The Royal Society of New Zealand

MARSDEN FUND

Te Pūtea Rangahau a Marsden

PRELIMINARY RESEARCH PROPOSAL APPLICATION RECEIPT RECORD

(STANDARD PROPOSAL)

Only **one** Application Receipt Record is to accompany the copies of your Preliminary Research Proposal.

This page will be sent back to you as confirmation of receipt from the Marsden Fund. Please fill in the details.

Name & Title:	Professor Chris Wild
Full Address:	Department of Statistics The University of Auckland Private Bag 92019 Auckland

Standard Proposal	Application number:	Panel:
	04-UOA-	MIS

From: Administration Officer
Marsden Fund
The Royal Society of NZ
PO Box 598
WELLINGTON

The Royal Society of New Zealand

MARSDEN FUND

Te Pūtea Rangahau a Marsden

STATISTICAL INFORMATION

(STANDARD PROPOSAL)

This information will be used for statistical purposes only. It will be removed from the application and it will NOT be sent out to panellists or referees.

Only **one** "Statistical Information" page should accompany the copies of your Preliminary Research Proposal.

Please fill in the details for all principal and associate investigators.

Name	Do you consider yourself to be of Māori descent? (Yes/No)	Year of award of highest postgraduate degree (excluding DSc)	Gender (F/M)
A/Prof Alan Lee	No	PhD 1974	M
Prof Alastair Scott	No	PhD 1965	M
Prof Chris Wild	No	PhD 1979	M

Standard Proposal	Application number:	Panel:
	04-	MIS

(NB The application number is included here for ease of administration only and will not be linked to the statistical information provided.)

Standard Proposal	Contact PI's surname	Initials	Application Number 04-	Panel

MARSDEN FUND PRELIMINARY RESEARCH PROPOSAL

STANDARD APPLICATION FORM

1. TITLE OF RESEARCH PROPOSAL

Model fitting with complex sampling structures

2. IDENTIFICATION

Contact Principal Investigator

Name (with title):	Professor Chris Wild
Full Address: Telephone: Fax: Email:	Department of Statistics The University of Auckland Private Bag 92019 Auckland (09) 373 7599 x 88797 (09) 373 7018 c.wild@auckland.ac.nz
Permission for RSNZ to use contact details (Yes/No)	Yes

Other Principal Investigator(s)

Name (with title):	Institution	Country
Professor Alastair Scott	The University of Auckland	NZ
A/Prof. Alan Lee	The University of Auckland	NZ

Associate Investigator(s)

Name (with title):	Institution	Country

3. FIELDS OF RESEARCH

Please enter up to three codes from the list of research codes supplied in "Fields of Research Classification Codes" document, along with up to 5 keywords for each code. Use codes that are as specific as possible.

Code	Description	Keywords
230203	Statistical theory	Complex sampling, biased sampling, regression, Semiparametric efficiency
230204	Applied statistics	

4A. ABSTRACT OF RESEARCH PROPOSAL

Using only this page, describe your proposed research indicating aims and the importance of the work. Outline the merit of your proposal (incorporating originality, insight and research excellence); the potential of the researchers to contribute to the advancement of knowledge; the development of research skills; and how you intend to do the research. Typeface 12 point, Times or a similar size font. Do not remove instructions or margins.

Regression analysis is one of the most useful and widely used statistical techniques for explanation and prediction, finding applications in many branches of biological, social and medical science. Over the last 20 years, there has been a great deal of work done on developing methods for fitting models to data arising from complex sampling designs and to situations where information on some units is partially or fully missing, but efficient methods for fitting models are still not available for many designs in common use [2, 3, 6, 9, 12-15, 17. Moreove r, even for most situations where methods have been developed, satisfactory optimality results are still not available [4, 5].

Particular problems arise when the chance of getting information on a unit depends, either by accident or by design, on the potential response of that unit. The simplest example of such a "response-selective" design is a case-control study [1], where cases (people with a disease of interest, for example) are sampled at a much higher rate than people without the disease. More complicated versions of this are pervasive in medicine and social science. A local example incorporating biased sampling and missing data is a recent study of low birthweight in infants. All babies categorised as "low birthweight" (about 3% of all babies) were automatically selected for the study. Control babies (those whose birthweights are not low) are randomly selected for inclusion in the study within a short time interval after birth using a sampling rate of 1 in 30, designed to give approximately equal numbers of cases and controls. We are interested in identifying factors that influence birthweight. The sampling is "response-selective" since the chance of inclusion depends on birthweight, which is the response of interest in this study. The missing data arises from the fact that, in addition to more than 100 potential explanatory variables measured on all babies in the study, information on a number of other variables is collected routinely on all births by the hospitals and available from computer files. Thus for some variables we have data on 30,000 babies, while for others we have data only on the 2,000 or so babies in the study.

Other problems arise with data large scale sample surveys which, for reasons of cost and efficiency, almost always include variable selection probabilities and cluster sampling [8].]. Some measurement error problems can also be fitted into our framework [7].

The aims of the proposed research are fourfold:

1. To continue the development of methods for fitting regression models in situations where missing data and complex sampling, particularly response-selective sampling, cause problems.
2. To develop optimality results and rigorous asymptotic theory for the methods that we (and others) are developing.
3. To test the performance, efficiency and robustness of the methods using simulation.
4. To produce robust and user-friendly software that will make the methods available to researchers in medicine, biology and social science.

The PIs are well placed to carry out this programme successfully. Scott and Wild have a good track record in developing practical methods to handle response selective sampling and survey data (e.g. [10, 18-23]). Lee has the background in functional analysis that is required to develop the underlying theory [11] and has recently obtained some very promising results in problems that have appeared intractable until now. Both Lee and Wild have the necessary computational expertise.

Standard Proposal	Contact PI's surname	Initials	Application Number 04-	Panel

CONFIDENTIAL
PANEL MEMBER'S WORKSHEET

*****FOR ASSESSMENT ONLY*****
*****NOT TO BE FILLED IN BY APPLICANTS*****

Grading

Grade 1

Merit of the Proposal						
High Must	Must	Low Must	High Should	Should	Low Should	No

Grade 2

a) Potential of Researchers						b) Development of Research Skills						Total (a + b)
High 5	4	Med 3	2	Low 1	0	High 5	4	Med 3	2	Low 1	0	

Comments:

Possible Referees (highly ranked proposals only):

4B. REFERENCES

[1] Breslow, N.E. (1996). Statistics in epidemiology: the case-control study. [Fisher Lecture] *J. American Statist. Assoc*., 91, 14-28.

[2] Breslow, N.E. and Chatterjee, N. (1999). Design and analysis of two-phase studies with binary outcome applied to Wilms tumour prognosis. *Appl. Statist*., 48, 4:457-468.

[3] Breslow, N.E. and Holubkov, R. (1997). Maximum likelihood estimation of logistic regression parameters under two-phase, outcome-dependent sampling. *J. Roy. Statist. Soc.* B. **59**, 447-461.

[4] Breslow, N.E., McNeney, B., and Wellner, J.A. (2003). Large sample theory for semiparametric regression models with two-phase, outcome dependent sampling. *Annals of Statistics*, 31, 1110-1139

[5] Breslow, N.E., Robins, J.M. and Wellner, J.A. (2000). On the semi-parametric efficiency of logistic regression under case-control sampling. *Bernoulli*, 6, 3:447-455.

[6] Breslow, N.E., and Zhao, L.P. (1988). Logistic regression for stratified case-control studies. *Biometrics* **44**, 891–99.

[7] Carroll, R.J., Ruppert, D. and Stefanski, L.A. (1995). *Measurement Error in Nonlinear Models*, New York: Chapman and Hall.

[8] Chambers, R. and Skinner, C.J. (eds) (2003). *Analysis of Survey Data*, Chichester: John Wiley & Sons.

[9] Chatterjee, N., Chen, Y.H. and Breslow N.E. (2003). A pseudoscore estimator for regression problems with two-phase sampling. *J. American Statist. Assoc.,* 98, 158-168.

[10] Lawless, J.F, Kalbfleisch, J.D. and Wild, C.J. (1999). "Estimation for Response-selective and Missing Data Problems in Regression", *Journal of the Royal Statistical Society, B*, **61**, 413-438, 1999.

[11] AJ Lee (1978). Sampling theorems for non-stationary random processes. *Trans. Am. Math. Soc.* 242, pp 25-241.

[12] Lin, X. & Carroll, R. J. (2001). Semiparametric regression for clustered data using generalized estimating equations. *Journal of the American Statistical Association*, 96, 1045-1056.

[13] Little, R.J.A. and Rubin, D.B. (2002). *Statistical Analysis with Missing Data, 2nd Ed*. New York: John Wiley & Sons, Inc.

[14] Little, R.J.A. and Vartivarian, S. (2002). On weighting the rates in non-response weights. *Statistics in Medicine*, 22, 9:1589-1599.

[15] Neuhaus, J. & Jewell, N. (1990). The effect of retrospective sampling on binary regression models for clustered data. *Biometrics* **46**, 977–90.

[16] Neuhaus, J. Scott, A.J. & Wild, C.J.. (2001). The analysis of retrospective family studies. *Biometrika*, 89, 23-37.

[17] Satten, G. A. & Carroll, R. J. (2000). Conditional and unconditional categorical regression models with missing covariates. *Biometrics*, 56, 384{400.

[18] Scott, A.J. and Wild (1986). "Fitting logistic models under case-control or choice based sampling", *Journal of the Royal Statistical Society,* **48**, 170-182.

[19] Scott, A.J. and Wild (1991). "Fitting logistic regression models in stratified case-control studies", *Biometrics,* **47**, 497-510.

[20] Scott, A.J. and Wild (1997). "Maximum Likelihood Estimation for Case-Control Data", *Biometrik*a, , **84**, 57-71.

[21] Scott, A.J. and Wild, C.J. (2001a)., "Maximum Likelihood For Generalised Case-Control Studies", *Journal of Statistical Planning and Inference,* **96**, 3–27, 2001.

[22] Scott, A.J. and Wild, C.J. (2001b). Fitting Logistic Regression Models in Case-control Studies with Complex Sampling. *Journal of the Royal Statistical Society, C,* **50**, 389–401.

[23] Wild, C. J. (1991). "Fitting prospective regression models to case-control data", *Biometrika*, **78**, 705-717.

Standard Proposal	Contact PI's surname	Initials	Application Number 04-	Panel

5. DURATION OF RESEARCH PROPOSAL

Years:	3 years	(Initial maximum of 3 years for standard application. Please refer to the section on *Duration of Research Proposal* in the *2004 Preliminary Research Proposal Guidelines for Applicants*).

6. MĀORI RESPONSIVENESS

If your research is of relevance for Māori, provide a brief summary of the consultation process so far and plans for ongoing consultation. Māori applicants may wish to demonstrate their linkages and what processes they have used in developing and designing their proposal. Please refer to the relevant sections in the *2004 Preliminary Research Proposal Guidelines for Applicants.*

This project is developing statistical theory and methods and does not have special relevance to Māori.

7A. PERSONNEL

List the time involvement of all personnel in terms of a Full Time Equivalent (FTE). Give names for all personnel (except when they are as yet unknown for such people as postdoctoral fellows and postgraduate students). Please refer to the *2004 Preliminary Research Proposal Guidelines for Applicants* for recommended minimum time for Principal Investigators.

Name	**FTE** Year 1	**FTE** Year 2	**FTE** Year 3
Principal Investigator(s)			
Alan Lee	.2	.2	.2
Alastair Scott	.2	.2	.2
Chris Wild	.1	.1	.1
Associate Investigator(s)			
Postdoctoral fellow(s)			
Research/Technical Assistant(s)	.5	.5	.5
Others (name)			
Sub-contractor(s)			
Postgraduate student(s)	3	3	3
TOTAL	**4**	**4**	**4**

7B. DEVELOPMENT OF RESEARCH SKILLS

Please explain briefly how the intended research programme would contribute to the development or broadening of research skills in New Zealand.

The proposed research will contribute to the training of two PhD students and one Masters student. The project itself will develop new statistical methods that will allow New Zealand researchers to design more efficient studies and analyse data from a wider variety of data structures more efficiently.

8. CONTACT DETAILS OF OTHER INVESTIGATORS

Principal Investigator(s) other than Contact Principal Investigator

Name (with title):	A/Professor Alan Lee	Professor Alastair Scott
Full Address: Telephone: Fax: Email:	Department of Statistics The University of Auckland Private Bag 92019 Auckland (09) 373 7599 x 88749 (09) 373 7018 lee@stat.auckland.ac.nz	Department of Statistics The University of Auckland Private Bag 92019 Auckland (09) 373 7599 x 88751 (09) 373 7018 a.scott@auckland.ac.nz
Permission for RSNZ to use contact details (Yes/No)		

Name (with title):		
Full Address: Telephone: Fax: Email:		
Permission for RSNZ to use contact details (Yes/No)		

Associate Investigator(s)

Name (with title):		
Full Address: Telephone: Fax: Email:		
Permission for RSNZ to use contact details (Yes/No)		

Name (with title):		
Full Address: Telephone: Fax: Email:		
Permission for RSNZ to use contact details (Yes/No)		

Standard Proposal	Contact PI's surname	Initials	Application Number 04-	Panel

9A. CURRICULUM VITAE

Please attach a curriculum vitae for each of the named research personnel (other than technical assistants) who are relevant to this application. Use one page per person.

Name of Researcher: Alastair John Scott

Tertiary Education:

1961	BSc	University of Auckland
1962	MSc (1st class Hons)	University of Auckland
1965	PhD	University of Chicago

Distinctions/Honours: Fellow of:
Royal Society of New Zealand;
Royal Statistical Society;
Institute of Mathematical Statistics;
American Statistical Association
Honorary Life Member of the New Zealand Statistical Association
Elected Member of the International Statistical Institute

1981 JASA paper on the analysis of complex survey data with J.N.K.Rao selected as one of the 19 landmark papers in the history of survey sampling by the International Assoc. of Survey Statisticians for its 2001 centenary volume.

Employment Record:

1962-65	Scientific Officer, Applied Maths Div., DSIR., Wellington
1965 - 71	London School of Economics
1972 -	University of Auckland (currently Professor of Statistics)

Other Information:

This and the next page can be duplicated for additional research personnel

9B. PUBLICATIONS

Name of researcher:	Alastair John Scott

Indicate **total** number of books, refereed journal articles or refereed conference papers published (including before 1999)	115

List the following: (i) **refereed** publications published since, but not including 1998, and (ii) prior publications that relate directly to the proposal. For each section, please give full titles and co-authors, and separate refereed publications into the following categories: a) journal articles; b) books; c) book chapters; d) books edited; e) refereed conference proceedings; f) other. For each category please list the most recent publications first.

(i) **Publications (1999 - present)**

(a) Journal Articles

J.N.K. Rao, A.J. Scott and E. Benhin (2003),"Undoing Complex Survey DataStructures: Some Theory and Applications of Inverse Sampling". *Survey Methodology* 29, 107-128.

P.D.Davis, R. Lay-Yee, S. R. Briant and A.J.Scott (2003), 'Preventable in-hospital medical injury under a no-fault system' *Qual.Saf. Health Care*, 251-256.

P. Davis, R. Lay-Yee, A. Scott, R. Briant, S. Schug (2003) 'Acknowledgement of "no fault" medical injury: review of hospital records in New Zealand', *British Medical Journal*, 326, 79-80.

A.J. Scott and C.J. Wild (2002). 'On the robustness of weighted methods for fitting models to case-control data' *Journal of the Royal Statistical Society B*, 64, 207-219.

P.Davis, B. Gribben, R. Lay-Yee, A. Scott (2002) Co-morbidity and health outcomes in three Auckland hospitals results *NZ Medical Journal*, 115, 2002, 211-216.

J. Neuhaus, A.J. Scott and C.J. Wild. (2002). 'The analysis of retrospective family studies. *Biometrika*, 89, 23-37.

P.Davis, B. Gribben, R. Lay-Yee, A. Scott (2002) 'How much variation is there between general practitioners? A multi-level analysis of decision making in primary care.' *Journal of Health Services Research and Policy*, 202-209.

A.J. Scott and C.J. Wild (2001). 'Maximum likelihood for generalized case-control designs' *Journal of Statistical Inference & Planning*, 96, 3-27.

A.J. Scott and C.J. Wild (2001). 'The analysis of clustered case-control studies' (with C.J. Wild), *Journal of the Royal Statistical Society C*, 50, 389 – 401.

P.D.Davis, R. Lay-Yee, S. R. Briant, S,Schug and A.J.Scott (2001), 'Adverse events regional feasibility study: methodological results, *NZ Medical Journal*, 114, 200-2.

P.D.Davis, R. Lay-Yee, S. R. Briant, S,Schug and A.J.Scott (2001),), 'Adverse events regional feasibility study: indicative finding, *NZ Medical Journal*, 114, 203-5.

P. Davis, B. Gribben, A. Scott, R. Lay-Yee (2000) 'Do physician practice styles persist over time? Practice variation in primary care', *Journal of Health Services Research*, 5, 200-7.

J.N.K. Rao and A.J. Scott (1999), 'A simple method for analyzing clustered Poisson',*Statistics in Medicine*, 18, 1373-1386.

P. Davis, B. Gribben, A. Scott, R. Lay-Yee (1999) 'The supply hypothesis and medical practice variation in primary care' *Social Science & Medicine*, 4, 1-12.

(c) Book chapters

A.J. Scott and C.J. Wild (2003), 'The analysis of multi-stage case-control studies', in *Analysis of Complex Surveys*, C.J. Skinner and R. Chambers (eds). John Wiley and Sons: New York, 109-122

A.J. Scott and P.D. Davis (2001) 'Estimating interviewer effects for binary responses' in *Achieving Data Quality in a Statistical Agency: a Methodological Perspective*. Statistics Canada; Ottawa, 107-114

(e) Refereed conference proceedings

A.J. Scott and C.J. Wild (2002). Maximum likelihood methods for stratified cluster sampling with informative stratification. in *Proceedings of the International Conference on Recent Advances in Survey Sampling*, G. Roberts and D. Bellhouse (eds), 207-215.

J.N.K. Rao and A.J. Scott (2000), 'Undoing complex survey data structures: Some theory of inverse sampling (with J.N.K. Rao*), Proceedings of the Statistical Society of Canada*, 10, 39-45.

A.J. Scott and C.J. Wild (1999), 'Complex Sampling and Case-control Studies', *Bulletin of the International Statistical Institute*, 58, 327-330

(f) Other

A. Raymont, R. Lay-Yee, P.D.Davis and A.J. Scott (2003). '*The National Primary Medical Care Survey. Report 1: Family Doctors - methodology and description of the activity of private GPs*. Report to the NZ Ministry of Health , 96pp.

P Davis, R. Lay-Yee, R. Briant, S. Schug, A. Scott, S. Johnson, W. Bingley (2001). '*Adverse events in New Zealand public hospitals*, NZ Ministry of Health, Occasional Papers 3, 1-88.

(ii) **Prior publications relevant to this proposal**

J.N.K. Rao, A.J.Scott and C.J. Skinner (1998). 'The analysis of survey data', *Statistica Sinica*, 8, 1059-1070.

A.J. Scott and C.J. Wild (1997). 'Fitting regression models to case-control data by maximum likelihood', *Biometrika*, 83, 57-72

A.J. Lee, L.A. McMurchy and A.J. Scott (1997). 'Re-using data from case-control studies', *Statistics in Medicine*, 16, 1377-1389

S.G. Gange, K.L.P. Linton, D.L.DeMets and R.Klein) and A.J. Scott (1995). 'Methods for correlated ordinal measures with opthalmic applications', *Statistics in Medicine*, 14, 1961-74.

J.N.K. Rao and A.J.Scott (1992). 'A simple method for the analysis of clustered binary data' (with J.N.K. Rao), *Biometrics*, 48, 577-585.

A.J. Scott and C.J. Wild (1991) 'Fitting logistic regression models in stratified case/control studies' (with C.J.Wild), *Biometrics*, 41, 705-717.

A.J. Scott and C.J. Wild (1989). 'Likelihood ratio tests in case/control studies'. *Biometrika*, 76, 806-809.

J.N.K. Rao and A.J.Scott (1987). 'On simple adjustments to chi-squared tests with sample survey data'. *The Annals of Statistics*, 15, 385-397.

A.J. Scott and C.J. Wild (1986). 'Fitting logistic models under case-control or choice-based sampling', *Journal of the Royal Statistical Society, B*, 48, 170-182.

J.N.K. Rao and A.J.Scott (1981). 'The analysis of categorical data from complex sample surveys: Chi-squared tests for independence and goodness of fit', *Journal of the American Statistical Association*, 76, 221-230.

D. Holt, A.J. Scott and P.D. Ewings (1980). 'Chi-squared tests with survey data', *Journal of the Royal Statistical Society, A*, 143, 302-320.

(c) Book chapters

A.J. Scott and C.J. Wild (1989). 'Selection based on the response variable in logistic regression', Chapter 9 in *Analysis of Complex Surveys*, ed. by D. Holt, C.J. Skinner and T.M.F. Smith, Wiley, 191-205.

(e) Refereed conference proceedings

'Score tests for survey data' (with J.N.K. Rao), *Proceedings of the Statistical Society of Canada*, 7, 1997, 33-39.

This and the previous page can be duplicated for additional research personnel

Standard Proposal	Contact PI's surname	Initials	Application Number 04-	Panel

9A. CURRICULUM VITAE

Please attach a curriculum vitae for each of the named research personnel (other than technical assistants) who are relevant to this application. Use one page per person.

Name of Researcher: Christopher John Wild

Tertiary Education:
1974 BSc (Mathematics and Statistics) University of Auckland
1976 MSc (1st class honours) University of Auckland
1979 PhD (Statistics) University of Waterloo

Distinctions/Honours: Commonwealth Scholar 1976-9

President of the International Association for Statistics Education
Fellow of the Royal Statistical Society
Elected Member of the International Statistical Institute

Employment Record: 1979- University of Auckland (currently Professor of Statistics)

Other Information:

This and the next page can be duplicated for additional research personnel

9B. PUBLICATIONS

Name of researcher:	Christopher John Wild

Indicate **total** number of books, refereed journal articles or refereed conference papers published (including before 1999)	62

List the following: (i) **refereed** publications published since, but not including 1998, and (ii) prior publications that relate directly to the proposal. For each section, please give full titles and co-authors, and separate refereed publications into the following categories: a) journal articles; b) books; c) book chapters; d) books edited; e) refereed conference proceedings; f) other. For each category please list the most recent publications first.

(i) **Publications (1999 - present)**

(a) Journal Articles

Merry, S., McDowell, H., Wild, C.J., Julliet Bir, J., and Cunliffe, R. "A randomized placebo controlled trial of a school-based depression prevention program", *Journal of the American Academy of Child and Adolescent Psychiatry*, to appear.

E.A Mitchell, E.A., Robinson, E., Clark, P.M., Becroft, D.M.O., Glavish, N., Pattison, N.S., JE Pryor, J.E., Thompson, J.M.D., and Wild, C.J. "Maternal nutritional risk factors for small for gestational age babies in a developed country: a case control study", *Archives of Disease in Childhood*, to appear.

Pryor J.E., Thompson J.M.D., Robinson E., Clark P.M., Becroft D.M.O., Pattison N.S., Galvish N., Wild, C.J., Mitchell E.A. Stress and lack of social support as risk factors for small-for-gestational-age birth. *Acta Pædiatr*, **92**, 62–64, 2003.

Scott, A.J. and Wild, C.J., "On the robustness of weighted methods for fitting models to case-control data ", *Journal of the Royal Statistical Society, B*, **64**, 207–219, 2002.

Neuhaus, J., Scott, A.J. and Wild, C.J., "The Analysis of Retrospective Family Studies", *Biometrika*, **89**, 23-37, 2002.

Mitchell, E.A., Thompson, J.M.D., Robinson, E. , Wild, C.J., Becroft, D.M.O., Clark, P.M., Glavish, N. , Pattison, N.S., and Pryor, J.E., "Smoking, nicotine and tar and risk of small-for-gestational-age babies", *Acta Paediactr,* **91**, 323–328, 2002.

Pfannkuch, M., Seber, G.A.F and Wild, C.J. "Probability with less pain", *Teaching Statistics*, **24**, 24–30, 2002.

Scott, A.J. and Wild, C.J., "Maximum likelihood for generalised case-control studies", *Journal of Statistical Planning and Inference*, **96**, 3–27, 2001. [*Lead article, invited*]

Scott, A.J. and Wild, C.J., "Fitting Logistic Regression Models in Case-control Studies with Complex Sampling", *Journal of the Royal Statistical Society, C*, **50**, 389–401, 2001.

Yee, T.W. and Wild, C.J., Invited discussion of "Smoothing spline ANOVA for multivariate Bernoulli observations with application to ophthalmology data" by Gao, F., Wahba, G., Klein, R. and Klein, B., *Journal of the American Statistical Association,* **96**, 147–151, 2001.

Thompson, J.M.D., Clark,, P.M., Robinson, E., Becroft, D.M.O., Pattison, N.S., Glavish, N., Pryor, J.E., Wild, C.J., and Rees, B.A. and ., Mitchell, E.A. "Risk factors for small-for-gestational-age babies: the Auckland Birthweight Collaborative (ABC) Study", *Journal of Paediatrics and Child Care,* **37**, 369–375, 2001.

Pfannkuch, M. and Wild, C.J., "Statistical Thinking and Statistical Practice: Themes gleaned from interviews with professional statisticians," *Statistical Science*, **15**, 132-152, 2000.

Kalbfleisch, J.D., Lawless, J.F., and Wild, C.J. "Estimation for Response-selective and Missing Data Problems in Regression", *Journal of the Royal Statistical Society, Series B*, **61**, 413-438, 1999.

Wild, C.J. and Pfannkuch, M., "Statistical thinking in empirical enquiry" (with discussion). *International Statisitical Review*, **67**, 221-266, 1999.

(b) Books

Wild, C.J. and Seber, G.A.F. *Chance Encounters: A first course in data analysis and inference*, Pub: Wiley, New York, 2000, 611 pages

Wild, C.J. and Seber, G.A.F, *Instructors Manual for 'Chance Encounters'*, Pub: Wiley, New York, 2000, 207 pages

(c) Book Chapters

Pfannkuch, M. and Wild, C.J., "What do we know about statistical thinking?" In D. Ben-Zvi and J. Garfield (eds.), *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking*. Dordrecht, The Netherlands: Kluwer Academic Publishers, 2004.

Scott, A.J. and Wild, C.J., "Fitting Logistic Regression Models in Case-control Studies with Complex Sampling." In *Analysis of Survey Data*, Chambers, R.L. and Skinner, C.J. (eds.), Chichester: Wiley, 109-121, 2003.

(e) Refereed Conference Proceedings

Scott, A.J. and Wild, C.J. (2003). Maximum likelihood methods for stratified cluster sampling with informative stratification. *Proceedings of the International Conference on Recent Advances in Survey Sampling,* G. Roberts and D. Bellhouse (eds), 207-215.

Cunliffe, R.V., Regan. M. & Wild, C. 'Flexible Learning and Large Numbers (A Case Study).' *Proceedings of the IASE Satellite Conference on Statistics Education: Statistics Education and the Internet*, CD-Rom. Voorberg Netherlands, International Statistical Institute, (10 pages), 2003.

(ii) **Prior publications relevant to this proposal**

(a) *Journal Articles*

Scott, A.J. and Wild, C.J., "Maximum Likelihood Estimation for Case-Control Data", *Biometrika*, **84**, 57-71, 1997.

Yee, T.W. and Wild, C.J., "Vector Generalised Additive Models", *Journal of the Royal Statistical Society, Series B*, **58**, 481-493, 1996. [*Lead article*]

Wild, C.J. and Yee, T.W., "Additive Extensions to Generalized Estimating Equation Methods", *Journal of the Royal Statistical Society, Series B*, **58**, 711-712, 1996.

Wild, C.J. "Fitting prospective regression models to case-control data", *Biometrika*, **78**, 1991, 705-717. [*Lead article*]

Scott, A.J. and Wild, C.J. "Fitting logistic regression models in stratified case-control studies", *Biometrics,* **47**, 497-510, 1991.

Scott, A.J. and Wild, C.J. "Hypothesis testing in case-control studies", *Biometrika,* **76**, 806-809, 1989.

Scott, A.J. and Wild, C.J. "Fitting logistic models under case-control or choice based sampling", *Journal of the Royal Statistical Society,* **48**, 170-182, 1986.

(b) ***Books***

Seber, G.A.F. and Wild, C.J. *Nonlinear Regression*, Pub: Wiley, New York, 1989, 768 pages.

[Paperback Edition 2003]

(c) Book Chapters

Seber, G.A.F. and Wild, C.J. "Least Squares", Chapter 9 of "Statistical Methods for Physical Science", John Stanford and Steve Vardeman (eds), Academic Press, 1994.

Scott, A.J. and Wild, C.J. "Selection based on the response variable in logistic regression", Chapter 9 in *Analysis of Complex Surveys,* ed. by D. Holt, C.J. Skinner and T.M.F. Smith, Pub: Wiley, 191-205, 1989.

This and the previous page can be duplicated for additional research personnel

Standard Proposal	Contact PI's surname	Initials	Application Number 04-	Panel

9A. CURRICULUM VITAE

Please attach a curriculum vitae for each of the named research personnel (other than technical assistants) who are relevant to this application. Use one page per person.

Name of Researcher: A/Prof Alan James Lee

Tertiary Education: University Junior Scholar
MSc (1st Class Honours) in Mathematics from the University of Auckland.
Postgraduate Scholar and Senior Scholar in Mathematics.

PhD (1974) University of North Carolina

Distinctions/Honours:

Employment Record: 1973-1974: Visiting Assistant professor, Department of Mathematics, Indiana University USA

1974 – present: Department of Statistics, University of Auckland, New Zealand.

Head of Department, 1997-2002

Aug. 1980 - July 1981: Visiting Professor, Department of Statistics, University of North Carolina, Chapel Hill, USA.

Jan. - Dec. 1986: Visiting Professor, Faculty of Mathematics, University of Southhampton, Southampton, UK

Jan. - May. 1992: Visiting Professor, Faculty of Mathematics, University of Southhampton, Southampton, UK

Other Information:

This and the next page can be duplicated for additional research personnel

9B. PUBLICATIONS

Name of researcher:	A/Prof Alan James Lee

Indicate **total** number of books, refereed journal articles or refereed conference papers published (including before 1999)	50

List the following: (i) **refereed** publications published since, but not including 1998, and (ii) prior publications that relate directly to the proposal. For each section, please give full titles and co-authors, and separate refereed publications into the following categories: a) journal articles; b) books; c) book chapters; d) books edited; e) refereed conference proceedings; f) other. For each category please list the most recent publications first.

(i) **Publications (1999 - present)**

(a) Journal Articles

Lee, A.J, Nyangoma, O and Seber, G.A.F. (2002) Confidence regions for multinomial parameters. *Journal of Computational Statistics and Data Analysis*, 39,329-342.

Lambie, I., Seymour, F., Lee, A. and Adams, P. (2002) Resiliency in the victim-offender cycle in male sexual abuse. *Sexual Abuse: A Journal of Research and Treatment*, 14, p31-48.

Lee, A.J (2002). Effect of list errors on the estimation of population size. *Biometrics*, 185-191.

Lee, A. J., Seber, G. A. F., Holden, J. K., and Huakau, J. T. (2001). Capture-recapture, epidemiology, and list mismatches: Several lists. *Biometrics*, 57, 707-713.

Lee, A (1999). Modelling Rugby league data via bivariate negative binomial regression. *Australian & New Zealand J. Statistics* 41, 141-152.

Belami, A. and Lee, A.J. (1999) Some properties of the Liang–Zeger method applied to clustered binary regression. *Australian & New Zealand J. Statistics*. 41, 43-58

(b) Books

Seber, G.A.F. and Lee, A.J. (2003) *Linear Regression Analysis, 2nd Ed.* Wiley, New York, 567 pages.

(c) Book Chapters

Lee, A. J. and Seber, G. A. F. (2001). Residuals. In A. El-Shaarawi (ed.), *Encyclopedia of Environmetrics*. Wiley: New York.

(ii) **Prior publications relevant to this proposal**

(a) Journal Articles

AJ Lee (1998). Some simple methods for generating correlated categorical variates. *Computational Statistics and Data Analysis*, 26, 133-148.

AJ Lee, Lovina McMurchy and AJ Scott (1997) Reusing data from case-control studies. *Statistics in Medicine*, 16, 1377-1389.

AJ Lee, AJ Scott and SC Soo. (1993) Comparing Liang-Zeger estimates with maximum likelihood in bivariate logistic regression. *Journal of Statistical Computation and Simulation*, 44, 133-148.

AJ Lee (1993) Generating random binary deviates having fixed marginal distributions and specified degrees of association. *The American Statistician*, 47, 209-215.

AJ Lee (1978). Sampling theorems for non-stationary random processes. *Trans. Am. Math. Soc.* 242, pp 25-241.

AJ Lee (1977). Continuous path approximations to measurable stochastic processes. *J. Math. Anal. Aplns*, 61, pp 1-6.

(b) Books

AJ Lee (1990) U-statistics, Theory and Practice. 310 pp. Marcel Dekker, New York.

This and the previous page can be duplicated for additional research personnel

Standard Proposal	Contact PI's surname	Initials	Application Number 04-	Panel

10. REFEREES

Please nominate the names and full contact details of 5 people who are experts in the proposed field(s) of research. Where possible at least 3 should be from outside New Zealand. Referees will be contacted if you are invited to submit a Full Proposal. Applicants do not need to contact referees beforehand. **Avoid listing referees with whom you may have a possible conflict of interest** (see the section on referees in *2004 Preliminary Research Proposal Guidelines for Applicants*). Initial approaches to referees will be by email. **Please indicate the expertise of each referee next to their name.**

Name & Title; Expertise:	Professor Norman Breslow; Biostatistics
Full Address: Telephone: Fax: Email:	Department of Biostatistics University of Washington Seattle, WA 98195-7232, U.S.A 1-206-543-2035 1-206-616-2724 norm@u.washington.edu

Name & Title; Expertise:	Distinguished Professor R.J. Carroll; General statistics
Full Address: Telephone: Fax: Email:	Department of Statistics, Texas A&M University College Station TX 77843-3143 USA. 1-409- 845-3141 1-409- 845-3144 carroll@stat.tamu.edu

Name & Title; Expertise:	Professor R. Chambers; Survey sampling
Full Address: Telephone: Fax: Email:	Department of Social Statistics University of Southampton Southampton SO17 1BJ UK 44 - 23 8059 4311 44 - 23 8059 3846 rc6@soton.ac.uk

Name & Title; Expertise:	Professor Nicolas Jewell; Biostatistics
Full Address: Telephone: Fax: Email:	Department of Biostatistics University of California at Berkeley Berkeley, CA 94720 U.S.A 1-510- 642 4627 1-510- 643 5163 jewell@stat.berkeley.edu

Name & Title; Expertise:	Dr Katrina Sharples; Biostatistics
Full Address: Telephone: Fax: Email:	Department of Preventive and Social Medicine Dunedin School of Medicine, Otago University PO Box 913 Dunedin 0-3 -479 7221 0-3-479 7298 katrina.sharples@stonebow.otago.ac.nz

11. DECLARATION BY DULY AUTHORISED AGENT

The Royal Society of New Zealand and the Marsden Fund Council undertake to collect, use and store the information you provide to enable your application to be evaluated. This information will not be supplied to any other organisation.

The personal information supplied in this Preliminary Research Proposal form will be used in accordance with the principles of the Privacy Act 1993.

I acknowledge the information contained herein is accurate and can be used in the manner described.

I will indemnify the Royal Society of New Zealand and the Marsden Fund Council from any claims, demands, costs, action or proceedings of any nature which may arise at any time in relation to this application.

I confirm that the consent of each individual referred to in this application has been obtained for the provision of personal information in support of this application.

Name	Signature	Date
Prof. Chris Wild Principal Investigator (Contact person)		4 Feb 2004
  Duly authorised agent from host organisation		