THE SIMULATION OF LINKAGE IN HUMAN POPULATIONS

by

Alexander Frederick Wilson

a Thesis

Submitted to the faculty of the Graduate School in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Medical Genetics, Indiana University

September, 1980

Accepted by the faculty of the Department of Medical Genetics and the Graduate School of Indiana University in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Advisory Committee:

CHAIRMAN:

P. Michael Conneally, Ph.D.

Professor
Department of Medical Genetics

MEMBER:

James A. Norton Jr., Ph.D.

Professor of Biostatistics
Department of Psychiatry

MEMBER:

Pao-lo Yu, Ph.D.

Professor
Department of Medical Genetics

MEMBER:

John M. Gersting Jr., Ph.D.

Professor
Departments of Mathematical Sciences and Engineering Sciences

ABSTRACT

THE SIMULATION OF LINKAGE IN HUMAN POPULATIONS

Alexander F. Wilson

A mathematical model is developed to simulate the recombination of qualitative traits in two and multi-generation human pedigrees. Repeated sampling techniques are employed to empirically determine the distribution of the maximum likelihood estimate (MLE) of the recombination fraction in two generation double backcross matings. Estimates of the mean, variance, skewness and kurtosis of the MLE's are computed. The distribution is used to construct empirical confidence intervals, significance levels and power curves. Approximate normal curve tests for departures from the expected normal values of skewness and kurtosis, and chi square tests for departures from expected normal frequencies are calculated. All MLE's are considered regardless of their log likelihoods. Several transformations are employed in an attempt to normalize the distribution of the MLE of the recombination fraction.

The model is designed to emulate the actions of a single investigator conducting a linkage study. A sample (i.e. data collected in a linkage study) consists of a specified number of families of either constant sibship size or sibship size based on a negative binomial distribution. The data from the sample are analyzed (non-sequentially) and a decision made on the null hypothesis i.e. no linkage.

The control set has fully penetrant loci, and no genetic heterogeneity. Various degrees of undetected penetrance or genetic heterogeneity are introduced into the model in order to observe their effects on the distribution of the MLE of the recombination fraction, and on empirically derived confidence intervals, significance levels and power. The relative efficiency of linkage analysis under these various models is also considered.

The distribution of the MLE of the recombination fraction is generally non-normal when the true recombination fraction is less than 0.10 or greater than 0.40. The MLE is slightly positively biased when the true recombination fraction is small and negatively biased when the true recombination fraction is large.

When confidence intervals are constructed based on data with log likelihoods similar in magnitude to those reported in the literature the confidence intervals for all MLE's are quite wide, narrowing only as the log likelihoods become much higher (>20). Under these same circumstances the power of the likelihood ratio test to detect linkage is excellent over the range 0.01 to 0.15-0.20 with little power when the recombination fraction is greater than 0.30. Empirically derived significance levels appear to be approximately one half (or less) of those obtained by the log likelihood ratio evaluated as a chi square with one degree of freedom.

Undetected reduced penetrance or genetic heterogeneity positively biases the MLE of the recombination fraction, the degree of bias depending on the degree of penetrance or amount of heterogeneity. Small

to moderate amounts of undetected reduced penetrance or genetic heterogeneity have large effects on empirically derived confidence intervals, power, and relative efficiency, but virtually no effect on significance levels.

Under certain conditions the arcsin transformation approximately normalizes the distribution of the MLE of the recombination fraction over the range 0.01 to 0.20 with respect to skewness and kurtosis. Confidence intervals based on this transformation can be constructed analytically.

TABLE OF CONTENTS