to: J. M. Friedman, MD, PhD; Department of Medical Genetics; The University of British Columbia; 6174 University Boulevard; Vancouver, BC V6T 1Z3; CANADA

Telephone: (604) 822-2749

Fax: (604) 822-5348

E-mail: frid@interchange.ubc.ca

ABSTRACT

Café-au-lait spots and neurofibromas are defining features of neurofibromatosis 1 (NF1), but they vary greatly in number, size, and clinical importance from patient to patient. The cause of this variability is unknown. We tested the hypotheses that development of these lesions is influenced by local or familial factors.

The presence or absence of café-au-lait spots, cutaneous neurofibromas, and diffuse plexiform neurofibromas was recorded for each of ten divisions of the body surface in 547 NF1 patients, including 117 affected individuals in 52 families. We used stratified Mantel-Haenszel tests to look for local associations between the presence of diffuse plexiform neurofibromas, cutaneous neurofibromas, and café-au-lait spots in individual body segments of NF1 patients. We used a random effects model to obtain intrafamilial correlation coefficients for the age-adjusted number of body divisions affected with each of the three lesions.

No significant association was observed between the occurrence of cutaneous and diffuse plexiform neurofibromas, between café-au-lait spots and cutaneous neurofibromas, or between café-au-lait spots and plexiform neurofibromas in the same body segment. The correlation among relatives in the number of body segments affected with café-au-lait spots was 0.45 (95% confidence interval [CI] = 0.18, 0.71), with cutaneous neurofibromas, 0.37 (95% CI = 0.15, 0.55), and with plexiform neurofibromas, 0.35 (95% CI = 0.15, 0.57). We conclude that the development of café-au-lait spots, cutaneous neurofibromas, and plexiform neurofibromas are spatially independent in NF1 patients but that the development of all three lesions is influenced by familial factors.

Keywords: Neurofibromatosis 1, familial correlation, café-au-lait spots, neurofibromas

INTRODUCTION

Neurofibromatosis 1 (NF1) is an autosomal dominant condition characterized by extremely variable expressivity. Café-au-lait spots and neurofibromas are the defining features. Neurofibromas are complex benign tumors arising in the fascicles of peripheral nerves (Korf, 1999). Histologically, a local increase in endoneurial matrix of the fascicle is accompanied by a thickened perineurium, increased size and number of Schwann cells (Harkin and Reed, 1969; Woodruff, 1999), and increased numbers of mast cells and fibroblasts (Giorno et al., 1989). Cutaneous neurofibromas are confined to a single fascicle within a nerve, while diffuse plexiform neurofibromas involve multiple fascicles (Burger and Scheithauer, 1994).

Cutaneous neurofibromas begin to appear in mid-childhood and eventually develop in almost all NF1 patients (Friedman and Riccardi, 1999; DeBella et al., 2000). Cutaneous neurofibromas tend to increase in number and size with age. Some adults with NF1 have hundreds or thousands of these lesions; other NF1 patients develop only a few cutaneous neurofibromas throughout life.

Diffuse plexiform neurofibromas are almost always, if not always, congenital (Friedman and Riccardi, 1999). Many are apparent on surface examination, although they often extend into deeper tissues. Some diffuse plexiform neurofibromas involve only deeper tissues and are not apparent on physical examination. Plexiform neurofibromas tend to be larger than cutaneous neurofibromas, sometimes involving an entire limb or other part of the body. Plexiform neurofibromas may give rise to malignant peripheral nerve sheath tumours, but discrete cutaneous neurofibromas rarely, if ever, do.

Café-au-lait spots are pigmented macules. Histologically, they contain melanocytes with abnormally large pigment particles (Fitzpatrick, 1981). Café-au-spots may be present at birth,

and by one year of age almost all children with NF1 have 6 or more of these lesions (Friedman and Riccardi, 1999; DeBella et al., 2000).

The number and location of café-au-lait spots and neurofibromas are highly variable, even among NF1 patients of similar age. The cause of this variability is unknown. Here we test the hypotheses that the development of these lesions is influenced by local or familial factors.

SUBJECTS AND METHODS

Subjects. 547 NF1 patients, including 117 affected individuals in 52 families, who had information recorded on spatial distribution of skin lesions were available in the NF Institute Database (Riccardi 1992). All of these patients were evaluated by Dr. Vincent Riccardi, and all meet the NIH diagnostic criteria for NF1 (Gutmann et al. 1997; National Institutes of Health Consensus Development Conference 1988). For each patient, the presence of one or more café-au-lait spots, one or more cutaneous neurofibromas, and one or more diffuse plexiform neurofibromas was recorded for each of the ten divisions of the body surface shown in Figure 1.

Analysis of local effect. We used two-layered Mantel-Haenszel tests (SPSS 1998) to look for local associations between the presence of diffuse plexiform neurofibromas and cutaneous neurofibromas in individual body segments of each NF1 patient. We stratified simultaneously by the body segment being considered and by the number of other body segments with one or more cutaneous neurofibromas (a categorical variable with range 0 to 9). This stratification was used to adjust for the fact that an NF1 patient who has a larger total number of body segments

with one or more neurofibromas is more likely to have at least one neurofibroma in any particular segment than an NF1 patient who has fewer total body segments affected. Confidence intervals for the summary odds ratio were obtained using a jackknife based on 20 different subgroups – a number that is sufficiently large to produce a stable estimate (Miller 1974). Homogeneity was assessed using the Breslow-Day test (SPSS 1998). Local associations between café-au-lait spots and cutaneous neurofibromas and between café-au-lait spots and plexiform neurofibromas were analyzed in the same manner.

Skin surface area. The body divisions used in this study cover varying amounts of skin surface area, so we checked for an association between the surface area and the presence of one or more cutaneous neurofibromas in a segment. Using logistic regression, we set the segment area as the independent variable and the presence or absence of cutaneous neurofibromas as the dependent variable. We tested in a similar manner for associations between surface area and the presence of diffuse plexiform neurofibromas and café-au-lait spots in a segment. Since the median age of our patients was 13 years, we approximated the surface area of the body segments by using standard percentages for 10-14 year-old individuals (McManus and Pruitt 1996). The proportions of total surface area assigned to each body segment were: head = 11%, neck = 2%, right upper torso = 12%, left upper torso = 12%, right lower torso = 4%, left lower torso = 4%, right arm = 9.5%, left arm = 9.5%, right leg = 18%, and left leg = 18%.

Total number of neurofibromas. In addition to data on whether each body segment was affected by one or more cutaneous neurofibromas, complete counts of cutaneous neurofibromas were available for 44 of the patients. The total number of neurofibromas in these patients ranged

from none to several hundred and appeared to increase logarithmically with the number of affected segments. We used linear regression (SPSS 1998) to test the relationship between log-transformed counts of the total number of cutaneous neurofibromas in an individual and the number of body segments that included one or more cutaneous neurofibromas. Counts of the total number of café-au-lait spots were not made, and few subjects had more than one plexiform neurofibroma, so these variables were not analyzed in this manner.

Familial analysis. For the familial analysis, we stratified subjects into 5-year age intervals, calculated the deciles for the total number of segments affected with cutaneous neurofibromas in each stratum, and ranked each subject by decile for the stratum in which he or she lay. We then used random effects models to obtain maximum likelihood estimates and confidence intervals for intrafamilial correlation coefficients for rank (Donner et al. 1989; Spjotvoll 1967). Café-au-lait spots and plexiform neurofibromas were analysed in the same manner.

RESULTS

We studied the distribution of café-au-lait spots, cutaneous neurofibromas, and diffuse plexiform neurofibromas in 10 segments of the body surface (Figure 1) in each of 547 patients with NF1. Two hundred eighty-one (51.4%) of the subjects were female, and 266 (48.6%) were male. Four hundred twenty-six (77.9%) were white, 67 (12.2%) were Hispanic, 44 (8.0%) were black and 10 (1.8%) were of other or mixed origin. Mean age was 17.5 years, and median age was 13 years.

Lesion frequency by body segment. Table 1 shows the frequency of these lesions in each of the 10 body segments. Two hundred ten patients had no cutaneous neurofibromas in any segment, and 337 patients had one or more cutaneous neurofibromas. Plexiform neurofibromas were noted in 216 patients. Cutaneous and plexiform neurofibromas occurred with similar frequencies in all ten body segments. Café-au-lait spots were observed in almost all patients and had similar frequencies in all segments except the head, where these lesions were less frequent.

No associations between lesion types within individual body segments. Table 2 shows the ten body segments examined and the odds ratios for associations of each pair of lesions for each segment. No association was observed between the occurrence of cutaneous and diffuse plexiform neurofibromas in the same body segment. The summary odds ratio was 1.20 (95% confidence interval [CI] = 0.81, 1.79). There was no evidence for heterogeneity across body segments (p=0.37).

Similarly, there was no association between the presence of café-au-lait spots and either cutaneous or diffuse plexiform neurofibromas within a single body segment. The summary odds ratios were 1.26 (95% CI = 0.82, 1.93) for café-au-lait spots and cutaneous neurofibromas and 1.25 (95% CI = 0.74, 2.12) for café-au-lait spots and plexiform neurofibromas. There was significant (p=0.03) heterogeneity in the occurrence of cutaneous neurofibromas and café-au-lait spots, with a positive association seen in the neck (odds ratio=2.94; 95% CI = 1.20, 7.20). No evidence of heterogeneity across body segments was found for the occurrence of plexiform neurofibromas and café-au-lait spots (p=0.52).

Log-linear relationship between segment size and number of cutaneous neurofibromas. The number of body segments affected with one or more cutaneous neurofibromas was strongly correlated with the total number of cutaneous neurofibromas in 44 NF1 patients in whom both total counts and data on the number of affected body segments were available ($r=0.95$, $p<0.001$). The relationship is log linear; the regression equation is

$$Log(\text{total number of neurofibromas} + 1) = 0.23 * (\text{number of segments affected}) + 0.014.$$

We observed no significant association between the relative size of the body surface area in a segment and the presence of one or more cutaneous neurofibromas ($p=0.18$) or of a diffuse plexiform neurofibroma ($p=0.23$). In contrast, an association was found between the presence of one or more café-au-lait spot in a body segment and its surface area expressed as a percentage of the body's total ($p<0.001$, odds ratio = 1.030, 95% CI = 1.015, 1.046).

All three lesions are correlated among relatives with NF1. We estimated intrafamilial correlations in the age-adjusted number of body segments that included one or more café-au-lait spots, one or more cutaneous neurofibromas, or one or more plexiform neurofibromas in 117 affected members of 52 families. We found significant intrafamilial correlations for the number of body segments affected by each of these clinical features. The intrafamilial correlation coefficient for the number of body segments affected with café-au-lait spots was 0.45 (95% CI = 0.18, 0.71). The correlation among relatives with NF1 for the number of body segments affected with cutaneous neurofibromas was 0.37 (95% CI = 0.15, 0.55). The correlation coefficient

among relatives for the number of body segments affected with plexiform neurofibromas was 0.35 (95% CI = 0.15, 0.57).

DISCUSSION

Lesions in body segments of individual patients. The number of body segments affected by one or more cutaneous neurofibromas appears to provide a good measure of how severely each of these NF1 patients is affected by this disease feature. We found a very high correlation (r = 0.95) between the number of body segments in which one or more cutaneous neurofibromas was present and the total number of cutaneous neurofibromas in 44 patients in whom counts were available. It seems likely that a similar relationship exists between the number of body segments affected with café-au-lait spots or plexiform neurofibromas and the severity of each of these disease features, but we did not have information on total counts of these lesions available to demonstrate this.

We have shown previously that individuals with diffuse plexiform neurofibromas are more likely also to have dermal neurofibromas (Szudek et al. 2000a; Szudek et al. Submitted for publication-a), but this association did not take into account the location or number of these lesions. The current study is the first to examine this association within body divisions. Since almost all, if not all, diffuse plexiform neurofibromas are of congenital origin (Friedman and Riccardi 1999), we wanted to find out if they influence the subsequent development of cutaneous neurofibromas. Our findings indicate that the occurrence of cutaneous neurofibromas in NF1 patients is not strongly influenced by the local presence of a diffuse plexiform neurofibroma. In

fact, we found that all three of the lesions studied (café-au-lait spots, cutaneous neurofibromas, and plexiform neurofibromas) occurred independently of each another in almost all of the body segments analyzed (Table 2).

We found a significant association between café-au-lait spots and cutaneous neurofibromas only in the neck. One possible reason the neck might be affected by both lesions is recurrent minor trauma to the skin associated with flexion, extension, and rotation of the head (Riccardi 1990). Clearly, however, other factors are also involved in the pathogenesis of café-au-lait spots and neurofibromas, as indicated by the familial correlations we observed for the age-adjusted number of body segments affected by each of the three lesions studied.

Familial correlations. The intrafamilial correlations we observed for cutaneous neurofibromas and café-au-lait spots in NF1 patients are consistent with the findings of a previous study (Easton et al. 1993). The number of familial patients and the prevalences of all three lesions were similar in these two studies. Our study found a similar correlation for café-au-lait spots but higher correlation coefficients for cutaneous neurofibromas than Easton and his associates did. We also found a significant familial correlation for plexiform neurofibromas. Easton et al. only analyzed this feature as a discrete (present/absent) trait and found no familial association.

We have also studied the familiality of café-au-lait spots, cutaneous neurofibromas, and plexiform neurofibromas as discrete traits in an independent series of NF1 patients using multivariate probit regression analysis with adjustment for age and the presence of associated clinical features (Szudek et al. 2000b; Szudek et al. Submitted for publication-b). The results of that study are consistent with the current one and with the study of Easton and associates (1993) despite the differences in design and methodology: We again found strong intrafamilial

correlations for café-au-lait spots (r = 0.43, 95% CI 0.29-0.57) and cutaneous neurofibromas (r = 0.49, 95% CI 0.33-0.65). Like Easton et al., we did not find a correlation for the occurrence of plexiform neurofibromas considered as a discrete trait when all relatives were considered, but we did find a significant sib-sib correlation for the occurrence of this clinical feature (r = 0.18, 95% CI 0.04-0.32). These observations provide further evidence for the importance of familial factors in the development of café-au-lait spots and neurofibromas in people with NF1.

The genetic basis for these familial associations has not been determined, but contributing factors may include effects of the mutant *NF1* allele itself, effects of the normal *NF1* allele, or modifying effects of other loci. The moderate magnitudes of the intrafamilial correlation coefficients show that familial factors alone are insufficient to predict the degree to which a patient will be affected by these lesions.

Our results are consistent with the possibility that different pathogenic mechanisms are involved in development of the three lesions studied. Chimeric mice composed in part of $Nf1^{-/-}$ cells develop plexiform neurofibromas but not cutaneous neurofibromas (Cichowski et al. 1999; Vogel et al. 1999). On the other hand, insertion of *tax* into the germline of mice leads to the development of multiple cutaneous neurofibromas but not plexiform neurofibromas (Feigenbaum et al. 1996). It is, therefore, clear that these two types of neurofibromas can develop by independent pathways, at least in mice. Some families with *NF1* mutations develop café-au-lait spots but no tumours (Abeliovich et al. 1995), consistent with different pathogenic factors being involved in the development of café-au-lait spots and neurofibromas.

In summary, multiple factors appear to be involved in the pathogenesis of café-au-lait spots as well as of both plexiform and cutaneous neurofibromas in patients with NF1. Some of

these factors are familial, but others are not. Some pathogenic factors may be shared among these three lesions, but other pathogenic mechanisms appear to differ.

ACKNOWLEDGEMENTS

We thank Patricia Birch for her help with this study.

Table 1: Number and percentage of 547 NF1 patients who have one or more cutaneous neurofibromas, diffuse plexiform neurofibromas or café-au-lait spots in each of 10 body segments.

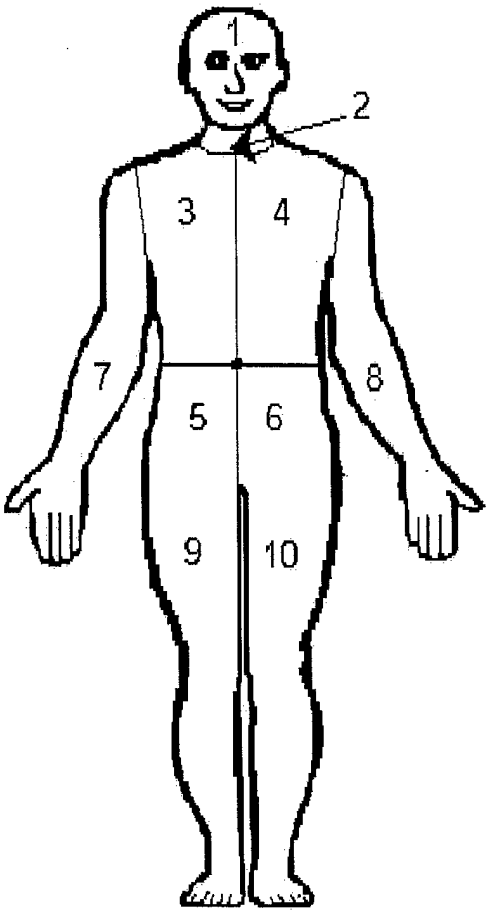
Segment	Cutaneous Neurofibromas		Plexiform Neurofibromas		Café-au-lait Spots	
	Total	(%)	Total	(%)	Total	(%)
1 Head	179	(33%)	47	(8%)	101	(18%)
2 Neck	168	(31%)	29	(5%)	397	(73%)
3 Right Upper Torso	259	(47%)	32	(6%)	532	(97%)
4 Left Upper Torso	258	(47%)	21	(4%)	531	(97%)
5 Right Lower Torso	285	(52%)	55	(10%)	537	(98%)
6 Left Lower Torso	287	(52%)	41	(7%)	533	(97%)
7 Right Arm	206	(38%)	21	(4%)	514	(94%)
8 Left Arm	208	(38%)	19	(3%)	511	(93%)
9 Right Leg	219	(40%)	54	(10%)	527	(96%)
10 Left Leg	220	(40%)	45	(8%)	525	(96%)
Total	***337***	***(62%)***	***216***	***(39%)***	***543***	***(99%)***

Table 2: Associations between cutaneous neurofibromas, diffuse plexiform neurofibromas and café-au-lait spots by body segment in 547 NF1 patients. Odds ratios could not be calculated for the association of café-au-lait spots and plexiform neurofibibromas in the right upper torso, right lower torso, or right arm because there were no patients who had plexiform neurofibromas but did not have café-au-lait spots in these segments.

	Cutaneous and Plexiform Neurofibromas		Cutaneous Neurofibromas and Café-au-lait Spots		Café-au-lait spots and Plexiform Neurofibromas	
Segment	Odds Ratio	(95% C.I.)	Odds Ratio	(95% C.I.)	Odds Ratio	(95% C.I.)
1 Head	0.95	(0.34-2.68)	1.34	(0.67-2.67)	1.26	(0.60-2.65)
2 Neck	2.39	(0.51-11.20)	2.59	(1.23-5.47)	2.42	(0.71-8.24)
3 Right Upper Torso	0.83	(0.23-3.02)	0.26	(0.01-10.34)	-	-
4 Left Upper Torso	0.39	(0.06-2.49)	0.12	(0.01-7.07)	1.29	(0.02-83.37)
5 Right Lower Torso	0.85	(0.32-2.24)	0.98	(0.01-84.41)	-	-
6 Left Lower Torso	0.91	(0.35-2.36)	1.13	(0.19-6.94)	0.06	(0.01-0.99)
7 Right Arm	1.17	(0.09-14.43)	1.91	(0.29-12.67)	-	-
8 Left Arm	1.01	(0.18-5.60)	0.91	(0.18-4.65)	0.22	(0.02-1.97)
9 Right Leg	3.91	(1.02-15.06)	0.2	(0.04-1.15)	0.35	(0.05-2.34)
10 Left Leg	3.6	(0.99-13.08)	1.1	(0.24-5.00)	2.7	(0.17-44.12)
Summary	***1.2***	***(0.81-1.79)***	***1.36***	***(0.91-2.03)***	***1.25***	***(0.74-2.12)***

FIGURE LEGEND

Figure 1: Body segment scheme used by Neurofibromatosis Institute Database.


1
2
3
4
7
8
5
6
9
10

REFERENCES

Abeliovich D, Gelman-Kohan Z, Silverstein S, Lerer I, Chemke J, Merin S, Zlotogora J. 1995. Familial cafe au lait spots: a variant of neurofibromatosis type 1. J Med Genet 32: 985-986.

Burger P, Scheithauer B. 1994. Tumors of the central nervous system. Tumors of the central nervous system. Armed Forces Institute of Pathology, Washington, DC.

Cichowski K, Shih TS, Schmitt E, Santiago S, Reilly K, McLaughlin ME, Bronson RT, Jacks T. 1999. Mouse models of tumor development in neurofibromatosis type 1. Science 286: 2172-2176.

DeBella K, Szudek J, Friedman JM. 2000. Use of the national institutes of health criteria for diagnosis of neurofibromatosis 1 in children. Pediatrics 105:608-614.

Donner A, Wells G, Eliasziw M. 1989. On two approximations to the F-distribution: application to testing for intraclass correlation in family studies. Canadian Journal of Statistics 17: 209-215.

Easton D, Ponder M, Huson S, Ponder B. 1993. An analysis of variation in expression of neurofibromatosis (NF) type I (NFI): Evidence for modifying genes. Am J Hum Genet 53: 305-313.

Feigenbaum L, Fujita K, Collins FS, Jay G. 1996. Repression of the NF1 gene by Tax may expain the development of neurofibromas in human T-lymphotropic virus type 1 transgenic mice. J Virol 70: 3280-3285.

Fitzpatrick TB. 1981. Melanin synthesis pathways in the pathogenesis of neurofibromatosis. Adv Neurol 29: 209-211.

Friedman JM, Riccardi VM. 1999. Clinical and epidemiological features. *In* J. M. Friedman, D. H. Gutmann, M. MacCollin and V. M. Riccardi (eds), Clinical and epidemiological features. Johns Hopkins University Press, Baltimore, pp. 29-86.

Giorno R, Lieber J, Claman HN. 1989. Ultrastructural evidence for mast cell activation in a case of neurofibromatosis. Neurofibromatosis 2: 35-41.

Gutmann D, Aylsworth A, Carey J, Korf B, Marks J, Pyeritz R, Rubenstein A, Viskochil D. 1997. The diagnostic evaluation and multidisciplinary management of neurofibromatosis 1 and neurofibromatosis 2. JAMA 278: 51-57.

Harkin J, Reed R. 1969. Tumors of the peripheral nervous system. Tumors of the peripheral nervous system. Armed Forces Institute of Pathology, Washington, DC, pp. 67-100.

Korf BR. 1999. Plexiform neurofibromas. Am J Med Genet 89:31-37.

Miller R. 1974. The jackknife - a review. Biometrika 61: 1-15.

National Institutes of Health Consensus Development Conference. 1988. Neurofibromatosis: Conference statement. Arch Neurol 45: 575-578.

McManus WF, Pruitt BA Jr. 1996. Thermal injuries. *In* D.V. Feliciano, E.E. Moore and K.L. Mattox (eds), Trauma, 3rd ed. Appleton & Lange, Stamford, Connecticut, pp. 937-949.

Riccardi V. 1990. The potential role of trauma and mast cells in the pathogenesis of neurofibromas. *In* Y. Ishibashi and Y. Hori (eds), The potential role of trauma and mast cells in the pathogenesis of neurofibromas. Elsevier, Amsterdam, pp. 167-190.

Riccardi VM. 1992. Neurofibromatosis: Phenotype, natural history, and pathogenesis. The Johns Hopkins University Press, Baltimore.

Spjotvoll E. 1967. Optimum invariant tests in unbalanced variance component models. Ann Math Statist 38: 422-428.

SPSS. 1998. SPSS for Windows. Ver. 9.0.0.

Szudek J, Birch P, Riccardi VM, Evans DG, Friedman JM. 2000a. Associations of clinical features in neurofibromatosis 1 (NF1). Genet Epidemiol 19: 429-439.

Szudek J, Evans D, Friedman JM. (Submitted for publication-a). Logistic regresssive models of neurofibromatosis 1 (NF1) clinical features.

Szudek J, Joe H, Friedman JM. 2000b. Familial aggregation of neurofibromatosis 1 (NF1) clinical features. Am J Hum Genet 67: 211.

Szudek J, Joe H, Friedman JM. (Submitted for publication-b). Analysis of intra-familial phenotypic correlation in neurofibromatosis 1 (NF1). Submitted for publication

Vogel KS, Klesse LJ, Velasco-Miguel S, Meyers K, Rushing EJ, Parada LF. 1999. Mouse tumor model for neurofibromatosis type 1. Science 286: 2176-2179.

Woodruff JM. 1999. Pathology of tumors of the peripheral nerve sheath in type 1 neurofibromatosis. Am J Med Genet 89:23-30.

Appendix 2

Outline for Yinshan Zhao's thesis (from Section 1.4 of thesis).

Chapter 2 consists of a review of variance component models for quantitative traits (continuous values). This chapter provides insight in how the underlying genetic mechanism determines the correlation structure of familial data, and it is fundamental to the development of the chapter that follows.

In Chapter 3, we discuss three different modelling approaches for familial data. We first present two existing approaches which have been applied to familial data or other multivariate data: the first approach, to construct dependence structure, is by introducing random effects and the second is by using the multivariate normal copula. In this thesis we provide a comprehensive summary of these models under the context of familial data analysis. It also serves as a background for Chapter 5, in which estimating procedures are addressed. We then propose a family of new models called conditional independence models. In this approach, models are constructed based on the assumption that the trait values of two non-sibling relatives are independent conditional on
their parents. This approach reduces the task of modelling a complex pedigree to modelling a family unit containing only the parents and their offsprings. After the general introduction of the models, we study some specific models for binary, count and survival responses.

In Chapter 4, we propose a method for a binary trait to estimate intraclass and interclass odds ratios using relative pairs, and derived the asymptotic variance of the estimate. Asymptotic efficiency is compared with the maximum likelihood estimate (MLE) based on a multivariate probit model.

In Chapter 5, we propose two likelihood-based estimating methods. The first approach is a two-stage method in which univariate marginal parameters and dependence parameters are estimated separately based on the likelihoods of the univariate marginal distributions and bivariate
marginal distributions. In the second approach, all the parameters are estimated simultaneously based on the likelihoods of the bivariate marginal distributions. Both methods yield asymptotically consistent parameter estimates. In Chapter 6, we investigate the performance of the two methods. The asymptotic efficiency of the estimates were compared with the MLEs when the latter can be obtained. Our results show that the two-stage method can be inefficient with the covariate coefficients when the correlation is strong.

In Chapter 7, the models and inferential approaches studied in the previous chapters are applied to datasets of patients with neurofibromatosis type 1 or type 2. In Chapter 8, the final chapter, we discuss some future research topics related to this thesis.

Appendix 3.

Excerpt from summary section on estimation methods from Zhao's thesis.

Estimating Procedures and Efficiency Comparison

The major difficulty in implementing models with multinormal random effects and multivariate normal (MVN) copula models is parameter estimation when the family size exceeds 4. The MLE is generally computationally difficult to obtain since it involves high dimensional integration. Therefore, developing estimating procedures that are less computationally demanding is important.

The models we mentioned share a common feature: the parameters which specify the models can be classified as univariate marginal parameters and dependence parameters, the former characterizes the univariate margins, such as the means and variances in the MVN model, while the latter, joint with the univariate marginal parameters, fully specifies the features of multivariate law, such as the correlations in the MVN model. This feature allows us to form likelihood type estimation methods based on the univariate and bivariate marginal distributions. Such estimation methods are called composite likelihood (CL) methods by Lindsay (1988). We considered two such approaches for familial data: the first is based on both CL of the univariate margins and bivariate margins and estimate the marginal parameters and the dependent parameters in two steps while the second only uses the bivariate CL and estimates the parameters simultaneously. Weighting schemes are also considered to improve the efficiency.

In this section, we first introduce the general properties of estimating procedures based on composite likelihood, then present the two approaches mentioned above followed by methods to estimate the covariance matrix of the estimates from CL methods. Finally, we give a summary of the results of some efficiency comparisons.

General Properties

A composite likelihood (CL), sometimes called pseudo-likelihood, is formed by adding together individual component log likelihoods, each of which is a log likelihood of a marginal distribution of a multivariate model (Lindsay, 1988). CL is appealing for the following reasons: Firstly, it inherits some properties of the ordinary likelihood. Under regularity conditions, the estimates based on CL are asymptotically consistent and unbiased. Secondly, the estimates are much easier to compute under many circumstances compared to the ML estimates.

The standard theory for inference functions (Godambe, 1991) can be applied to derive the asymptotic properties of estimators.

[Mathematics not included, original written in LaTeX]

A Two-stage Estimating Procedure

[Mathematics not included, original written in LaTeX]

The next question is that how well this method performs when the data are correlated. Our investigation shows that it can be inefficient comparing with the MLE when the data are highly correlated. To improve the efficiency, we also considered adding weights to the estimating functions.

Estimating Approach Based on Bivariate Composite Likelihood (BCL)

[Mathematics not included, original written in LaTeX]

Methods to Estimate the Asymptotic Covariance Matrix

Different methods can be considered to estimate the asymptotic covariance matrix of the parameter estimate.

(a) Evaluate the Godambe information matrix analytically. This method can be computationally expensive or even not be possible sometimes. For example, with survival data, the matrix cannot be evaluated without specifying the censoring distribution.

(b) Use resampling methods such as jackknife (Xu, 1996) or bootstrapping. Naturally, the sampling uint is family.

(c) Evaluate the Godambe information matrix empirically or using resampling techniques.

Efficiency

In this section, we compare the efficiency of the CLEs with the MLE in terms of asymptotic variances.

Since it is impossible to conduct the efficiency comparison analytically for all models in general, our investigation was carried out on four different types of models: multivariate normal (MVN), multivariate probit (MVP) and Poisson log-normal mixture (PLNM) and MVN with right censoring.

For MVN and MVP, var(theta_MLE) and var(theta_CL) are derived from the inverse Fisher and Godambe information matrices respectively. Different dependence structures are considered, including structures with one dependence parameter (exchangeable), with two dependence parameters (parent-offspring and sib-sib correlations in a type-3 family) and with three dependence parameters (parent-offspring, sib-sib and parent-parent correlations in a type-4 family).

For PLNM and MVN with right censoring, simulation studies were carried out. Only exchangeable dependence structure is considered.

From different models and different dependence structures, we obtained similar results. The following is a summary of the main points.

Among univariate marginal parameters, we separate the regression parameters from the other parameters.

(a) Regression parameters: the two-stage method is easily affected by the following factors. (1) dependency: it tends to lose more efficiency when dependence becomes stronger. The efficiency can be 0 in certain cases. (2) data type: there is more efficiency loss for continuous responses than discrete responses. (3) censoring rate: when right censoring occurs, there is less efficiency loss when the censoring rate increases. (4) family size: the efficiency decreases with the average family size. The BCL method is less affected by the above factors. It is generally better than the two-stage method and the efficiency of the BCL estimate is close to 1 most of the time.

(b) Other univariate parameters: both methods are reasonable.

(c) Dependence parameters: the BCL method is often better for stronger dependence, the two-stage method is better for weaker dependence.

(d) Effect of family size: the efficiency is negatively associated with the mean and relative dispersion of family size (measured by the variance-mean ratio).

Some final words about these two approaches. They are not limited to familial data. We can apply them to other correlated data. We recommend the BCL method when the dependency is strong. It provides better overall estimation of the parameters, especially the regression parameters. However, it is numerically harder to implement since numerical optimization gets harder as the total number of parameters increases. Therefore we recommend the two-stage method when the dependency is weaker.

References

Godambe, VP (ed.) (1991). Estimating Functions. Oxford University Press, Oxford.

Lindsay, BG (1988). Composite likelihood methods. In Statistical Inference from Stochastic Processes, ed. by Prabhu, NU, American Mathematical Society: RI, pp. 221-239.

Xu, JJ (1996). Statistical Modelling and Inference for Multivariate and Longitudinal Discrete Response Data. Ph.D. thesis, Department of Statistics, University of British Columbia.

Appendix 4 (2003 abstracts accepted)

(a) Phylogenetic Footprinting of the NF1 5'' Upstream Region (5UR).
(b) The location of constitutional neurofibromatosis 2 (*NF2*) splice-site mutations is associated with the number of intracranial meningiomas: results from an international NF2 database.
(c) Genotype-phenotype correlations for cataracts in neurofibromatosis 2.
(d) Genotype-phenotype correlations for spinal tumors in neurofibromatosis 2
(e) Genotype-phenotype correlations for peripheral nerve tumors in neurofibromatosis 2
(f) Revised diagnostic criteria for neurofibromatosis 2.

Phylogenetic Footprinting of the NF1 5'' Upstream Region (5UR).

Lee, Bernard

(Abstract accepted for publication at 2003 American Society of Human Genetics)

The 5UR of the human neurofibromatosis 1 gene was defined as the 59756 bp region between the NF1 translation start site and the end of the first upstream GenScan prediction (NT_010799.114). The 5URs of mouse and rat were defined as 59756 bp upstream of the translation start site of the NF1 homologs in these species. The 5UR in Fugu was defined as the 1488 bp segment between the known 5'' flanking gene (FN5) and the NF1 translation start site. Sequence alignments were established by mVista, and windows of identity greater than that of the coding regions and extending 50 bp or more in length among all 3 mammalian species were identified with Frameslider, a Perl program written for this research. These highly homologous regions (HHRs) were compared to the Fugu 5UR using Pairwise BLAST and analyzed for potential transcription factor binding sites and other promoter-associated sequences with MATCH, MatInspector, Eurkaryotic Promoter Database and TRRD.

Three HHRs were discovered in the NF1 5UR. HHR1, located 42626-42696 bp upstream of translation start site, contains an AP-1 site shared by all four species. HHR2, located 640-689 bp upstream of translation start site, has no promising predictions for recognized transcription factor binding sites. HHR3, located 233-519 bp upstream of the NF1 translation start site, contains a previously-described CREB site that is shared by all three mammalian species.

HHR3 also includes a 24 bp sequence 310-333 bp upstream of the translation start that is identical in human, mouse and rat and differs by only1 bp in Fugu. Bioinformatic analysis and correlation with previously-published in vitro transcription studies indicate that this sequence, which we call NF1 Highly Conserved Sequence (NF1HCS), is likely to be involved in transcriptional regulation. NF1HCS lies 151bp downstream from the NF1 major transcriptional start site but appears to be a strong candidate for the NF1 core promoter element despite its position further downstream than any previously-described eukaryotic downstream core promoter element.

The location of constitutional splice-site neurofibromatosis 2 (*NF2*) mutations is associated with the number of intracranial meningiomas: results from an international NF2 database.

Michael E. Baser,[1] Ryan Woods,[2] Harry Joe,[2] Lisa Kuramoto,[2] J. M. Friedman,[3] Andrew J. Wallace,[4] Emilia Bijlsma,[5] Sylviane Olschwang,[6] Laura Papi,[7] Dilys M. Parry,[8] Richard T. Ramsden,[9] Guy A. Rouleau,[10] D. Gareth R. Evans.[4] [1]Los Angeles, CA, U.S.A., [2]Department of Statistics and [3]Department of Medical Genetics, University of British Columbia, Vancouver, B.C., Canada, [4]Department of Medical Genetics, St. Mary's Hospital, Manchester, U.K., [5]Department of Clinical Genetics, Academic Medical Centre, University of Amsterdam, The Netherlands, [6]INSERM U434, Fondation Jean-Dausset-CEPH, Paris, France, [7]Department of Clinical Physiopathology, University of Florence, Italy, [8]Genetic Epidemiology Branch, National Cancer Institute, Bethesda, MD, U.S.A., [9]Department of Otolaryngology, Manchester Royal Infirmary, Manchester, U.K., [10]Center for Research in Neuroscience, McGill University and the Montreal General Hospital Research Institute, Montreal, Canada

It has been hypothesized that the location of constitutional *NF2* splice-site mutations is associated with NF2 disease severity (Am J Med Genet 1998;77:228-233). The purpose of this study was to evaluate genotype-phenotype correlations for the location of splice-site *NF2* mutations. The study had 199 patients from 85 families with splice-site mutations in an international NF2 database (splice-sites flanking exons 1 and 9 were not included in the analysis because there were no mutations flanking exon 1 and only one mutation flanking exon 9). A gamma mixture of negative binomials model with an exchangeable correlation within families was used to model the association of the number of intracranial meningiomas with the location of splice-site mutations; the other covariate was inheritance (people with new mutations/ inherited cases). The locations of splice-site mutations were categorized by their correspondence to domains in the *NF2* protein: mutations flanking exons 2-8 (FERM domain) or exons 10-15 (α-helical domain). The mean $\pm$ SD number of meningiomas in people with mutations flanking exons 2-8 was 1.3 ± 2.0, and in people with mutations flanking exons 10-15, 0.4 ± 0.8. Within exons 2-8, the number of meningiomas in people with splice-site mutations flanking exons 2-5 was 2.0 ± 2.4, and in people with mutations flanking exons 6-8, 0.8 ± 1.3. These results indicate that there is a decreasing 5 to 3 gradient for the number of meningiomas in people with splice-site *NF2* mutations.

Genotype-phenotype correlations for cataracts in neurofibromatosis 2.

Michael E. Baser,[1] Harry Joe,[2] Lisa Kuramoto,[2] J. M. Friedman,[3] Andrew J. Wallace,[4] Richard T. Ramsden,[5] D. Gareth R. Evans.[4] [1]Los Angeles, CA, U.S.A., [2]Department of Biostatistics and [3]Department of Medical Genetics, University of British Columbia, Vancouver, B.C., Canada, [4]Department of Medical Genetics, St. Mary's Hospital, Manchester, U.K., [5]Department of Otolaryngology, Manchester Royal Infirmary, Manchester, U.K.

Genotype-phenotype correlations are well-established for central nervous system tumors in neurofibromatosis 2 (NF2), but such correlations have not been established for non-tumor manifestations of the disease, such as presenile cataracts. The purpose of this study was to evaluate genotype-phenotype correlations for cataracts in NF2. The study had 255 people from 190 families in the United Kingdom NF2 registry who were screened for consitutional *NF2* mutations using SSCP and examined for cataracts (posterior subcapsular or cortical) using slitlamp biomicroscopy. There were 90 people with nonsense or frameshift mutations (including 17 somatic mosaics defined at the molecular level), 47 with splice-site mutations, 15 with missense mutations, 25 with large deletions, and 78 with unidentified mutations. A multivariate probit model with an exchangeable correlation structure within families was used to estimate regression coefficients and calculate relative risks (RR) and confidence intervals (CI) for presence of cataracts. The RR of cataracts was nearly constant with increasing age at diagnosis of NF2, probably because the study population was relatively young. People with classical NF2 and nonsense or frameshift mutations were the reference group in comparisons between different types of *NF2* mutations; the prevalence of cataracts was lower in people with each other type of *NF2* mutation. The RR of cataracts was significantly lower in somatic mosaics (RR = 0.15, 95% CI = 0.04 - 0.51), in people with large deletions (RR = 0.39, 99% CI = 0.16 - 0.98), and in people with new unfound mutations and older onset of symptoms (ages $\geq$ 20 years), who are likely to have somatic mosaicism or large deletions (RR = 0.07, 95% CI = 0.01 - 0.35). These results extend the genotype-phenotype correlations that have been reported for the tumor manifestations of NF2.

Genotype-phenotype correlations for spinal tumors in neurofibromatosis 2.

Michael E. Baser,[1] Dilys M. Parry,[2] Harry Joe,[3] Lisa Kuramoto,[3] J. M. Friedman,[4] James E. Gillespie,[5] Andrew J. Wallace,[6] Richard T. Ramsden,[7] D. Gareth R. Evans.[6] [1]Los Angeles, CA, U.S.A., [2]Genetic Epidemiology Branch, National Cancer Institute, Bethesda, MD, U.S.A., [3]Department of Biostatistics and [4]Department of Medical Genetics, University of British Columbia, Vancouver, B.C., Canada, [5]Department of Radiology and [7]Department of Otolaryngology, Manchester Royal Infirmary, Manchester, U.K., [6]Department of Medical Genetics, St. Mary's Hospital, Manchester, U.K.

Genotype-phenotype correlations have not been established for spinal tumors in neurofibromatosis 2 (NF2), although there is suggestive evidence (Radiology 2001;218:434-442). The purpose of this study was to evaluate genotype-phenotype correlations for spinal tumors in NF2. The study had 336 people from 229 families in the United Kingdom NF2 registry who were screened for constitutional *NF2* mutations using SSCP and had full spine MRI scans. There were 111 people with nonsense or frameshift mutations (including 19 somatic mosaics defined at the molecular level), 63 with splice-site mutations, 24 with missense mutations, 41 with large deletions, and 97 with unidentified mutations. A gamma mixture of negative binomials model with an exchangeable correlation within families was used to model the association of the number of spinal tumors with the type of constitutional *NF2* mutation; the other covariates were age at spinal MRI scan and type of treatment center (specialty or non-specialty). People with classical NF2 and nonsense or frameshift mutations were the reference group in comparisons between different types of *NF2* mutations (mean $\pm$ SD number of spinal tumors, 7.9 $\pm$ 12.2); the number of spinal tumors was lower in people with each other type of *NF2* mutation. The number of spinal tumors was significantly lower in people with large deletions (1.1 $\pm$ 2.0) and missense mutations (1.8 $\pm$ 2.7). In a subset of 160 patients from 125 families who also had data on the number of intramedullary tumors, there were not genotype-phenotype correlations for these tumors, which are much less common than intradural extramedullary tumors (overall, 0.3 $\pm$ 0.6 tumors and 4.0 $\pm$ 7.9 tumors). These results indicate that there are genotype-phenotype correlations for intradural extramedullary spinal tumors in NF2.

Genotype-phenotype correlations for peripheral nerve tumors in neurofibromatosis 2.

Michael E. Baser,[1] Harry Joe,[2] Lisa Kuramoto,[2] J. M. Friedman,[3] Andrew J. Wallace,[4] Richard T. Ramsden,[5] D. Gareth R. Evans.[4] [1]Los Angeles, CA, U.S.A., [2]Department of Biostatistics and [3]Department of Medical Genetics, University of British Columbia, Vancouver, B.C., Canada, [4]Department of Medical Genetics, St. Mary's Hospital, Manchester, U.K., [5]Department of Otolaryngology, Manchester Royal Infirmary, Manchester, U.K.

Studies of several large neurofibromatosis 2 (NF2) patient populations have found that the prevalence of peripheral nerve tumors is associated with broad categories of NF2 disease severity, but genotype-phenotype correlations have not been established. The purpose of this study was to evaluate genotype-phenotype correlations for peripheral nerve tumors in NF2. The study had 328 people from 229 families in the United Kingdom NF2 registry who were screened for constitutional *NF2* mutations using SSCP and had information on the number of peripheral nerve tumors. There were 111 people with nonsense or frameshift mutations (including 17 somatic mosaics defined at the molecular level), 60 with splice-site mutations, 25 with missense mutations, 34 with large deletions, and 98 with unidentified mutations. A gamma mixture of negative binomials model with an exchangeable correlation within families was used to model the association of the number of peripheral nerve tumors with the type of constitutional *NF2* mutation. People with classical NF2 and constitutional nonsense or frameshift *NF2* mutations were the reference group in comparisons between different types of *NF2* mutations (mean $\pm$ SD number of peripheral nerve tumors, 3.7 ± 5.0). The number of peripheral nerve tumors was significantly lower in people with each other type of *NF2* mutation: splice-site mutations (1.2 ± 1.8), missense mutations (1.4 ± 3.9), large deletions (1.6 ± 2.0), somatic mosaics (1.3 ± 2.5), and in people with new unfound mutations and older onset of symptoms (ages $\geq$ 20 years), who are likely to have somatic mosaicism or large deletions (1.5 ± 2.5). These results extend the genotype-phenotype correlations that have been reported for central nervous system tumors in NF2.

Ramsden RT, Evans DGR, Wallace AJ, Joe H, Baser ME. Revised diagnostic criteria for neurofibromatosis 2. 53rd Annual Meeting, American Society of Human Genetics, 4-8 November 2003, Los Angeles (CA). Accepted.

We reported that each of the four sets of clinical diagnostic criteria for neurofibromatosis 2 (NF2) had low sensitivity at the time of the initial assessment (Neurology 2002;59:1579-1565). The purpose of this study was to determine the extent to which modifications to the Manchester diagnostic criteria increased sensitivity. The study had 221 NF2 patients in the United Kingdom NF2 registry who presented without bilateral vestibular schwannomas (155 people who did not have a family history of NF2 at initial assessment and 66 inherited cases). The modifications were: (1) in people without a family history of NF2, permitting the diagnosis when there are multiple meningiomas and only one, instead of two, other tumors or cataract (as in the NNFF criteria); in people with a 1st degree relative with NF2, permitting the diagnosis when there is only one, instead of two, tumors or cataract (as in the 1991 NIH criteria), but restricting 1st degree relatives to parents, (2) adding juvenile mononeuropathy ($\leq$ 15 years) as a diagnostic criterion, (3) in addition to clinical criteria, permitting the diagnosis when constitutional *NF2* mutations are identified. We used Kaplan-Meier analysis to determine the time course, from initial assessment to the most recent clinical evaluation, of the increasing proportion of people who would be diagnosed with NF2 using the Manchester criteria and the three modifications; the jackknife method was used to compute pointwise standard errors for differences in proportions of pairwise Kaplan-Meier curves between different sets of criteria. In people without a family history of NF2 at initial assessment (the most difficult group to diagnose), sensitivity was increased by incorporating features of the 1991 NIH criteria and the NNFF criteria (modification 1 above). The modified Manchester criteria were significantly more sensitive than the original Manchester criteria from four years after initial assessment to 17 years after initial assessment. In inherited cases, sensitivity was further increased by adding mononeuropathy as a diagnostic criterion and incorporating the positive results of mutation analysis. These results indicate that, in NF2 patients who present without bilateral vestibular schwannomas, modifications to the Manchester criteria can increase diagnostic sensitivity, although not at the time of the initial assessment.