

STATISTICAL SOCIETY OF OTTAWA

SOCIÉTÉ STATISTIQUE D'OTTAWA

CHAPTER - ASA
REGIONAL ASSOCIATION - SSC



SECTION - ASA
ASSOCIATION RÉGIONALE - SSC

ANNOUNCEMENT!

SSO 2003 Symposium: Statistical Issues in Genetics Research

Co-sponsored by the **Bureau of Biostatistics and Computer Applications,
Food Directorate, Health Canada**

Friday, February 28th, 2003,
9:15 a.m. to 15:30 (Registration: 8:30 – 9:15)

Natural Resources Canada

Camsell Hall, 588 Booth Street (At Booth & Carling), Ottawa, Ontario

Shelly B. Bull, Samuel Lunenfeld Research Institute of Mount Sinai Hospital,
Department of Public Health Sciences, University of Toronto
**Assessment of Genetic Heterogeneity Using Regression Models
for Allele Sharing**

David Hamilton, Department of Mathematics and Statistics, Dalhousie University;
David Cole, Professor, Departments of Laboratory Medicine & Pathobiology,
Medicine, and Paediatrics (Genetics), University of Toronto
**Bounds on a Composite Measure of Linkage Disequilibrium: Application to
Genetic Screening**

Paul A. White, Environmental & Occupational Toxicology Division, HECSB,
Health Canada, Ottawa; Andrew Williams, Biostatistics & Epidemiology Division,
HECSB, Health Canada, Ottawa.
**Application of DNA Microarray Technology in
Environmental Health Sciences**

J.C. Loredó-Ostí, Department of Human Genetics, McGill University
The Statistical Analysis of Quantitative Traits

Registration fees (include lunch and coffee):

Until February 20th: Members \$60, Non-members \$80, Students \$40.
After February 20th: Members \$70, Non-members \$90, Students \$50.

Send form and contact information with cheque payable to Statistical Society of Ottawa to :
Maggie Montague, AL 0900B1, Rm 0108B, Brooke Claxton Bldg, Tunney's Pasture,
Ottawa, K1A 0T6 (954-5148; fax: 957-2842; Margaret_Montague@hc-sc.gc.ca).

President/ Président

William Ross
(613) 957-8726 (work/travail)
(613) 957-1574 (FAX/télécopieur)
email: william_ross@hc-sc.gc.ca

Vice-President/ Vice-président

Mara Lee McLaren
(613) 947-2414 (work/travail)
(613) 952-8295 (FAX/télécopieur)
email: maralee.mclaren@cac.gc.ca

Secretary/ Secrétaire

Eric Rancourt
(613) 951-5046 (work/travail)
(613) 951-3100 (FAX/télécopieur)
email: eric.rancourt@statcan.ca

Treasurer/ Trésorier

Edward Chen
(613) 951-4769 (work/travail)
(613) 951-3100 (FAX/télécopieur)
email: chenedw@statcan.ca

President-Elect/ Président-désigné

John Nash
(613) 225-3781 (work/travail)
(613) 225-6553 (FAX/télécopieur)
email: jcnash@uottawa.ca

Past-President/ Président-sortant

Dan Harvey
(613) 824-0926 (work/travail)
(613) 824-2373 (FAX/télécopieur)
(Call 824-0926 before sending a fax)
email: danharvy@istar.ca

Program Coordinator/ Coordonnateur de programme

Dena Schanzer
(613) 946-0461 (work/travail)
(613) 954-5414 (FAX/télécopieur)
email: dena_schanzer@hc-sc.gc.ca

STATISTICAL SOCIETY OF OTTAWA

SOCIÉTÉ STATISTIQUE D'OTTAWA

CHAPTER - ASA
REGIONAL ASSOCIATION - SSC



SECTION - ASA
ASSOCIATION RÉGIONALE - SSC

ANNONCE!

Symposium 2003 de la SSO: Questions statistiques dans la recherche en génétique

Commendité en partie par le **Bureau de biostatistique et des applications informatiques, Direction des aliments, Santé Canada**

Vendredi le 28 février 2003,
9:15 à 15:30 (Inscriptions: 8:30 – 9:15)

Ressources Naturelles Canada

Camsell Hall, 588 Booth Street (coin Booth & Carling), Ottawa, Ontario

Shelly B. Bull, Institut de recherche Samuel Lunenfeld de l'hôpital Mount Sinai,
Département des sciences de la santé publique, Université de Toronto

Évaluation de l'hétérogénéité génétique à l'aide de modèles de régression pour le partage de l'allèle

David Hamilton, Département de Mathématiques et Statistique, Université Dalhousie; David Cole, Départements de médecine en laboratoire et de pathologie, médecine, et pédiatrie (génétique), Université de Toronto

Des bornes pour une mesure composite du déséquilibre localisé : application au dépistage génétique

Paul A. White, Division de toxicologie environnementale et occupationnelle, HECSB, Santé Canada, Ottawa; Andrew Williams, Division de biostatistique et d'épidémiologie Division, HECSB, Santé Canada, Ottawa.

Application de la technologie de micro-alignements d'ADN dans les sciences de la santé environnementales

J.C. Loredano-Osti, Département de génétique humaine, Université McGill
L'analyse statistique de traits quantitatifs

Frais d'inscription (incluent le dîner et les cafés) :

Jusqu'au 20 février : Membres \$60, non-membres \$80, étudiants \$40.

Après le 20 février : Membres \$70, non-membres \$90, étudiants \$50.

Envoyer le formulaire accompagné d'un chèque à l'ordre de la Société Statistique d'Ottawa à : Maggie Montague, AL 0900B1, Rm 0108B, Brooke Claxton Bldg, Tunney's Pasture, Ottawa, K1A 0T6. (954-5148; fax: 957-2842; Margaret_Montague@hc-sc.gc.ca).

President/ Président

William Ross
(613) 957-8726 (work/travail)
(613) 957-1574 (FAX/télécopieur)
email: william_ross@hc-sc.gc.ca

Vice-President/ Vice-président

Mara Lee McLaren
(613) 947-2414 (work/travail)
(613) 952-8295 (FAX/télécopieur)
email: maralee.mclaren@cac.gc.ca

Secretary/ Secrétaire

Eric Rancourt
(613) 951-5046 (work/travail)
(613) 951-3100 (FAX/télécopieur)
email: eric.rancourt@statcan.ca

Treasurer/ Trésorier

Edward Chen
(613) 951-4769 (work/travail)
(613) 951-3100 (FAX/télécopieur)
email: chenedw@statcan.ca

President-Elect/ Président-désigné

John Nash
(613) 225-3781 (work/travail)
(613) 225-6553 (FAX/télécopieur)
email: jcnash@uottawa.ca

Past-President/ Président-sortant

Dan Harvey
(613) 824-0926 (work/travail)
(613) 824-2373 (FAX/télécopieur)
(Call 824-0926 before sending a fax)
email: danharvy@istar.ca

Program Coordinator/ Coordonnateur de programme

Dena Schanzer
(613) 946-0461 (work/travail)
(613) 954-5414 (FAX/télécopieur)
email: dena_schanzer@hc-sc.gc.ca

PROGRAM AND ABSTRACTS

9:15 Introductions

9:30-10:30 : **Assessment of Genetic Heterogeneity Using Regression Models for Allele Sharing**

Shelley B. Bull, Samuel Lunenfeld Research Institute of Mount Sinai Hospital,
Department of Public Health Sciences, University of Toronto

The objective of current genomewide studies of complex diseases is to find chromosomal regions that harbour susceptibility genes for the disease, but the presence of heterogeneity in the study sample will often reduce power to detect linkage. Differences in environmental exposures, gene-environment interaction, and gene-gene interaction represent important sources of heterogeneity among and within families. Regression models for allele-sharing assess linkage evidence in a likelihood framework in which overall tests, subgroup tests, and tests for heterogeneity can incorporate measured covariates suspected to be associated with heterogeneity. We will review recent approaches and current questions related to the use of covariates in allele-sharing linkage analysis.

10:30-11:00 Break

11:00-12:00 **Bounds on a Composite Measure of Linkage Disequilibrium: Application to Genetic Screening**

David Hamilton, Department of Mathematics and Statistics, Dalhousie University

David Cole, Professor, Departments of Laboratory Medicine & Pathobiology,
Medicine, and Paediatrics (Genetics), University of Toronto

In mendelian genetics, there is a one-to-one correspondence between mutation and disease phenotype. In the genetics of complex disorders (e.g., osteoporosis), multiple minor mutations may be responsible for a predisposition-to-disease phenotype. When such mutations are closely linked in the human genome (say within a single gene), the properties of their association (known as linkage disequilibrium or LD) are important in assessing the design of a screening test for genetic predisposition. This is particularly true if the utility of such a multi-gene (haplotype) screen is demonstrated for one subpopulation but is unknown for the second.

Measures of linkage disequilibrium typically assess lack of statistical independence through the discrepancy between the joint probability of two alleles and the product of their marginal probabilities. Such measures suffer from the disadvantage that their range of possible values depends on the marginal allele probabilities. This makes comparisons between subpopulations difficult. To obviate this problem, linkage disequilibrium measures are often expressed as a proportion of the most extreme value possible. In this talk we present bounds for a composite measure of linkage disequilibrium, which is appropriate when data is collected on genotypes and the phase of double heterozygotes cannot be determined. This extends earlier work of Lewontin and Weir and Cockerham who discuss relative measures of linkage disequilibrium for other types of data.

12:00-13:15 Lunch

13:15-14:15 Application of DNA Microarray Technology in Environmental Health Sciences

Paul A. White, Environmental & Occupational Toxicology Division, HECSB, Health Canada, Ottawa.

Andrew Williams, Biostatistics & Epidemiology Division, HECSB, Health Canada, Ottawa.

DNA microarrays (aka gene chips or gene arrays) provide research scientists with the ability to simultaneously track the expression level of every known gene in the genome of a selected organism. Many researchers have promoted the enormous utility of the technology in environmental toxicology and risk assessment. However, for the most part the claims remain untested. Successful use of DNA microarray technology is not simple. Microarray research must carefully consider experimental design, techniques for image capture, analysis, and interpretation, and methods for statistical data analysis. The sheer volume of data complicates each step. A single high-density microarray experiment can generate tens of thousands or even millions of values. Moreover, the data are subject to multiple sources of variation that are not biologically meaningful. The challenge for researchers is to remove systematic sources of variation, and identify interesting patterns of gene expression. Data analysis usually proceeds in four stages. The first stage involves data inspection, and removal of instrument-related rogue values. The second stage is normalization to remove systematic bias caused by variations in hybridisation efficiency, dye incorporation, and dye molecular fluorescence. The third stage involves data “filtering” and identification of expression values that are significantly altered relative to the control or reference RNA. The fourth stage involves the use of data analysis methods (e.g., clustering, PCA, discriminant analysis) to identify gene expression patterns. We are specifically interested in the hypothesis that toxic substances can be classified according to their influence on gene expression, and moreover, that techniques like the Naïve Bayesian classification can be employed to identify predictive gene sets.

14:15: -14:30 Break

14:30-15:30 The Statistical Analysis of Quantitative Traits

J.C. Loredó-Ostí, Department of Human Genetics, McGill University

Since it was outlined by Ronald Fisher and Sewall Wright, quantitative genetics has been playing a key role not only as the foundation for most of the plant and animal breeding programs but also in the understanding of the inheritance mechanism of complex human traits such as heart disease. The technical advances in molecular genetics and the explosion in quantitative traits data available pose the challenge of actually finding the genes that cause quantitative variation.

In this opportunity, we will talk briefly about the nature and modeling of quantitative variation, starting from segregation and linkage analysis in controlled crosses, sib-pair regression and other allele sharing based methods to quantitative trait analysis on large and complex pedigrees.

PROGRAMME ET RÉSUMÉS

9:15 Introduction

9:30-10:30 : **Évaluation de l'hétérogénéité génétique à l'aide de modèles de régression pour le partage de l'allèle**

Shelly B. Bull, Institut de recherche Samuel Lunenfeld de l'hôpital Mount Sinai,
Département des sciences de la santé publique, Université de Toronto

L'objectif des études de maladies complexes présentement en cours sur le génome complet est de trouver des régions chromosomiques qui portent des gènes favorisant la maladie. Cependant, la présence d'hétérogénéité dans l'échantillon à l'étude va souvent réduire la puissance de détection du lien. Parmi et à l'intérieur de familles, des différences d'exposition environnementale, des interactions gène-environnement et des interactions gène-gène représentent des sources importantes d'hétérogénéité. Les modèles de régression pour le partage de l'allèle permettent d'évaluer l'évidence de liens dans un cadre de travail utilisant la vraisemblance. Dans ce cadre, les tests globaux, les tests de sous-groupes et les tests d'hétérogénéité peuvent incorporer des co-variables mesurées qu'on soupçonne d'être associées à l'hétérogénéité. Nous passerons en revue les approches récentes et les questions de l'heure reliées à l'utilisation de co-variables dans les analyses de lien du partage de l'allèle.

10:30-11:00 Pause

11:00-12:00 : **Des bornes pour une mesure composite du déséquilibre localisé : application au dépistage génétique**

David Hamilton, Département de Mathématiques et Statistique, Université Dalhousie

David Cole, Départements de médecine en laboratoire et de pathologie, médecine, et pédiatrie (génétique),
Université de Toronto

Dans la génétique mendélienne, il y a une correspondance bi-univoque entre la mutation et le phénotype d'une maladie. Dans la génétique des désordres complexes (p. ex. l'ostéoporose), plusieurs mutations mineures peuvent être responsables d'un phénotype de prédisposition à la maladie. Lorsque de telles mutations sont très liées dans le génome humain (disons à l'intérieur d'un seul gène), les propriétés de leur association (linkage disequilibrium ou LD) sont importantes pour évaluer la conception d'un test de dépistage de prédisposition génétique. Ceci est particulièrement vrai si l'utilité d'un tel dépistage multi-gène (haplotype) est démontré pour une sous-population mais s'avère inconnue pour la seconde.

Habituellement, les mesures de déséquilibre localisé évaluent le manque d'indépendance statistique par la différence entre la probabilité jointe de deux allèles et le produit de leur probabilités marginales. De telles mesures ont malheureusement le désavantage d'avoir une étendue de valeurs possibles qui dépend des probabilités marginales des allèles. Ceci rend les comparaisons entre sous-populations difficiles. Pour éviter ce problème, on exprime souvent les mesures de déséquilibre local comme une proportion des valeurs possibles les plus extrêmes. Dans cet intervention, nous présentons des bornes pour une mesure composite du déséquilibre localisé; ce qui est approprié lorsqu'on recueille des données sur les génotypes et que la phase d'hétérozygotes doubles ne peut être déterminée. Ceci étend les travaux précédents de Lewontin et Weir et Cockerham qui discutent de mesures relatives de déséquilibre localisé pour d'autres types de données.

12:00-13:15 Dîner

13:15-14:15: Application de la technologie de micros-alignements d'ADN dans les sciences de la santé environnementales

Paul A. White, Division de toxicologie environnementale et occupationnelle, HECSB, Santé Canada, Ottawa

Andrew Williams, Division de biostatistique et d'épidémiologie Division, HECSB, Santé Canada, Ottawa

Les micro-alignements d'ADN (gene chips ou gene arrays) fournissent aux chercheurs scientifiques la capacité de retracer simultanément le niveau de chaque gène du génome d'un organisme donné. Plusieurs chercheurs ont fait la promotion de la très grande utilité de la technologie dans la toxicologie environnementale et l'évaluation des risques. Cependant, dans la plupart des cas, les affirmations demeurent non-testées. L'utilisation de la technologie des micro-alignements n'est pas simple. La recherche sur les micro-alignements doit tenir compte avec soin du plan d'expérience, des techniques de capture d'images, de l'analyse, de l'interprétation et des méthodes statistiques d'analyse de données. Le volume de données lui-même complique chaque étape. Une seule expérience de micro-alignements à haute densité peut générer des dizaines de milliers voire des millions de valeurs. De plus, les données sont sujettes à de multiples sources de variation qui n'ont pas de sens du point de vue biologique. Le défi des chercheurs consiste à retirer les sources systématiques de variation et à identifier les patrons intéressants d'expression des gènes. L'analyse des données comporte habituellement quatre étapes. La première étape consiste à explorer les données et enlever les erreurs grossières dues aux instruments. La deuxième étape est la normalisation, qui sert à enlever le biais systématique causé par des variations de l'efficacité d'hybridation, de l'inclusion de colorant et de la fluorescence moléculaire du colorant. La troisième étape consiste à « filtrer » et identifier des valeurs d'expression qui sont significativement modifiées par rapport au contrôle ou à l'ARN de référence. Enfin, l'utilisation de méthodes d'analyse de données (p. ex. classification, ACP, analyse discriminante) pour identifier les patrons d'expression génétique constitue la quatrième étape. Nous sommes intéressés spécifiquement à l'hypothèse que des substances toxiques peuvent être classifiées selon leur influence sur l'expression génétique, et de plus, que des techniques telles que la classification Bayésienne naïve peuvent être utilisées pour identifier les ensembles prédictifs de gènes.

14:15-14:30 Pause

14:30-15:30 : L'analyse statistique de traits quantitatifs

J.C. Loredó-Osti, Département de génétique humaine, Université McGill

Depuis que Ronald Fisher et Sewall Wright l'ont souligné, la génétique quantitative a joué un rôle clé non seulement comme fondation des programmes de reproduction végétale et animale, mais aussi dans la compréhension des mécanismes de transmission des traits complexes humains tels que les maladies cardiaques. Les progrès techniques de la génétique moléculaire et l'explosion de la disponibilité de données quantitatives sur les traits posent le défi de vraiment trouver les gènes qui causent des variations quantitatives.

Nous allons donc parler brièvement de la nature et de la modélisation de variations quantitatives, en commençant par les analyses d'isolement et de groupement dans des études croisées contrôlées jusqu'à l'analyse quantitative de traits sur des grandes et complexes lignées en passant par la régression « sib-pair » et d'autres méthodes basées sur le partage d'allèles.